



Book of extended abstracts

The Fifth Conference of  
the European Philosophy of Science Association

23 – 26 September 2015  
<http://www.philsci.eu/epsa15>

Duesseldorf Center for Logic and Philosophy of Science (DCLPS)  
Heinrich Heine University Duesseldorf, Germany

<http://dclps.phil.hhu.de/>

**EPSA15 is supported by**

Springer

De Gruyter

Mentis

Harvard University Press

Cambridge University Press

Landeshauptstadt Düsseldorf

Deutsche Forschungsgemeinschaft

Duesseldorf Center for Logic and Philosophy of Science

Heinrich Heine University Duesseldorf

European Philosophy of Science Association

**Edited by**

Alexander Christian, Christian J. Feldbacher, Alexander Gebharter, Nina Retzlaff, Gerhard Schurz & Ioannis Votsis

**Impressum**

Heinrich Heine University Duesseldorf

Department of Philosophy (Institut fuer Philosophie)

Universitaetsstrasse 1, Geb. 24.52

40225 Duesseldorf, Germany

## Table of Contents

<b>Abstracts</b> .....	<b>1</b>
Plenary Lectures .....	1
Symposia & Contributed Papers I.....	3
Symposia & Contributed Papers II.....	46
Symposia & Contributed Papers III.....	90
Symposia & Contributed Papers IV .....	147
Symposia & Contributed Papers V .....	189
Symposia & Contributed Papers VI .....	232
Symposia & Contributed Papers VII .....	275
Symposia & Contributed Papers VIII .....	321
Symposia & Contributed Papers IX.....	366
Poster Session .....	414
Pre-events .....	480
<b>Name Index</b> .....	<b>490</b>

## Abstracts

---

### Plenary Lectures

**Plenary Lecture I**

Chair: Stephan Hartmann

**Springer Lecture**

Room 5D, Wednesday 14:30 – 16:00

---

*Trendsetters and Social Change*

CRISTINA BICCHIERI

University of Pennsylvania

cb36@sas.upenn.edu

---

Trendsetters are the "first movers" in social change. To study the dynamics of change, we need to study the interplay between trendsetters' actions and individual thresholds. It is this interplay that explains why change may or may not occur.

**Plenary Lecture II**

Chair: Jan-Willem Romeijn

**De Gruyter Lecture**

Room 5D, Thursday 17:40 – 19:10

---

*Measuring Graded Membership: The Case of Color*

IGOR DOUVEN

CNRS (INSHS)

igor.douven@paris-sorbonne.fr

---

In my talk, I discuss Kamp and Partee's semantics for languages with vague predicates and especially the account of graded membership that is part of it. In its original presentation, the semantics is known to be incomplete, lacking a proposal for determining unique degrees of membership. It has recently been shown that the semantics can be completed by embedding it in the conceptual spaces framework, as developed in the cognitive sciences. It has also been shown that, in this version, the semantics is formally correct. However, the question of its material adequacy is still open. I report empirical work that addresses this question by testing the semantics in the domain of color. Specifically, a number of experiments are reported which are

meant to determine, on the one hand, the regions in color space where the typical instances of certain colors are located, and on the other hand, the degrees of membership in various color categories of a great number of different shades. From the locations of the typical regions in conjunction with Kamp and Partee's account follow degrees of membership for the color shades we are interested in. These predicted degrees are compared with the judged degrees, as obtained in the experiments.

**Plenary Lecture III**

Chair: Gerhard Schurz

**Plenary Lecture**

Room 5D, Saturday 16:30 – 17:30

---

*Causality in Dynamical Biological Mechanisms*

MARCEL WEBER

Université de Genève

marcel.weber@unige.ch

---

It is widely held that structural causal models based on interventionist criteria for causal asymmetry provide an adequate representation of causality in any kind of causal system. In this talk I examine a type of dynamical system that is typical for biological mechanisms in that it contains a causal feedback loop, namely a biological clock mechanism. Such mechanisms can be described qualitatively as well as quantitatively by using systems of coupled differential equations. While these equations cannot be solved analytically, they have approximate solutions using discrete time. I show that these discrete time model are fully representable as causal structural models. However, these models are not causally equivalent to the original differential model. In particular, the differential model shows a failure of modularity. This suggests that in such dynamical mechanisms interventionist causality is something that emerges only at coarse-grained, approximate descriptions of reality and not at the fundamental level of the mechanism.

## Symposia & Contributed Papers I

### Quo Vadis Selective Scientific Realism?

**Symposium**

Organizer: Peter Vickers

Chair: Ioannis Votsis

Room 5D, Wednesday 16:30 – 18:30

---

#### *Case Studies and Selective Realism*

ANJAN CHAKRAVARTTY

University of Notre Dame

chakravartty.1@nd.edu

---

#### *Does Realism Become More Reasonable as Theories Become More Successful?*

DAVID W. HARKER

East Tennessee State University

harkerd@mail.etsu.edu

---

#### *The Scientific Realism Debate in the Year 2015: A New Era of Realist Criteria and Non-Realist Historical Challenges*

TIMOTHY D. LYONS

Indiana University–Purdue University Indianapolis

tdlyons@iupui.edu

---

#### *Selective Realism: Theory Choice or Theory Synthesis?*

DEAN PETERS

University of Durham, University of Johannesburg

deandpeters@gmail.com

---

*Understanding the Selective Realist Defence Against the PMI*

PETER VICKERS

University of Durham

peter.vickers@durham.ac.uk

---

**General Description**

The challenge to scientific realism that emerges from the historical record continues to provoke significant discussion. The details may have changed since Laudan's canonical presentation, but the overall spirit remains the same. Simply put, radical theory change poses a problem for those scientific realists who want to make an inference from the success of science to the (approximate) truth of scientific hypotheses. Over the last few decades the principle realist response has been to develop some kind of *selective* realism, which seeks to distinguish those parts of a theory which warrant realist commitment from those that do not. Whilst there is wide agreement among realists that *some* version of selective realism is to be accepted, there is fierce disagreement as to *which* version of selective realism is most plausible. Candidates that are still actively discussed include the 'entity realism' associated with Hacking (1983) and Cartwright (1983), Worrall's (1989; 2006) 'structural realism' and Papineau's (2010) related 'Ramsey-sentence realism', Kitcher's (1993) 'working/presuppositional' distinction, and Psillos' (1999) defence of the selective strategy whereby realists attempt to '*divide et impera*'.

Participants in this symposium have contributed to the debate in various ways. Anjan Chakravartty (1998; 2007) has advocated 'semi-realism', designed to capture the core commitments shared by both entity realists and structural realists. David Harker (2012) has cautioned against certain versions of the selective realist thesis, and has argued for a comparative rather than absolute understanding of 'empirical success'. Recently, Peter Vickers (2013) and Dean Peters (2014) have attempted to make sense of the working posits idea by paying close attention to the internal logic of theoretical derivations, and how theoretical claims are deployed therein. However, while Vickers tends to regard as essential those theoretical propositions that are the result of combining multiple prior propositions, Peters argues for the essentialness of more general theoretical claims that contribute to multiple individual derivations. Finally, Tim Lyons has spent more than a decade publishing articles specifically on the topic of the historical challenge to selective scientific realism, and he has introduced a number of new historical case studies to the debate.

With regard to the historical challenge, selective realists embrace a common attitude: the historical fact that scientific theories, which we now believe to be largely false, generated highly significant predictive/explanatory successes can be rendered compatible with a realist philosophy of science, on the condition that the theoretical claims understood by the selective realist to be responsible for these successes

are not among those we regard as false. There remain, however, historical examples that create problems even for selective realists, since certain theoretical constituents appear central to a theory's success but are definitely not approximately true by current lights (on any reasonable theory of 'approximate truth'). A number of new examples have recently been introduced to the literature, including Kepler's predictions based on his theory of the *anima motrix* (Lyons 2006), Kirchhoff's theory of diffraction (Saatsi and Vickers 2011), Sommerfeld's prediction of the hydrogen fine structure (Vickers 2012), and Dirac's prediction of the positron (Pashby 2012). Vickers (2013) – drawing on previous work by Lyons and others – presents a long list of other possible examples. Thus the question remains whether scientific realism – even *selective* scientific realism – is consistent with the historical record.

Reconciling realism with history is a central challenge but realists are also sensitive to the importance of generating a thesis which is well motivated *independently* of the historical record. No doubt the historical record can teach us lessons about how current selective positions might be sensibly adjusted, but the realist's job would be too easy if she had *total* freedom to adjust her realism to match the history. Lyons's paper in the current symposium 'The scientific realism debate in the year 2015: a new era of realist criteria and non-realist historical challenges' addresses just these issues, identifying both (i) conditions any selective realism must meet, regardless of the history, and also (ii) case studies from the history which pose a serious threat to any such selective realism.

The debate has long been carried out with a focus on how the selective realist should identify the *working* assumptions, the bits of a theory we should commit to when that theory has enjoyed sufficient empirical success. Vickers's contribution to this symposium 'Understanding the selective realist defence against the PMI' takes issue with this approach, making a distinction between (i) the project of responding to historical challenges, and (ii) the project of identifying the 'working posits'. Whilst (ii) is one route to (i), a currently unexplored option for the selective realist is to tackle (i) by merely arguing that specific assumptions are 'idle'. Breaking the realist's job into two parts – projects (i) and (ii) – might help to make more transparent what the realist is trying to achieve, and whether it is possible at all.

David Harker's contribution 'Does scientific realism become more reasonable as theories become more successful?' further complicates the story, raising the question of how selective realism can be married with the fact that our confidence in scientific theories and hypotheses increases over time. Typically the selective realism debate has focused on individual scientific successes in the history of science, or collections of such successes, with little thought for the wider story of how the relevant science gradually developed in the months and years immediately before and after those successes. But, he argues, this broader story is crucial, since it is only by gaining an appreciation for the gradual growth of evidence that one can come to understand how scientists came to believe the corresponding theoretical hypotheses. The question to be explored here is whether the selective realist's simplification of the historical story – focusing on individual scientific successes – is an innocent simplification which is consistent with the more complicated historical



development of science, or whether the historical reconstructions of the selective realist are too far removed from real science to tell us much at all about how science works, or what we can reasonably believe.

This concern about the use of historical reconstructions can be pushed still further, threatening the very foundations of the debate. Within the philosophical community it has long been assumed that the history of science can bear very significantly on scientific realism – even to the extent that it can ‘falsify’ or ‘confirm’ a given realist position. Historians of science, however, often stress that historical reconstructions can never be fully ‘neutral’. In his paper ‘Case studies and selective realism’ Anjan Chakravartty questions the very framework of this debate on these grounds. It may be that the history of science can *bear* on scientific realism, but this hardly guarantees that it can settle disputes between realists and antirealists in the way that has so often been assumed on both sides of the debate.

Another fundamental issue concerns the very idea of *selective* realism if we think that even a very modest/partial ‘meaning holism’ is to be favoured. Selective realism is *prima facie* based on the idea that one can take any individual part of a scientific theory in isolation and – if that part is ‘doing work’ – make a doxastic commitment to it. However, many believe that the semantic content of an individual equation, or model, or hypothesis, differs significantly depending on the wider theoretical framework in which it is embedded. In his paper ‘Selective realism: theory choice or theory synthesis?’ Dean Peters addresses this concern, bringing to life the issue of ‘semantic content’ via discussions of (i) ‘particulate’ versus ‘blended’ accounts of inheritance at the beginning of the 20th century, and (ii) the ‘prion revolution’ of the 1980s and 90s. Peters defends selective realism here, on the grounds that there is a natural way to make selective realism consistent with at least a partial meaning holism, and such that selective realism is actually supported by the history of science.

Overall in this symposium two broad issues are to be addressed: (1) what general criteria should be satisfied by an epistemically acceptable variant of selective realism; and (2) how is the debate between the realist and anti-realist to be most profitably conducted? In addressing these issues, participants anticipate that this symposium will help to set the agenda for the short-term evolution of the selective realism debate.

## Abstracts

### 1. Anjan Chakravartty: *Case Studies and Selective Realism*

Case studies of past and present science, concerning both the interpretation of specific theories and the nature of theory change over time, are often presented as evidence for or against the viability of forms of *selective* realism: versions of scientific realism that advocate belief in certain components of theories, as opposed to their entire descriptive content. In this talk I consider the question of how probative case study evidence can be in this domain, focusing on three prominent versions of selectivity: explanationism; entity realism; and structural realism. In each case I suggest that while case studies are essential to philosophical

analysis, they are not compelling in the way that many suggest. I offer three arguments in support of this thesis.

The first argument concerns debates about explanationism, the attempt to ground selective realism in components of scientific theories that describe what is required to explain their empirical success. I argue that several worries, including concerns about the “neutrality” of historical narrative, problematize the role of case studies here. A second argument focuses on entity realism. While this version of selective realism is sometimes marshaled in response to antirealist arguments stemming from historical discontinuities in scientific theorizing over time, I contend that considerations for and against it are quickly and naturally dissolved into disputes about how best to characterize the semantics of theoretical terms, which are neatly insulated from the details of cases. A final argument focuses on structural realism. While the problematic here varies according to the particular structuralist hypothesis at issue, in each case, I maintain, the tenability of the proposal rests not on cases *per se*, but on a number of conceptual issues concerning the notions of structure involved.

## **2. David W. Harker: *Does Realism Become More Reasonable as Theories Become More Successful?***

Central to almost all defenses of scientific realism is the conviction that scientific success, appropriately defined, is evidence for the approximate truth of those parts of scientific theories that are somehow responsible for the success. Efforts to unearth the most plausible version of this inference have helped generate a significant literature. However, what’s not straightforwardly reconcilable with the underlying sentiment is the possibility that realist attitudes towards certain entities, structures, properties, dispositions, and so on, might become *more* reasonable over time. That’s to say, whether particular constituents of a scientific theory stand in the right relationship to that theory’s successes is typically advanced as a straightforward dichotomy. Hence, whether one should adopt a realist attitude towards those constituents appears similarly dichotomous. Yet often it appears that our confidence in scientific conclusions strengthens over time, as more qualified individuals have chance to review available evidence, as more evidence - of perhaps very different types - becomes available, as new technologies and methods enable us to improve upon the quality of available evidence, and as these conclusions become central to frameworks of increasing scope, precision and fecundity.

In this paper I consider how selective realists might capture the intuition that realist attitudes can become more reasonable. I suggest this requires paying more attention to the ways in which scientific concepts are refined in response to anomalies and new data, and the ways those refinements help us explain both the success and failures of earlier models. I argue that analyzing scientific successes -without regard for the scientific context in which those successes appear - is to misrepresent an important aspect of scientific argumentation. The prospects for scientific realism will improve, I suggest, if we attend to patterns within the evolution of lineages of scientific research programs.

### **3. Timothy D. Lyons: *The Scientific Realism Debate in the Year 2015: A New Era of Realist Criteria and Non-Realist Historical Challenges***

The scientific realism debate has now reached an entirely new level of sophistication. Faced with increasingly focused challenges, epistemic realists have appropriately revised their basic meta- hypothesis that successful scientific theories are approximately true: they have emphasized criteria that render realism far more selective and, so, plausible. Mindful of these advances, I articulate a set of conditions that must be met for a selective realist criterion to be viable. The theoretical constituents it picks out must be, not only ascertainable/identifiable, they must also be explanatorily relevant—genuinely *responsible* for success— and sufficiently realist, reaching to a level *deeper* than the empirical data.

With these conditions in hand, I target what I take to be the set of most influential realist criteria now on offer. After briefly flagging some that fail to meet the above requirements, I point to a subset that nonetheless lives up to them. With the latter identified, however, I advance novel case studies to challenge these forms of realism. Although historical challenges to realism have tended to elaborate on those theories in Laudan's infamous list (theories positing phlogiston, caloric, ether) and to focus on the late 18th and early 19th centuries, I break from this path: I look to the history of astrophysics, and I emphasize 20th century advances. Moreover, I offer a novel account of the nature of the historical threat to realism, articulating a set of purely deductive arguments. I contend that both the content and form of these novel challenges severely threaten even the least demanding variants of selective epistemic realism.

I conclude on a positive note, however, arguing that scientific realism need not be rejected outright, that a number of central realist tenets can be retained unproblematically even in the face of such threats to the epistemic component of scientific realism.

### **4. Dean Peters: *Selective realism: theory choice or theory synthesis?***

In this paper, I explore a tension between "selective" scientific realism and the incommensurability thesis associated with Kuhn and Feyerabend. The incommensurability thesis states that the meaning of a theoretical term is constituted by its relations to other concepts within that theory, so that terminological continuity between theories conceals more radical conceptual discontinuity. Hence the Kuhnian picture of theory change as the outcome of stark choice or "revolution".

Selective realism claims that realist commitment should be reserved only for *parts* of scientific theories - usually those essential to their empirical success - and that these parts tend to be preserved in later theories. Since it emphasises continuity between *parts* of theories, often in isolation from their original context, the formulation of a specifically *selective* realism foregrounds the tensions between realism and Kuhnian holism. Moreover, if more than one theory is empirically successful at a given time, the selective realist must believe that each is at least approximately and/or partially true, which seems to rule out stark choice between them.

Arguably, selective realism is so at odds with holism that it suggests the elimination of "theory" talk altogether (cf. Vickers 2013), in favour of more fine-grained talk regarding the postulation, confirmation/refutation and combination of theoretical "parts".

This view is apparently supported by historical episodes where a "revolutionary" theoretical realignment is best described as the result of *synthesis* between competing sets of theoretical claims. The first case I discuss is the neo-Darwinian synthesis of the early twentieth century, which succeeded in reconciling "particulate" and "blended" views of inheritance, while also giving a satisfactory account of natural selection. The second case I discuss is the "prion revolution" of the 1980s and 90s, wherein the "central dogma" of molecular biology was reconciled with the idea that biological information can be stored and transmitted by proteins.

### **5. Peter Vickers: *Understanding the Selective Realist Defence against the PMI***

Responding to the PMI, selective scientific realists make a distinction between the 'working' hypotheses of a theory, which warrant realist commitment, and the 'idle' parts of a theory, which do not. This strategy certainly helps the realist answer *some* historical challenges – cases where false theories enjoyed significant predictive (possibly explanatory) success – but there remain historical examples where what very much *seem* to be working hypotheses are definitely not approximately true. A number of new examples have recently been introduced to the literature, including Kirchoff's theory of diffraction, Sommerfeld's atomic theory, and Dirac's prediction of the positron. Thus the question remains whether scientific realism is consistent with the historical record.

In recent years much of the discussion has concerned how (and whether) the selective realist can define 'working', such that the definition is properly motivated (not ad hoc), and can be used to rebut PMI-style objections. However, two distinct realist projects have been conflated: (i) the project of responding to PMI-style objections, and (ii) the project of explaining what realists should commit to/what the working posits are. Engaging in project (ii) is certainly one way to tackle project (i), but there is an easier, more direct way. Any historical challenge to selective realism will consist of a claim that some posit is both working and not approximately true. Thus to respond the realist need only argue that the posit in question is *not* working, and it is quite possible to have a good argument that a hypothesis is idle without having any general theory of what it takes for a hypothesis to be 'working'. This also leads to an interesting new possibility: even if realists cannot hope to predict what will be retained in future scientific theories, they might be able to make predictions concerning what will *not* be retained.

**Measure Sensitivity in the Study of  
Reasoning and Cognition**

**Symposium**

Organizer: Gustavo Cevolani, Vincenzo Crupi & Roberto Festa

Chair: David Atkinson

Room 5E, Wednesday 16:30 – 18:30

---

*Criteria for the Deciding Between Confirmation Measures*

PETER BRÖSSEL

Ruhr-University Bochum

peter.broessel@rub.de

---

*Measure Sensitivity in Verisimilitude Theory*

GUSTAVO CEVOLANI

University of Turin

g.cevolani@gmail.com

---

*Shannon and Beyond: Generalized Entropies and Rational Information Search*

VINCENZO CRUPI

University of Turin

vincenzo.crupi@unito.it

---

*Coherentism, Pluralism and Measure Sensitivity*

MICHAEL SCHIPPERS

University of Oldenburg

mi.schippers@uni-oldenburg.de

---

*Probabilistic Explications of Causal Strength*

JAN SPRENGER

Tilburg University

j.sprenger@tilburguniversity.edu

**General Description**

**The problem of measure sensitivity.** Formal models of rational inference are increasingly used in philosophy and related fields, including formal epistemology, cognitive science, and computer science. These models are usually put to work in the following way: first, a target problem is stated in terms of one or more central notions; second, a model is introduced which (supposedly) captures the central features of the relevant notion(s); finally, a solution to the initial problem is proposed by showing that, once adequately applied to the problem at issue, the formal model chosen demonstrably favors a unique possible answer over the others. To mention but one example, the Bayesian approach in philosophy of science (Howson and Urbach 2006), epistemology (Bovens and Hartmann 2003), cognitive science (Oaksford and Chater 2007) and artificial intelligence (Korb and Nicholson 2004) tackles central issues in these areas relying on various probabilistic models of reasoning and rational belief.

While the role of model-based argumentation in current philosophical and epistemological analysis is widely acknowledged, a crucial problem with this common practice remains prominent. This is the problem of *measure sensitivity*: several theoretical arguments, it turns out, are not invariant across different and otherwise plausible models. More precisely, the soundness of these arguments is critically sensitive to the choice of a particular formal explication. Measure sensitivity is arguably a challenging feature of current philosophical argumentation, and represents an important issue in different fields. On the other hand, it naturally fosters the search for *robust* arguments, i.e., arguments which remain sound regardless of the choice of a specific model within a sizable set of relevant variants.

**The exemplary case of measure sensitivity in Bayesian confirmation theory.** A vivid illustration of the problem of measure sensitivity has attracted the attention of philosophers of science in the context of Bayesian confirmation theories (Festa 1999, Fitelson 1999, Brössel 2013, and Crupi 2014). Here, a (micro)model of inductive support is given in the form of a probabilistic measure of confirmation, expressing how much a piece of evidence  $e$  supports or undermines a given hypothesis  $h$ . According to the “incremental” view of confirmation, the degree of confirmation of  $h$  on  $e$  measures the change in the probability of  $h$  brought about by learning  $e$ ; i.e., how much the final probability  $P(h|e)$  is increased or decreased as compared to the initial probability  $P(h)$ . A variety of measures of this kind have been put forward, which are now well-known to be non-equivalent, even in ordinal terms (Fitelson 1999; Crupi, Tentori and Gonzalez 2007; Crupi, Festa, and Buttasi 2010; Festa 2012). As an example, consider two of the most widely known incremental measures, i.e, the *difference* measure

$$D(h,e) = P(h|e) - P(h),$$

and the *ratio* measure

$$R(h,e) = P(h|e)/P(h).$$

Of course, measures  $D$  and  $R$  will typically assign different numerical values to the degree of confirmation of  $h$  on  $e$ . More importantly, however,  $D$  and  $R$  can impose *distinct orderings* over hypothesis/evidence pairs. Consider for instance two hypotheses  $h$  and  $h^*$  such that  $P(h) = .1$  and  $P(h|e) = .9$ , whereas  $P(h^*) = .01$  and  $P(h^*|e) = .1$ . As it is easy to check,

$$D(h,e) = .8 > .09 = D(h^*,e),$$

while

$$R(h,e) = 9 < 10 = R(h^*,e).$$

In other words,  $e$  confirms  $h$  more than  $h^*$  according to measure  $D$ , but the ordering is reversed if  $R$  is assumed as the relevant measure instead. To cite but one other example, consider the following *likelihood ratio* confirmation measure:

$$L(h,e) = P(e|h)/P(e|\neg h)$$

Whereas both  $D$  and  $R$  express the increase in the degree of belief in the truth of  $h$  given evidence  $e$ ,  $L$  measures the discriminatory power of  $e$  with respect to  $h$  and its negation.  $L$  then captures a remarkably different intuition about confirmation, and it comes as no surprise that comparative assessments of the confirmation of different hypotheses can be reversed when  $L$  is adopted as the relevant measure, instead of  $D$  or  $R$ .

The fact that some incremental measures are not ordinally equivalent was already acknowledged by Carnap (1962) and other earlier confirmation theorists. Only more recently, however, the implications of such non-equivalence for Bayesian confirmation theory and inductive logic in general have been fully appreciated. In particular, it can be shown that a number of arguments surrounding confirmation theory are sensitive to the choice of measure, in the sense that their soundness depends on which measure of confirmation is adopted. These include Bayesian solutions of popular problems such irrelevant conjunction (see Crupi and Tentori 2010, Fitelson 2012), the paradox of the ravens, the grue paradox, and more besides (Fitelson 1999, Huber 2008, Festa 2012, Brössel 2013). In these and others cases, the proposed solution only holds as far as some specific confirmation measures are employed, but fails to deliver the desired result when different, non-equivalent measures are adopted. Accordingly, measure sensitivity is seen as “one of the biggest challenges” confronting contemporary Bayesian confirmation theory (Brössel 2013, 378).

**Other examples.** Given the widespread use of formal models in the study of reasoning and cognition — often in the form of measures of some kind —, it has to be expected that the problem of model sensitivity is a significant issue in other areas, too, besides confirmation theory. Indeed, this is what recent research reveals, by highlighting the consequences of the plurality of non-equivalent models in a number of different fields.

Some examples will serve as an illustration. The problem of optimal information search, a key topic in statistics, philosophy of science and the analysis of clinical reasoning, is presently addressed by means of a number of models based on (some version of) epistemic utility theory. Much as for confirmation, these

models are non-equivalent and often deliver diverging results (Nelson 2005, 2008; Crupi and Tentori 2014). Similar problems arise when epistemic utility theory is applied to the justification of norms of probabilistic reasoning, given the plurality of different measures available to assess the overall accuracy of belief states, usually construed as the main epistemic goal in these literature (Leitgeb and Pettigrew 2010, Levinstein 2012, Pettigrew 2013). Finally, measure sensitivity turns out to be a significant issue in the analysis of some central notions in the philosophy of science and formal epistemology, including coherence (Schippers 2014, 2015), explanatory power (Schupbach 2011, Schupbach and Sprenger 2011, Crupi and Tentori 2012), and verisimilitude or truthlikeness (Oddie 2014, Niiniluoto 2003, Schurz and Weingartner 2010, Cevolani, Festa, and Kuipers 2013).

**Aims of the symposium.** In all the above mentioned areas, existing inquiries and results are significantly sensitive to the choice of a particular model, such a choice being sometimes left with no firm theoretical basis. This motivates the present symposium proposal, whose general aim is to explore the problem of measure sensitivity in a number of variations, to assess its general significance, and articulate its specific implications. The five proposed contributions will address this problem in the light of the following open issues.

- First, to appreciate the relevance of the problem of measure sensitivity, an exploration is needed of how widespread this problem is across different fields of inquiry: philosophy of science, formal epistemology, and cognitive science provide a highly valuable domain for an integrated assessment.
- Second, the viability and desirability of different possible solutions to the problem of measure sensitivity have to be assessed. In this connection, it is not even clear whether robustness — i.e., *insensitivity/invariance* to the choice of a model — is a realistic, general aim to strive after, or if a more modest, piecemeal strategy is in order to handle the plurality of formal models and of their different applications. An exploration of this issue will also shed new light on the relative merits of “monist” vs. “pluralist” positions in the debate on the adequacy of different modes in a variety of contexts.

Finally, the problem of measure sensitivity raises some doubts concerning the practice of model-based argumentation itself. Are there intrinsic limitations to this kind of inquiry? Is measure sensitivity an insurmountable barrier or a plain, unavoidable feature of sound philosophical argumentation?

## Abstracts

### 1. Peter Brössel: *Criteria for the deciding between confirmation measures*

In confirmation theory the problem of measure sensitivity (Festa 1999, Fitelson 1999, Brössel 2013) boils down to this. There are various non-ordinally equivalent confirmation measures that cannot measure the same quantity. If they measured the same quantity, then they would be at least ordinally equivalent. In this paper we evaluate various approaches how to decide between various non-ordinally equivalent measures of confirmation. In particular, we investigate whether considerations of intuitive desiderata, of the level of



measurement, of the purpose of confirmation theories, or of conformity with other important measures of epistemology (such as explanatory, unificatory, and systematic power) might resolve the problem. Contra most philosophers we argue that intuitive desiderata cannot solve the problem of measure sensitivity. Contra Vassend (2014) we also argue that considerations of the level of measurement cannot help to narrow down the field of potential confirmation measures. However, considerations of the purpose of confirmation theory and of the conformity or coherence of confirmation measures with various other epistemic measures can narrow down the field of confirmation measures considerably. In particular, for every purpose of confirmation theory one can hope to find a distinguished confirmation measure that is up to ordinal equivalence the best explication of the quantitative notion of confirmation for that purpose. However, this also shows that there is not just one best explication of our everyday notion of confirmation, but different explications for different purposes.

## **2. Gustavo Cevolani: *Measure sensitivity in verisimilitude theory***

Verisimilitude or truthlikeness — i.e., closeness or similarity to the whole truth about a given domain — plays a central role in prominent discussions of scientific progress, the aims of science, and the realism/antirealism issue. The problem of adequately defining this notion, first raised by Popper in the early sixties, has proved harder than initially thought. Today, at least three different approaches to verisimilitude have been distinguished and a number of measures exist, aiming at explicating what does it mean for a hypothesis or theory to be closer to the truth than another one. The ongoing debate on the adequacy of competing verisimilitude measures has highlighted some key principles, allegedly expressing crucial intuitions concerning the very idea of truth approximation. I focus on three such principles, which make explicit some basic conceptual connections among truth (and falsity), informative content, and verisimilitude. I then emphasize that each of these principles is violated by at least one well-known measure of verisimilitude in the literature. Since most philosophical arguments surrounding verisimilitude theory rely on one or more of these principles, their soundness turns out to depend on the specific measures adopted to explicate the idea of closeness to the truth. This is the problem of measure sensitivity, which arguably challenges the theory of verisimilitude, as well as Bayesian confirmation theory and other areas of formal philosophy of science. After considering some illustrations of this problem in the verisimilitude literature, I briefly discuss its relations to the well-known problem of language dependence, and its implications for current theoretical work on competing approaches to measuring truthlikeness.

## **3. Vincenzo Crupi: *Shannon and beyond: Generalized entropies and rational information search***

We present a biparametric family of measures known as Sharma-Mittal entropies. This is a very comprehensive framework in contemporary information theory. It includes a number of important special cases arising across different fields such as generalized statistical thermodynamics, the study of biological

and ecological diversity, and analyses of political and demographic concentration/fragmentation. Mathematically, relevant parameter settings in the Sharma-Mittal formalism generate classical Boltzmann/Shannon entropy as well as Rényi and Tsallis entropies, and more besides. We discuss applications for human reasoning, where entropy represents epistemic uncertainty and uncertainty reduction, in turn, is a major kind of epistemic utility. We argue that the Sharma-Mittal parameters can be given meaningful theoretical interpretations in this domain. We also explain how a variety of models ensue concerning the expected informational value of experiments, thus of information search options. Finally, we explicate formal and conceptual connections with related branches of the literature, including probabilistic theories of confirmation and scoring rules.

#### **4. Michael Schippers: *Coherentism, pluralism and measure sensitivity***

Propositions cohere to the extent they agree or dovetail with each other. The concept of coherence plays an important role in the theory of epistemic justification and in legal reasoning. The last 15 years have seen a large number of probabilistic proposals trying to explicate the notion of coherence, which is notorious for its elusiveness. In evaluating these proposals, the reference to particular test cases has more and more been replaced by a study of adequacy constraints. Unfortunately, however, it turned out that for each adequacy constraint there is at least one extant measure violating it. Moreover, it can easily be shown that the set of common adequacy constraints, albeit intuitively well-motivated, is plainly inconsistent. In this talk I discuss some recent results that are intimately connected with the problem of measure-sensitivity, as prominently discussed in the literature on Bayesian confirmation theory. To this end I focus on adequacy constraints highlighting the relationship between coherence on the one hand and concepts such as probabilistic independence, logicity, truth- and reliability-conduciveness, inconsistency and disagreement on the other. After presenting some formal results I address the question of how to interpret them. More precisely, I argue that the problem of measure sensitivity, as it affects probabilistic measures of coherence, should be considered an argument for (a moderate) pluralism with respect to the underlying explicatum.

#### **5. Jan Sprenger: *Probabilistic explications of causal strength***

Is IQ dependent more on nature or nurture? Is lack of water or lack of sun more responsible for the death of your garden plant? What is the relative impact of the referee's and the coaches' decisions on the outcome of a football match? Questions like these are ubiquitous in science and everyday life and motivate the search for a measure of causal strength. Since probability is the primary scientific tool for uncertain reasoning, our intended measure of causal strength should be probabilistic, too.

First, following I.J. Good's axiomatic approach, we develop adequacy conditions for a measure of causal strength that take into account recent lessons from the structural equations approach to modeling causation (e.g., pathwise vs. generic causation). On the basis of those conditions, we evaluate various probabilistic

measures of causal strength from the psychology and cognitive science literature. In this way, we shall assess the prospect of uniqueness results for a measure of causal strength.

However, these results may not converge on a single measure. Empirical data are required in order to find out which of several theoretically defensible measures shows the greatest agreement with the way scientists and laypersons reason. Although probabilistic models of causal judgments have been assessed empirically, such studies leave open a number of methodological and theoretical issues. One of the most prominent ones is the worry that participants in these experiments might not correctly understand their task and conflate cognate concepts with each other. Plausible candidates for such a conflation are, among others, the concepts of causal strength and explanatory power. Therefore, we elicit causal judgments in probabilistic scenarios, and map those judgments on measures of causal strength. We analyze the results in two ways: First, a statistical analysis will reveal whether the measures of causal strength that have survived the theoretical scrutiny indeed fit the empirical data. Second, we determine whether participants clearly separate causal strength and related concepts (e.g., explanatory power). Apart from clarifying the notion of causal strength, our results also imply valuable insights into the measure sensitivity of causal reasoning.

## References

- Bovens, L. and Hartmann, S. (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.
- Brössel, P. (2013). The Problem of Measure Sensitivity Redux. *Philosophy of Science* 80 (3): 378-397.
- Cevolani, G., Festa, R., and Kuipers, T.A.F. (2013). Verisimilitude and belief change for nomic conjunctive theories. *Synthese* 190 (16): 3307-3324.
- Crupi, V. (2014). Confirmation. *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2014/entries/confirmation/>>.
- Crupi, V. and Tentori, K. (2010). Irrelevant conjunction: Statement and solution of a new paradox. *Philosophy of Science* 77 (1): 1-13.
- Crupi, V. and Tentori, K. (2012). A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems). *Philosophy of Science* 79 (3): 365-385.
- Crupi, V., Tentori, K., and Gonzalez, M. (2007). On Bayesian measures of evidential support: Theoretical and empirical issues. *Philosophy of Science*, 74, 229-252.
- Festa, R. (1999). Bayesian Confirmation. In M. C. Galavotti and A. Pagnini (eds.), *Experience, Reality, and Scientific Explanation*, Dordrecht: Kluwer, pp. 55–87.
- Festa, R. (2012). “For unto every one that hath shall be given”. Matthew properties for incremental confirmation. *Synthese* 184 (1):89-100.
- Fitelson, B. (1999). “The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity”. *Philosophy of Science* 66 (Proceedings): S362–S378.
- Fitelson, B. (2012). Contrastive Bayesianism. In Martijn Blaauw (ed.), *Contrastivism in Philosophy: New Perspectives*. Routledge
- Howson, C. and Urbach, P. (2006). *Scientific Reasoning: The Bayesian Approach*, Open Court, La Salle.
- Huber, F. (2008), “Assessing theories, Bayes style”, *Synthese*, 161: 89-118.
- Korb, K.B. & Nicholson, A.E. (2004). *Bayesian Artificial Intelligence*. Chapman & Hall/CRC.
- Leitgeb, H. and Pettigrew, R. (2010). An Objective Justification of Bayesianism I-II. *Philosophy of Science* 77 (2): 201-235, 236-272.
- Levinstein, B. (2012). Leitgeb and Pettigrew on Accuracy and Updating. *Philosophy of Science* 79 (3): 413-424.

- Nelson, J.D. (2005). Finding useful questions: On Bayesian diagnosticity, probability, impact, and information gain. *Psychological Review*, 112, 979-999.
- Nelson, J.D. (2008). Towards a rational theory of human information acquisition. In N. Chater & M. Oaksford (eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. OUP Oxford.
- Niiniluoto, I. (2003). Content and Likeness Definitions of Truthlikeness. In *Philosophy and Logic in Search of the Polish Tradition*, a cura di J. Hintikka et al., 27–35. Springer.
- Oaksford, M. & Chater, N. (2007). *Bayesian Rationality: The probabilistic approach to human reasoning*. OUP Oxford.
- Oddie, G. (2014). Truthlikeness, *The Stanford Encyclopedia of Philosophy*, E. N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2014/entries/truthlikeness/>>.
- Pettigrew, R. (2013). Epistemic utility and norms for credences. *Philosophy Compass*, 8, 897-908.
- Schippers, M. (2014). Probabilistic measures of coherence: from adequacy constraints towards pluralism. *Synthese* 191 (16): 3821-3845.
- Schippers, M. (2015). Towards a grammar of Bayesian coherentism. Accepted for publication in *Studia Logica*.
- Schupbach, J.N. (2011). "Comparing Probabilistic Measures of Explanatory Power." *Philosophy of Science* 78 (5): 813–29.
- Schupbach, J.N. & Sprenger, J. (2011). The Logic of Explanatory Power. *Philosophy of Science* 78 (1): 105-127.
- Schurz, G. & Weingartner, P. (2010). Zwart and Franssen's impossibility theorem holds for possible-world-accounts but not for consequence-accounts to verisimilitude. *Synthese* 172 (3): 415-436.

---

*What is a Ceteris Paribus Law?*

CARSTEN HELD

Universität Erfurt

carsten.held@uni-erfurt.de

---

**1.**

According to Hempel and Oppenheim's classic deductive-nomological (DN) account, a scientific explanation is an argument for the explanandum using a law of nature. It is ideal in the sense that the argument for the explanandum is logically sound; an everyday explanation is a sketch of such an ideal explanation. Accordingly, in an everyday explanation of the form 'B because A', where A and B are propositions, citing the factor expressed by 'A' is thought to be replaceable by citing 'A & L', where L expresses a law and is of such a form that A & L entails B; the explanans A is said to be nomically sufficient for explanandum B.

A well-worn example from classical mechanics is the mathematical pendulum: from Newton's Second Law and using familiar idealisations for the pendulum we derive an equation of motion ( $d^2 \vartheta / dt^2 + (g/l) \sin \vartheta = 0$ ), which (using the small-angle approximation  $\sin \vartheta \approx \vartheta$  and choosing  $\vartheta(t_0) = \vartheta_0$  and  $(d\vartheta/dt)(t_0) = 0$ ) we solve as:  $\vartheta(t) = \vartheta_0 \cos((g/l)^{1/2} t)$ . Understanding  $\vartheta(t)$  and  $\vartheta_0$  as limiting cases of events we can interpret this as a kind of process law: For arbitrary t, if the pendulum bob is released at  $\vartheta(t_0) = \vartheta_0$ , then it moves through  $\vartheta(t)$  at t. With this law in hand, we can *deduce* the bob's position at t, for arbitrary t, from the initial conditions and *thereby* can explain a real pendulum's position from law and initial conditions, given that it approximately conforms to the idealizations.

The DN account has two well-known defects. First, it allows us to cite factors nomically sufficient but not necessary for the explanandum, while an acceptable explanation requires such a necessary factor; the classic counterexamples all exploit this defect. Second, the account asks us to argue for the explanandum using a process law: a proposition universally connecting two types of events. But it has long been suspected that such process laws are never generally true. Suppose, e.g., that a real pendulum approximately conforms to the relevant idealizations, then still there are countless possibilities for the bob to be released at  $\vartheta(t_0) = \vartheta_0$  and not move through  $\vartheta(t)$  at t, for any t. These possibilities just depend on the fact that the envisaged situation, even given the idealizations, is compatible with countless outside interferences. We do not expect that just by chance these possibilities never materialize. So we think that in its generality our process law is false and thus cannot be a premise in a sound argument, as the DN account would have it. The DN account, we see, is fundamentally defective but this leaves open to which extent it is misguided. Here, I argue that an

ideal explanation is indeed a sound argument using a natural language (NL-) conditional, suitably interpreted. The latter can be viewed as an instance of a *Ceteris Paribus* (CP-) law.

## 2.

Let  $A$  and  $B$  be propositions and let their truthmakers  $A$  and  $B$  be situations (parts of possible worlds). One situation can literally contain another and two situations are disjoint iff none contains any part of the other. A situation makes  $A$  true iff its existence is necessary and sufficient for  $A$ 's being true. Moreover, a situation makes  $B$  true (false) by making  $A$  true if it makes both  $A$  and  $B$  ( $\neg B$ ) true and, for every situation,  $B$ 's ( $\neg B$ 's) being false in it entails  $A$ 's being false in it. A situation *subordinates*  $B$  to  $A$  iff it makes  $B$  true or false by making  $A$  true and it *coordinates* propositions  $A$  and  $B$  iff it makes them true without subordinating either one to the other.

NL-conditionals, it turns out, are no material conditionals; instead the conditional clause is best understood as an adverbial clause, containing a tacit quantifier (Lycan 2001). As a consequence, 'If  $A$ , then  $B$ ' is best interpreted as ' $B$ , in any situation in which  $A$ ' or ' $(\forall x) (\text{In}(x, A) \rightarrow \text{In}(x, B))$ ', where ' $\text{In}(x, A)$ ' abbreviates ' $\text{In situation } x, A$ '. (' $\rightarrow$ ' is the material conditional.) The quantifier domain is assumed *to exclude situations making  $B$  false without subordinating it to  $A$* . NL-conditionals, thus analyzed, can be used to construct sound arguments functioning as explanations. So our process law should, within the context of explaining the pendulum bob's motion, be replaced by an explicated NL-conditional: 'If, in  $x$ , the pendulum bob is released at  $\vartheta(t_0) = \vartheta_0$ , then, in  $x$ , it moves through  $\vartheta(t)$  at  $t'$ , where  $x$  is a situation as specified. But our process law is a CP-law. It now suggests itself that CP-laws should be understood as explicated NL-conditionals functioning in explanations.

## 3.

My proposal has two immediate implications that have not been fully appreciated in the literature. (See, among many others, e.g. Pietroski/Rey 1995, Woodward 2002.) First, CP-laws should not be considered in isolation without the context of explanation that brings them about. There are no CP-laws out there that mysteriously hold without holding in full generality and then, just by chance, come handy for our explanatory needs. Second, it is misguided to look for falsifiers of CP-laws and then buttress the laws against them by inserting antecedents claiming the presence of some completer or the absence of some interferer. Within the context of an explanation, there simply are no falsifiers for a specific reason. The situations making the explanandum false *without* making the explanans true are excluded from the NL-conditional's domain and rightly so because, by presupposition, the explanandum is true and these situations certainly cannot contribute to its explanation. The situations making the explanandum false *by* making the explanans true are included in the domain and rightly so, because if any one of them exists the NL-conditional is false, is not a CP-law and does not contribute to a sound argument for the explanandum. Of course, not any NL-conditional

is a CP-law – but the missing additional factor is easy to come by. CP-laws are NL-conditionals derived from functional laws of nature.

---

*Empirical Problems for Explanationism*

RUNE NYRUP  
Durham University  
nyrup.rune@gmail.com

---

Explanationism is the view that inference to the best explanation (IBE) is a reliable guide to true scientific theories (or at least theories likely to be approximately or partially true). This view faces a prominent problem (e.g. Cartwright 1983): why should the explanatory qualities of theories be a reliable guide to their truth? *Prima facie*, virtues like being simple, unifying or in some other sense a good explanation do not tell us anything about the (likely, approximate) truth of theories.

In responding to this objection, explanationists often appeal to the descriptive adequacy of their view vis-a-vis scientific inferential practice. This empirical claim is supposed to provide an empirical argument for the reliability of IBE. Two overall strategies for these arguments can be distinguished: *direct* and *indirect* arguments (Douven 2011). Most discussion of these arguments have been framed within the context of the scientific realism debate. Consequently, commentators have on whether these arguments can be formulated in a way that avoids begging the question against sophisticated forms of antirealism. This paper raises a more fundamental problem: the empirical premises needed for these argument to support their conclusion are simply *false*. This problem remains even if we assume the truth of scientific realism.

According to the direct strategy, successful applications of IBE provide evidence for the general reliability of IBE. This argument strategy can be formulated using various frameworks – e.g. as a simple enumerative induction, using Bayesian confirmation theory (Douven 2005) or Kitcher's (2001) “Galilean Strategy”. Regardless, in order for the direct strategy to support explanationism, it is not enough that IBE has *often* led scientists to infer (approximately) true theories. Rather, it must be the case that IBE has led to the truth *more often than not*.

However, as already Duhem (1954) and later Laudan (1981) pointed out, even if theories we currently accept provide very good explanations, these were preceded by numerous *false* theories which also provided very good explanations. This problem is distinct from the pessimistic induction against scientific realism. In order to defend the claim that successful theories tend to be true, most realists have responded by (i) narrowing their conception of “successful theories” to include novel predictive success and (ii) restricting their realist commitments to the “working posits” of the theory (Psillos 1999). Both realist responses raise problems for explanationism. First, move (i) suggests that it is novel predictive success, rather than explanatory quality, which is doing all the truth-tracking epistemic work. Second, the way (ii) is implemented



exacerbates this problem, since it usually the central *explanatory* posits – caloric, phlogiston, etc. – of past theories which are deemed to be “idle wheels” (Chang 2003).

The indirect strategy, in contrast to the direct strategy, does not rely on claims about the success of individual instances of IBE. It only relies on the premise that IBE plays an important role in scientific inquiry. This, together with the assumption that scientific inquiry is generally reliable, is taken to support the reliability of IBE (e.g. Thagard 1988, ch. 8; Lipton 2004, ch. 9). Notice that this argument can side-step the challenges from past unsuccessful applications of IBE if “generally reliable” is interpreted to mean something like “tends to lead to approximately true theories in the long run”. Even so, the direct strategy, in this simple formulation, commits a fallacy of division: even if scientific inquiry *as a whole* is generally reliable, it does not follow that any *individual* inference-pattern used in scientific inquiry is reliable as well. In particular, it fails to rule out that explanatory reasoning plays a different role in scientific inquiry from being a guide to the approximate truth of theories.

In order to support explanationism, the indirect strategy requires a stronger empirical premise, viz. that IBE plays an important role in scientific inquiry *as a means to determining which theories to regard as (likely to be approximately) true*. I propose an alternative view according to which explanatory reasoning first and foremost plays the role of *generating or selecting hypotheses that it would be worthwhile to pursue* (cf. Paavola 2006, McKaughan 2008). To pursue a theory is to investigate whether it is true, e.g. developing it theoretically and testing it empirically. Deciding which theory to pursue is essentially a decision-theoretic problem of how to best spend our time and resources available for scientific research. On this account, even if explanatory virtues do not tell us anything about the truth of *H*, the fact that *H* would be the most satisfying explanation if it were true can still be a good reason to investigate whether it is actually true.

I argue that the pursuitworthiness view provides a more descriptively adequate account of the role of explanatory reasoning in science. I illustrate this claim with two case-studies often taken to support explanationism: Semmelweis' investigations of childbed fever (Lipton 2004) and Le Verrier's discovery of Neptune (Dovuen 2011).

### References:

- Cartwright, Nancy (1983): *How the Laws of Physics Lie*. Oxford: Oxford University Press. Chang, Hasok (2003): “Preservative Realism and Its Discontents: Revisiting Caloric”, *Philosophy of Science* 30:902-912.
- Duhem, Pierre (1954): *The Aim and Structure of Scientific Theory*, Princeton: Princeton University Press
- Dovuen, Igor (2005): “Evidence, Explanation, and the Empirical Status of Scientific Realism”, *Erkenntnis* 63:253-291.
- Dovuen, Igor (2011): “Abduction”, in: Zalta, E. (ed.): *The Stanford Encyclopedia of Philosophy*. Kitcher, Philip (2001): “Real Realism: The Galilean Strategy”, *The Philosophical Review* 110:151-197.
- Laudan, Larry (1981): “A Confutation of Convergent Realism”, *Philosophy of Science* 48:19-49. Lipton, Peter (2004): *Inference to the Best Explanation* (2nd ed.), London: Routledge. McKaughan (2008): “From Ugly Duckling to Swan: C. S. Peirce, Abduction, and the Pursuit of Scientific Theories”, *Transactions of the Charles S. Peirce Society* 44:446-468.

Paavola, Sami (2006): “Hansonian and Harmanian Abduction as Models of Discovery”, *International Studies in the Philosophy of Science*, 20:93-108.

Psillos, Stathis (1999): *Scientific Realism*, London: Routledge.

Thagard, Paul (1988): *Computational Philosophy of Science*, Cambridge, MA: The MIT Press.

---

*Theoretical Fertility McMullin-style*

SAMUEL SCHINDLER

Aarhus University

samuel.schindler@css.au.dk

---

A theory's fertility, fruitfulness, or fecundity has been identified as one of the five most important theoretical virtues besides empirical accuracy, simplicity, unifying power, and external/internal consistency (Kuhn 1977; Okasha 2011). These days theoretical fertility is usually understood in terms of novel success, i.e., in terms of a theory's confirmed predictions of novel phenomena. Novel success has played a central role not only in discussions about theory choice, but also in the realism debate: realists are normally willing to commit to the truth of only those theories which have managed to produce novel success. There is however another form of fertility which, Ernan McMullin has claimed in several works (1968, 1976, 1982, 1984, 1985), provides just as strong an argument for realism, if not even a better one. This form of theoretical fertility, which I will call M-fertility, will be the subject of this paper. In particular, I will defend M-fertility as a "virtue in its own right" against a reductionist challenge put forth by Nolan (1999). At the same time, however, I will question the realist rationale offered by McMullin for M-fertility being a virtue.

McMullin (1968, 391; 1976, 423) characterizes a theory's fertility as the ability to "cope with the unexpected", in particular, "as new evidence becomes available". In another place, McMullin specifies:

... the theory proves to have the *imaginative resources* [...] *to enable anomalies to be overcome and new and powerful extensions to be made*. Here it is the long-term proven ability of the theory or research program to *generate fruitful additions and modifications* that has to be taken into account. (McMullin 1982, 16)

To put it schematically, at first pass, a theory T is M-fertile if T has resources for accommodating evidence E, inconsistent with T, by suggesting modifications of T so that a modified version of T, T\*, can accommodate E. This definition will be made more precise in the following.

Nolan (1999) has questioned that there is a feasible rationale for viewing M-fertility as a virtue "in its own right": "on the face of it", Nolan explains, "it can be a little hard to see why the liability of a theory to require improvements, or to raise new problems, should be considered a good thing" (267). Instead Nolan has argued that, insofar M-fertility is a virtue, it is a virtue only because it can be reduced to novel success. In the paper I will show that Nolan is wrong about this; M-fertility cannot be so reduced.

M-fertility is incompatible with both of the standard notions of novel success, namely temporal and use-novel success. For temporally novel success, evidence E must minimally be unknown at the time at which theory T is devised. E must also be anticipated by T. But that is clearly not the case for what McMullin has in mind. In M-fertility, modified T\* accommodates evidence that could not be accommodated by the original T (an anomaly for T is by definition not anticipated by T). In M-fertility also use-novelty is violated: T is modified *in order to* accommodate E. In other words, E was used in the construction of T\*. Hence, in so far as M-fertility is a virtue, it is a virtue that is independent of the virtue of novel success. Nolan's arguments to the contrary, I will show, fail. But is M-fertility a virtue in the first place?

McMullin offers the following realist rationale for M-fertility: the modifications of a theory in response to anomalies are virtuous so long as those modifications are not ad hoc, which, for McMullin is equivalent to the modifications being motivated by de-idealizations of the theory. Although that is plausible in principle, I will argue that even McMullin's own favoured example does not support this view. Instead, I will argue, it supports the view that scientists are happy to modify their theories to establish empirical adequacy even when those changes cannot be construed as de-idealizations.

Although this conclusion seems to support antirealism, I will point out an important caveat that the antirealist might be hard pressed to accommodate.

## References

- Kuhn, T.S. 1977. Objectivity, Value Judgment, and Theory Choice. In *The Essential Tension*. Chicago: University of Chicago Press.
- McMullin, E. 1968. What do Physical Models Tell us? *Studies in Logic and the Foundations of Mathematics* 52:385-396.
- — —. 1976. The fertility of theory and the unit for appraisal in science. In *Essays in the Memory of Imre Lakatos*, edited by R. S. Cohen. Dordrecht: D. Reidel Publishing Company.
- — —. 1982. Values in science. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association Two: Symposia and Invited Papers* 3-28.
- — —. 1984. A case for Scientific Realism. In *Scientific Realism*, edited by J. Leplin. Berkeley: University of California Press.
- — —. 1985. Galilean idealization. *Studies in History and Philosophy of Science Part A* 16 (3):247-273.
- Nolan, D. 1999. Is fertility virtuous in its own right? *The British Journal for the Philosophy of Science* 50 (2):265-282.
- Okasha, S. 2011. Theory choice and social choice: Kuhn versus Arrow. *Mind* 120 (477):83-115.

---

*Is Interaction Conducive to Scientific Objectivity?*

DUNJA SESELJA

Ruhr-University Bochum

dunja.seselja@rub.de

---

In the recent literature on scientific pluralism interaction among scientists has often been argued for as a necessary condition for scientific objectivity. In the Millian spirit, Helen Longino takes interaction among the proponents of rivaling hypotheses in a given domain essential for the advancement of both the individual inquiries and for the scientific domain itself (Longino (2002a), Longino (2002b), Kellert et al. (2006)). Similarly, Hasok Chang argues for the benefits of interaction among scientists facing a disagreement, which supersede the benefits of passive toleration (Chang (2012)). The main idea underlying these arguments is that the ideal of scientific objectivity includes interactive objectivity, which poses the requirement on scientific claims to be shared, discussed, open to examination and criticism, etc. (see Douglas (2009)).

This conduciveness of interaction for scientific objectivity has been challenged by Kevin Zollman (Zollman (2007), Zollman (2010)). Using a game-theoretic approach Zollman has constructed a formal model which represents processes of learning and decision making that occur in scientific communities. The model is based on an analogy between the learning processes typical for scientific communities and the so-called bandit problems (devised in statistics and used in economics). It concerns the question whether a scientist should at a given point keep on gathering information about different methods of solving a problem in order to see which of the methods is superior, or whether she should rather pick out the method that seems the most promising and stick with that one. In this respect a scientist is similar to a gambler, confronted with different slot machines which payoff at different rates, who has to decide at which point to stop testing the machines and stick with the one that seems to have the best payoff.

In view of this model Zollman concludes that there are two (mutually exclusive) possible ways in which a scientific community can assure an optimal acquisition of reliable knowledge:

1. either by limiting the information flow and restricting interaction in the community;
2. or by endowing the scientists with extreme beliefs (i.e. extremely high epistemic confidence) regarding their respective pursued hypotheses.

In other words, unless scientists have extreme (epistemically unwarranted) beliefs regarding their hypotheses, interaction among them may hinder reliable acquisition of scientific knowledge.

The aim of the present paper is to challenge Zollman's result by showing that his modelling is based on unwarranted assumptions about how scientists evaluate their hypotheses and how they respond to new evidence. My criticism consists of two parts.

In the first part I argue that Zollman's model is based on a conflation of two types of evaluation regarding scientific hypotheses, which Thomas Nickles has termed epistemic appraisal (EA) and heuristic appraisal (HA). EA is a *retrospective* assessment that regards truth-conducive features of justification and decision-making, and stands for the traditional idea of theory confirmation. In contrast, HA is a *prospective* assessment that regards heuristic and pragmatic considerations concerning the fruitfulness of research directives (Nickles, 2006, p. 159) (see also Nickles (1996), Nickles (2009)). Hence, EA addresses the question: "What should we believe/endorse/accept as reliable knowledge claims?", while HA addresses the question: "Which ideas/hypotheses/theories are worthy of pursuit?". As it has been argued by Nickles and others (see e.g. Šešelja et al. (2012)) these two kinds of questions are addressed by means of different types of considerations.

Now, Zollman's model aims at representing scientists' decisions that regard the question which research avenue each of them should pursue. Such decisions are clearly a matter of an HA. Nevertheless, I will show that Zollman describes these decisions as if they were a matter of an EA.

In the second part of my discussion I look at a possible escape route for Zollman's model. A natural question to ask is: can his model be saved from the above objection by simply being re-interpreted in such a way that the decisions of scientists are taken to represent belief updating in terms of HA? For this to be the case, the information exchanged by scientists has to concern considerations relevant for HA as well. The results that scientists obtain via their inquiries (such as their experimental results) can indeed play such a role. Nevertheless, I argue that the way Zollman's model represents the process of belief updating does not correspond to the way scientists update their beliefs in terms of HA. In particular, when faced with counter-evidence, scientists performing HA will update their stance not only in view of this evidence, but also in view of other considerations, such as the presence of positive and negative heuristics (to use a Lakatosian framework, see Lakatos (1978)) regarding the given hypothesis. Thus, even though their epistemic confidence regarding the hypothesis might, for example, drop or remain low (in view of EA), they might still hold it worthy of pursuit (HA). As a result – contrary to Zollman's view – unrestricted information exchange does not that easily lead to a convergence of research inquiries despite the fact that scientists remain epistemically optimal agents (i.e. without them adopting extreme, epistemically sub-optimal beliefs).

A case in point is the very case study which Zollman presents in his paper: the research on the causes and the treatment of peptic ulcer disease (PUD). As it has recently been argued in Šešelja and Straßer (2014), the bacterial hypothesis of PUD was still worthy of pursuit in the 1950s, when it was largely abandoned. While it did face challenges that Zollman takes to be crucial for its abandonment, the authors argue that this research avenue had sufficiently strong positive and negative heuristics. In other words, the bacterial hypothesis was not abandoned by non-biased scientists whose information flow was unrestricted. Rather, the large-scale study used to refute the bacterial hypothesis was based on a method that was previously shown to be unreliable. Hence, either this methodological problem was *too less* communicated among the

sci-entists at the time, or scientists were unjustified in taking the results of the study to be reliable (assuming no other non-epistemic factors played a role in this case).

### References:

- Chang, H.: 2012, *Is Water H<sub>2</sub>O? Evidence, Pluralism and Realism*. Springer.
- Douglas, H. E.: 2009, *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.
- Kellert, S. H., H. E. Longino, and C. K. Waters (eds.): 2006, *Scientific pluralism*. Minneapolis: University of Minnesota Press.
- Lakatos, I.: 1978, *The methodology of scientific research programmes*. Cambridge: Cambridge University Press.
- Longino, H.: 2002a, *The Fate of Knowledge*. Princeton: Princeton University Press.
- Longino, H.: 2002b, 'Reply to Philip Kitcher'. *Philosophy of Science* **69**, 573–577.
- Nickles, T.: 1996, 'Deflationary Methodology and Rationality of Science'. *Philosophica* **58**, 9–50.
- Nickles, T.: 2006, 'Heuristic Appraisal: Context of Discovery or Justification?'. In: J. Schickore and F. Steinle (eds.): *Revisiting Discovery and Justification: Historical and philosophical perspectives on the context distinction*. Netherlands: Springer, pp. 159–182.
- Nickles, T.: 2009, 'Life at the Frontier: The relevance of Heuristic Appraisal to Policy'. *Axiomathes* **19**, 441–464.
- Šešelja, D., L. Kosolovsky, and C. Straßer: 2012, 'Rationality of Scientific Reasoning in the Context of Pursuit: drawing appropriate distinctions'. *Philosophica* **86**, 51–82.
- Šešelja, D. and C. Straßer: 2014, 'Heuristic Reevaluation of the Bacterial Hypothesis of Peptic Ulcer Disease in the 1950s'. *Acta Biotheoretica* **62**, 429–454.
- Zollman, K. J. S.: 2007, 'The communication structure of epistemic communities'. *Philosophy of Science* **74**(5), 574–587.
- Zollman, K. J. S.: 2010, 'The epistemic benefit of transient diversity'. *Erkenntnis* **72**(1), 17–35.

---

*The Equivalence Principle and Dynamical Explanations*

ADÁN SUS

Universidad de Valladolid

adansus@gmail.com

---

As it is well known, Einstein introduction of the Equivalence Principle (EP) in 1907 is standardly considered the beginning of the path that led him to the formulation of the theory of General Relativity. Even if this episode has been largely studied, the interpretation of EP, as well as its role in the conception of the theory, is still an issue open to discussion.

In this paper I have two main objectives, both related to the physical interpretation of the Equivalence Principle. The first one aims directly at contributing to the discussion about the right formulation of the Equivalence Principle in the context of General Relativity. This question has a considerably long history and, consequently, different versions of the principle have been proposed as candidates to encode the one actually implemented in General Relativity. In relation to this, I argue that most of the proposals focus in capturing in which particular manner the matter fields couple to the metric, and they leave out something that I consider essential for the physical interpretation of EP, namely, that the structure that determines inertial motion cannot be itself determined independently from matter fields. I examine the prospect of incorporating this feature in the formulation of the principle and test whether this helps in the clarification of the relation of EP to other principles that were claimed to be essential in the search for the equations of the theory (like energy-momentum conservation or General Covariance). I also defend that this strategy provides a criterion that separates General Relativity from other relativistic theories of gravitation.

My second aim is to translate the insights from the previous discussion to a debate that belongs to the interpretation of Special Relativity. In recent times, the kinematical character of the explanation of certain phenomena, provided in Special Relativity, has been questioned. In its place, Harvey Brown has defended that a particular version of a Lorentzian dynamical explanation of such phenomena might be operating in Special Relativity. Nevertheless, we can say that most commentators have not been convinced by Brown's strategy; instead, more sophisticated accounts of the so-called kinematical, or sometimes mathematical, character of the explanation have been advanced. There are two different issues implied in these discussions: one relates to the question about the status of different models of scientific explanation, the other concerns the physical interpretation of Special Relativity. I contend that both dimensions might benefit from making converge this discussion with the previous one about the interpretation of EP in General Relativity. At the



bottom of this there is the fact that EP, and other principles, have a two-sided character: they can be seen as dynamical principles and as symmetry ones. From the dynamical point of view, they set restrictions on the shape of the physical interactions and, at the same time, as symmetry principles, they encode the geometrical features of the spacetime structures of the theory in question. Taking this into account, and providing a particular way of formulating the explanation of special relativistic phenomena in the theory that makes use of a restricted version of EP, I argue that the different positions in the debate about the character of explanations in SR can be reconciled. Furthermore, I extract some morals for the general problem of providing a model of scientific explanation.

---

*Naturalising Recombination*

ALASTAIR WILSON

University of Birmingham

a.j.wilson@bham.ac.uk

---

Various realist theories of the nature of modality appeal to some version of a principle of recombination. The informal 'patchwork principle' of Lewis' *On the Plurality of Worlds* and the more rigorous treatment in Armstrong's *A Combinatorial Theory of Possibility* are familiar examples, but recombination principles feature in most views according to which possible worlds are real structured entities. However, surprisingly little attention is usually paid to the epistemic status of these principles. They are paradigm mysterious examples of the putative synthetic *a priori* - highly substantive truths about the nature of modal reality, our way to knowledge of which we are somehow supposed to be able to reason. Now that transcendental idealism is out of fashion, and conventionalism has had its day, the most plausible treatment of the epistemic status of putative synthetic *a priori* truths is the Quinean one: such truths are justified through their indispensable contributions to the formulation of our best scientific theories. Nevertheless, the aura of mystery remains: the justification for such truths seems to be different in kind and much less direct than the justifications we have for truths about goings-on in the actual world.

The best way to render putative synthetic *a priori* truths unmysterious is to naturalize them. This means finding for them a home within the scientific worldview, rather than merely showing that they are necessary to underpin that worldview. In this paper, I offer a naturalistic treatment of recombination in the context of the Everettian, or many-worlds, approach to quantum mechanics. According to my proposal, the unitary evolution described by the Schrödinger equation is best understood as more akin to a recombination principle than to a law of any individual world. The Schrödinger equation grounds the truth of a macroscopic recombination principle, and thereby ensures that we have what David Lewis has called a *plenitude of possibilities*.

The Schrödinger equation has always had a puzzling status; it does not fit neatly into familiar frameworks for understanding laws of nature. Taken literally, it seems to describe an entity - the wavefunction - which evolves in an unfamiliar infinitely-high-dimensional Hilbert space. Even if the wavefunction is interpreted in some more palatable way - for example, Bohmians often think of it as something like a law - then it still encodes a remarkable amount of complexity: the wave-function of the universe, even for Bohmians, encodes enough information to reconstruct the entire space of physical possibilities (at least, those with the same initial conditions). If the Schrödinger equation is a law of nature, it is a law unlike any other.

Naturalizing recombination involves treating Everett worlds as distinct possible worlds: many-worlds quantum theory is then a theory of modality rather than merely a theory of the actual world. Unlike other versions of modal realism, Everettian modal realists have available a fully naturalistic story about the extent and contents of modal reality. This opens the way to a previously unsuspected possibility in metametaphysics: we can have empirical confirmation and disconfirmation for theories of the metaphysics of modality.

The reading of the Schrödinger equation as naturalizing recombination can in fact be motivated from within David Lewis' own view of the correct methodology for theory-building in science, as encoded in his 'best-systems' approach to laws of nature. If (as committed modal realists should be ready to) we expand the scope of this methodology to cover theories about the nature and contents of modal reality, then general facts about the space of worlds that strike good combinations of simplicity and strength will be strong candidates to be included in the laws of nature. This conclusion is not limited to Everettians: modal realists in general can think of their preferred principle of recombination as a *law of the plurality*.

Lewis appealed to the principle of recombination in order to capture what he called the requirement of *plenitude*: a modal realist should ensure that there are 'worlds enough, and no gaps in logical space'. The Schrödinger equation plays a very similar role in Everettian quantum mechanics: it can be thought of as a conservation principle for probability. It ensures that there is an Everett world for every outcome that - before the experiment - had non-zero probability, and it thereby ensures something very like Lewis' plenitude of possibilities. Still, the roles are not identical: whereas the Lewisian patchwork principle applies to fundamental entities, the plenitude of possibilities that the Schrödinger equation gives rise to is at the level of derivative entities.

The best modern versions of Everettian quantum mechanics lean heavily on aspects of the physics of decoherence. According to decoherence-based versions of the interpretation - associated in particular with Simon Saunders, David Wallace and Hilary Greaves - macroscopic worlds are derivative entities, somewhat indeterminate in their nature and number. Accordingly, Everettian modal realists ought to regard the plenitude of qualitative possibilities at the macroscopic level as a corollary of a more basic principle: fundamental reality evolves unitarily. Here, then, we have a putative truth of physics which is clearly more fundamental than a putative truth of metaphysics. The rationalist picture of metaphysics underlying physics, promoted by metaphysicians like George Bealer and E.J. Lowe, is fully inverted.

---

*On the Notion of A-Spatiotemporal Beables in Quantum Gravity, or: Can we Dispense with Space and Time as Fundamental Categories?*

ANTONIO VASSALLO

University of Lausanne

antonio.vassallo@unil.ch

---

One of the most remarkable contentions in the research for a theory of quantum gravity (QG) is that spacetime might not be fundamental, but “emergent” from an ontological ground floor made up of a-spatiotemporal elements of reality: causal set theory and non-commutative geometries approach are two major examples of theoretical frameworks advocating this view.

However, this picture of spacetime as a derivative entity can be challenged by noticing that the alleged building blocks of spacetime are usually represented as quantum superpositions of abstract mathematical objects, e.g. spin network states in loop QG. Given this fact, it is difficult to provide these elements with a sharp metaphysical characterization, and this lack of ontological clarity makes conceptually suspicious any attempt at establishing a connection between the a-spatiotemporal regime and the macroscopic world. With this respect, the claim that spacetime emerges from a probabilistic cloud of a-spatiotemporal elements is problematic for at least two reasons. Firstly, this characterization fails to deliver a clear picture of the appearance of a concrete and definite structure - spacetime - from quantum superpositions. Even invoking some sort of underlying decoherence mechanism is not enough, since decoherence is per se a source of interpretational controversy already in the standard quantum-mechanical setting. Secondly, the appeal to the concept of emergence has to face the fact that whatever sufficiently worked out account of emergence (e.g. in terms of causality, supervenience, or ontological grounding) heavily relies on pre-existing spatiotemporal notions (this line of argument is adopted, e.g. in Lam and Esfeld, 2013, especially sections 3-4).

Such a skeptical attitude is usually accused of being unreasonably attached to intuitions, and seeking to force a “folk” picture in terms of outmoded Aristotelian categories - like space and time - upon modern physics. Proponents of this objection claim that, if we are interested in a metaphysics motivated and informed by science in general, and physics in particular, we must take seriously what modern physics tells us. Hence, if it is among the contentions of a decently worked out theory of QG that space and time are not among the fundamental constituents of reality, then the “naturalized” metaphysician should hasten to drop these ontological categories (see Ladyman and Ross, 2007, section 1.2, for an articulated defense of this point in a more general context).

The aim of the paper is to enter the above sketched debate by considering the question whether a naturalized metaphysics that acknowledges the primacy of physics over the special sciences could dispense with space and time as fundamental categories, and by what means it might do so. As regards the method of enquiry, the paper will endorse - contra Ladyman and Ross' position - the view that a good naturalized metaphysical theorizing is not parasitic upon the physical practice but, rather, it goes hand in hand with it, in the spirit of natural philosophy.

The analysis will start by considering the thesis, defended in Huggett and Wüthrich (2013, section 3), that a sufficiently developed theory of QG that dispenses with spacetime but still offers complete and coherent formal derivations of empirically relevant aspects - such as spatiotemporal macroscopic degrees of freedom - has to be taken as physically genuine simpliciter. This thesis will be discussed under the light of the difference, put forward in Maudlin (2007), between *informational* and *ontological* completeness. By considering some concrete cases taken from the current physical literature on QG, it will be argued that a spacetimeless theory that recovers mere macroscopic empirical predictions (e.g. cross sections of scattering processes) by no means can be automatically taken seriously from an ontological point of view: simply put, informational completeness does not imply ontological completeness. The second step will consist in introducing and spelling out in detail the notion of local beable, especially highlighting the role of beables in bridging ontological and informational (empirical) aspects of a physical theory. The third, and crucial, step will be to translate the concept of beable to QG, and consider what kind of modifications such a concept should undergo in order to fit in this context. Finally, a tentative proposal will be put forward concerning the minimal metaphysical requirements that beables for a theory of QG should meet in order to be considered genuine elements of reality as opposed to mere abstract elements of the formalism.

In conclusion, the paper will address the following moral: when coming to QG, the naturalized metaphysician with physicalist inclinations can trim a lot off the categories of space and time, but cannot give them up entirely.

### References

- Huggett, N. and C. Wüthrich (2013). Emergent spacetime and empirical (in)coherence. *Studies in History and Philosophy of Modern Physics* 44, 276–285.
- Ladyman, J. and D. Ross (2007). *Every thing must go: metaphysics naturalized*. Oxford University Press.
- Lam, V. and M. Esfeld (2013). A dilemma for the emergence of spacetime in canonical quantum gravity. *Studies in History and Philosophy of Modern Physics* 44, 286–293.
- Maudlin, T. (2007). Completeness, supervenience and ontology. *Journal of Physics A* 40, 3151–3171.

---

*Parts, Wholes and Potentials*

F.A. MULLER

Erasmus University Rotterdam

f.a.muller@uu.nl

KERRY MCKENZIE

University of Western Ontario

kmckenzie@ucsd.edu

---

P. van Inwagen (1987, p. 23) raised the question what condition physical objects have to meet in order to compose another physical object. Three answers to this so-called *Special Composition Question* (whenceforth: the Question) have been considered and debated: *mereological Nihilism*, according to which there are no composite objects, only elementary particles; *mereological Universalism*, according to which any number of arbitrary physical objects compose another physical object (Lewis 1991); and Van Inwagen's moderate answer (1990: 82), according to which physical objects compose another one iff their activity constitutes a life. All three versions scandalise our intuitive judgements: Nihilism and Van Inwagen deny that atoms, molecules, rocks, trucks, planets and galaxies qualify as *bona fide* physical objects, whereas Universalism affirms that, say, the cilia of Hypathia of Alexandria, the fingers of Madame Curie and the Erasmus suspension bridge in Rotterdam *do* by contrast form such an object. Whilst we agree that philosophical clarity and coherence trump common sense any time, we prefer a view that is clear and coherent but that trumps common sense as painless as possible. We propose such a view here (*pace* Healey 2013). Slightly more specifically, we would like to propose a fourth answer to the Question, one that both vindicates more of our intuitive judgments than the extant answers, and that is grounded in our physical knowledge of reality.

Consider an arbitrary but finite number of physical objects. They may or may not be interacting with one another. Whenever they interact, they may or may not be in a *bound state*, which by definition is a state in which the objects have a total energy that is negative ( $E < 0$ ). In that case, the potential energy of the composing objects (which is always  $< 0$ ) is larger in absolute value than their kinetic energy (which is always  $\geq 0$ ); their total energy is the sum of their kinetic and potential energy. (The categorical attribution of a quantitative property of energy to objects presuppose a background of classical physics. True enough. When we move to quantum-physical theories, notably quantum mechanics, an exactly similar story can be told in terms of expectation-values of the Hamiltonian, which is the energy operator.) Our proposal now reads: *physical objects form a composite object iff these physical objects interact and are in a common bound state*, where 'common bound state' means that the composing objects are in the potential well that results of their mutual physical interaction. An object *a* then is a *part* of object *b* iff *a* is among the objects that compose *b*. We call this the *Bound-State Proposal*. In order to find out whether objects are in a common bound state — which is an epistemic, not a metaphysical problem — one must find a physical theory in the currently

accepted body of scientific knowledge that describes these objects as interacting and being in a bound state as a result of it. Note that the mere fact of interaction is a necessary but not a sufficient condition for composition; what needs to be added is that the resulting state they are in is bound.

**References**

- Healey, R., 'Physical Composition', *Studies in History and Philosophy of Modern Physics* **44** (2013) 48–62.  
Inwagen, P. van (1987). 'When are objects parts?', *Philosophical Perspectives* **1**, 21–47.  
Inwagen, P. van (1990). *Material Beings*, Ithaca: Cornell University Press.  
Lewis, D.K. (1991). *Parts of Classes*, Oxford: Basil Blackwell.

---

*Building Integrated Explanatory Models of Complex Biological Phenomena: From Mill's Methods to a Causal Mosaic*

ALAN LOVE

University of Minnesota

aclove@umn.edu

---

After his discussion of the four methods of observation and experimental inquiry (Agreement, Difference, Residues, and Concomitant Variations), John Stuart Mill raised a concern about the applicability of these methods to complex causal relationships.

It has been necessary to suppose, ...for the sake of simplification, that this analytical operation is encumbered by no other difficulties than what are essentially inherent in its nature; and to represent to ourselves, therefore, every effect, on the one hand as connected exclusively with a single cause, and on the other hand as incapable of being mixed and confounded with another coexistent effect (Mill 2006 [1973], 434).

The simplifying suppositions are embedded *idealizations*: reasoning strategies that purposefully depart from features known to be present in nature (Weisberg 2007). “One cause, one effect” and “no mixing of effects” do not correspond to most causal relationships in nature, as Mill was fully aware, and are often encountered in biological phenomena: “This difficulty is most of all conspicuous in the case of physiological phenomena; it being seldom possible to separate the different agencies which collectively compose an organized body” (Mill 2006 [1973], 456). His labels for the respective situations were the *plurality of causes* and *intermixture of effects*.

Building explanatory models of complex phenomena remains a challenge, in part because similar idealizations are embedded in our causal models, such as difference making and production accounts (Hall 2004). The problem takes on a special significance when the aim is to offer an integrated account of causes across disciplinary approaches, such as in the attempt to combine physical and genetic explanations of embryogenesis. Genetic explanations appeal to changes in the expression of genes and interactions among their RNA and protein products to causally explain how cells differentiate or embryos grow and develop. Physical explanations appeal to mechanical forces resulting from the geometrical arrangements of soft condensed materials within the embryo to causally explain the same effects. There is no controversy about whether both are involved: “both the physics and biochemical signaling pathways of the embryo contribute



to the form of the organism” (Von Dassow et al. 2010, 1). The question is how to combine them in order to understand their joint contribution to the effect of organismal form.

In this paper I argue that an underappreciated dimension of building integrated explanatory models where the plurality of causes and intermixture of effects are present is *time*. I begin with Mill idealizations to clarify their significance and relevance when attempting to combine physics and genetics in explanatory models. The plurality of causes obtains, such as scales being formed from genetic specification or physical cracking (Milinkovitch et al. 2013), and the intermixture of effects obtains, such as genetic and physical factors contributing to the developmental origin of complex structures (Savin et al. 2011). Second, I argue that another—unstated—idealization in Mill’s methods is the absence of time apart from an ordinal relation of causes preceding their effects. This idealization is shared by contemporary difference making accounts of causation but is not present in production accounts (e.g., mechanisms). Third, I pose a dilemma: even though production accounts are better able to incorporate time than difference making accounts, the framework of actual difference making corresponds to the experimental reasoning involved in uncovering the operation of both physical and genetic causes (Waters 2007).

I propose a resolution to the dilemma by showing how explicit periodizations of ontogeny serve as a framework to combine genetic and physical causes understood as both difference makers and in terms of mechanistic production. The resources for this strategy are already present in the practices of developmental biology through its standardized use of normal stages (Hopwood 2005). These permit the linkage of difference makers into chains of productive continuity; relations of systematic dependence between genetic and physical factors can be keyed to different aspects of a sequence during which the mediation of cause and effect occurs. But this temporal structure is not anchored in a particular mechanistic description; rather, it is localized to the model organism under scrutiny. Thus, standardized periodizations coordinate genetic and physical mechanisms composed of genetic and physical difference makers. The result is a procedure for building integrated explanatory models of complex biological phenomena that does not privilege a particular conception of causality, but instead derives from a causal mosaic that organizes different accounts of causality to serve the ends of inference and explanation (Illari and Russo 2014). In closing, I discuss a counterintuitive result of this procedure: it narrows the scope of generalizations ascertained for the models of causation in isolation. This suggests an explanatory tradeoff between models that yield causal generalizations of wide scope and models that yield an integration of different types of causes to more comprehensively explain complex phenomena.

## References

- Hall, N. 2004. Two concepts of causation. In *Causation and Counterfactuals*, eds. L. Paul, N. Hall, and J. Collins, 225–276. Cambridge, MA: MIT Press.
- Hopwood, N. 2005. Visual standards and disciplinary change: Normal plates, tables and stages in embryology. *History of Science* 43:239–303.

- Illari, P., and F. Russo. 2014. *Causality: Philosophical Theory Meets Scientific Practice*. New York: Oxford University Press.
- Milinkovitch, M.C., L. Manukyan, A. Debry, N. Di-Poï, S. Martin, D. Singh, D. Lambert, and M. Zwicker. 2013. Crocodile head scales are not developmental units but emerge from physical cracking. *Science* 339:78–81.
- Mill, J.S. 2006 [1973]. *A System of Logic Ratiocinative and Inductive*. Collected Works of John Stuart Mill, Volume 7 (Books I-III). Indianapolis: Liberty Fund.
- Savin, T., N. A. Kurpios, A. E. Shyer, P. Florescu, H. Liang, L. Mahadevan, and C. Tabin. 2011. On the growth and form of the gut. *Nature* 476:57–62.
- Von Dassow, M., J. Strother, and L. A. Davidson. 2010. Surprisingly simple mechanical behavior of a complex embryonic tissue. *PLoS ONE* 5:e15359.
- Waters, C.K. 2007. Causes that make a difference. *Journal of Philosophy* 104:551–579.
- Weisberg, M. 2007. Three kinds of idealization. *Journal of Philosophy* 104:639–659.

---

*Reality as a Relational Property: The History of G-Protein coupled Receptors*

ANN-SOPHIE BARWICH

KLI Vienna

ann-sophie.barwich@kli.ac.at

KARIM BSCHIR

ETH Zurich

bschir@phil.gess.ethz.ch

---

This paper tells the story of one of the most exciting objects of investigation in contemporary biochemistry and molecular biology: G-protein coupled receptors (GPCRs). GPCRs constitute the largest family of cell membrane receptors in the mammalian genome; they are involved in a wide range of fundamental signalling processes such as immune responses, vision and olfaction. Despite their centrality for biological and pharmacological studies today, their existence was doubted deep into the 1970s. By analysing the ‘historical career’ of GPCRs over the past 40 years, and by looking at how they turned from a hypothetical entity into a real one, we demonstrate that the realism question requires a philosophical perspective in which scientific objects are analysed as active elements within a specific research context.

Our interest the recent history of GPCRs is twofold. First, GPCRs constitute one of the most important objects of investigation in contemporary biology. Second, their history has not been part of philosophical or historical science studies. Despite their centrality in contemporary proteomics, the study of GPCRs is a very recent endeavour. Since their existence was controversial deep into the 1970s, their historiography as a scientific object only spans the past 40 years. Within the last thirty years, however, six Nobel Prizes were awarded to studies surrounding signalling processes by G-proteins and GPCRs. It is thus philosophically intriguing to ask: What exactly happened throughout these past decades that turned the initially hypothetical receptors, or conceptual placeholders, into one of the most promising scientific objects of modern biology?

The idea that scientific objects have a history is not new. Works by Rheinberger (1997), Daston (2000) and Arabatzis (2005) have emphasised the importance of history for questions surrounding the emergence and disappearance of scientific objects. Building on this approach, we propose a model of three tentative stages through which hypothetical entities progress in their development when they turn into real scientific objects. Scientific objects, we argue, pass through these stages that mark their relative degrees of reality. Relative degree of reality does not imply that a certain amount of substantial knowledge about a scientific object has been achieved. Rather, it means the extent to which a scientific object is considered a more or less stable part of a currently employed scientific ontology.

In the first stage of our model, hypothetical entities turn into potential candidates for realism through their progressive entrenchment in an experimental context. Such entrenchment is characterised by several factors, encompassing an explanatory role of the hypothetical entity within a wider theoretical context, the

material tractability of effects associated with the hypothetical entity through various methods, and the stability of produced links between explanatory function and tractable effects.

Second, thereby formed epistemic objects open up further avenues of research. Their productivity as part of an experimental system might be described by their capacity to facilitate new empirical findings and unify previously separate phenomena under a joint concept, or allowing for novel research outlooks.

In the third stage, scientific objects become an integral part of a current scientific ontology to the extent that these objects acquire the role of a standard, in the sense of a touchstone, by which other effects, hypothetical entities or mechanisms are evaluated in their capacity to represent potential candidates for realism.

Our analysis of the history of GPCRs has implications for the debate on scientific realism. We draw attention to a fundamental aspect constituting the activity of scientific objects: their capacity to act as touchstones of the reality of other things. We show that the selection of criteria, whereby an object (such as GPCRs) is assigned varying degrees of reality throughout a scientific discourse, cannot be made independently of the question of how this object becomes a standard by which the reality of 'neighbouring elements' of enquiry (other entities, mechanisms, processes) are evaluated.

We take the capacity of a scientific object to represent this standard as the criterion for its reality. This understanding of realism thus does not limit itself to define reality *qua* manipulability (Hacking 1983). We argue that manipulability is an insufficient criterion to account for the historical formation of scientific objects. It does not explain how previously disparate phenomena crystallise into a new scientific object, instead of becoming assimilated to already known objects or to represent data coincidences. Only through its historical development into an integrative focal point within a scientific ontology can we speak of a scientific object as a unified real entity. Ultimately, the reality of a scientific object is a relational property acquired in a historical context of successive methods. The comparatively young history of GPCRs shows how they contributed to the integration of a wide variety of biological phenomena and at the same time are considered the 'holy grail' of contemporary biochemistry (Snogerup Linse 2012).

## References

- Arabatzis, Theodore. 2005. "Representing Electrons: A Biographical Approach to Theoretical Entities." Chicago: University of Chicago Press.
- Daston, Lorraine (Ed.). 2000. "Biographies of Scientific Objects." Chicago: Chicago University Press.
- Hacking, Ian. 1983. "Representing and Intervening: Introductory Topics in the Philosophy of Natural Science." Cambridge: Cambridge University Press.
- Rheinberger, Hans-Jörg. 1997. "Towards a History of Epistemic Things. Synthesizing Proteins in the Test Tube." Stanford: Stanford University Press.
- Snogerup Linse, Sara. 2012. "Studies of G-Protein-Coupled Receptors." The Royal Swedish Academy of Sciences.

---

*Causality in Pharmacology: Conceptual Analysis for a Changing Landscape*

BARBARA OSIMANI

University of Camerino

barbaraosimani@gmail.com

---

Other than most basic sciences, such as chemistry or biology, pharmacology works across levels of reality: whereas the first step in the causal chain leading to the therapeutic outcome takes place at the biochemical level, the end-effect is a clinically observable result. On the other hand, pharmacology also differs from other kinds of technologies mainly for the fact that it makes nature work at its place: the drug molecule just gives an input to the causal process, which then goes on, almost uncontrolled in the organism system. Another important aspect concerns the fact that causation in pharmacology works on a population of receptors in an integrated complex system. This not only implicates that the drug effects essentially depend on dosage, but also that they depend on the features of the population of affine receptors present in the individual consumer, as well as on how they are interconnected. Since knowledge of the “natural laws” at play in the organism and of the “initial conditions” holding in each context, is far from being complete, causal inference and prediction in pharmacology faces specific challenges which translate into 1) a strong uncertainty associated with the intended outcome and exact causal path leading to it (“process tracing” of causal mechanisms); 2) strong uncertainty as to the relevant background conditions and related subpopulations (external validity/heterogeneity) 3) strong unpredictability of side-effects (in analogy to what economists call “externalities”). Because of these multiple sources of uncertainty, until recently, drug approval has mainly relied on a black box methodology, grounded on hypothesis rejection (frequentist statistics) with related implications (categorical causal assessment where hypotheses are either rejected or not, with no degrees in-between; one indicator of causality mainly based on the difference between treated and untreated group in the sample population; focus on internal validity at the expense of external validity, with implications for extrapolation and prediction; finally, as a consequence of the focus on internal validity, tendency to abstract from heterogeneity, which is taken into account only to downgrade the study quality instead of considering it on its own right). The methodological landscape however is rapidly changing both through the gradual diffusion of Bayesian methods of statistical inference, not only for “pattern recognition” in signal detection for pharmacosurveillance, where they are already well established (see the work of the Uppsala Monitoring Centre) but also in clinical trials (Price, 2014), as well as by the development of so called “systems pharmacology”. These two paradigms respond on one side to the acknowledgement of the strong uncertainty intrinsic in pharmacology, and on the other to its multilevel scope. In fact, Bayesian methods are gaining ground because of their ability to optimise the use of available evidence by incorporating historical

(heterogeneous) knowledge in the prior, allowing diverse types of evidence to be integrated in the probability function, and by providing a probabilistic measure of the hypothesis under investigation, hence allowing decisions under uncertainty. On the other side, the various research projects devoted to knowledge discovery through the combination of multiple datasets across organism levels can be seen as a sort of “systematic process tracing” in that the drug molecule is matched against all possible proteins/receptors in the proteome database; these are then connected with each biological pathway where they are known to be involved, finally these are linked to all related organ subsystems and to phenotypic effects.

However whereas in the standard approach causal inference is based on a straightforward (unsophisticated) counterfactual account of causation, these new approaches require a renewed epistemological foundation with regard to how causality is conceptualized. The present paper aims to provide such an account by taking into account the specific features emerging in pharmacology with a particular focus on safety issues. In particular the following questions will be addressed: In what sense knowledge discovery techniques can be said to provide causal knowledge? How do they differ from epidemiological or experimental evidence?

These question will be addressed by taking into account the techniques developed for predicting side effects of drugs based on their biochemical features and on the integrated information of different databases. These methods promise to be highly informative on potential effects which may go undetected with traditional methods. In particular I will analyse a validation study which uses knowledge on side effects already available for the drug Trocetrapiib gained through Randomized Controlled Trials (Xie et al. 2009). The kind of method validated in the study explicitly connects a priori theoretical knowledge on biochemical features of drug binding sites and receptors with proteomic, genetic and clinical as well as epidemiologic databases. It will turn out, that, as it stands, although grounded on systematic search, the kind of causal knowledge derived from such methods is very fragile, and the reason for this is that it still relies on “universal” causal chains rather than on “context- dependent” causal webs, thereby abstracting again from heterogeneity and background conditions. Hence, whereas such methods may help reduce uncertainty concerning process tracing of mechanisms (point 1), and may also be helpful in early detection of side effects (point 3); still they are still poor advisers with respect to the problem of external validity/heterogeneity (point 3).

---

*Charles Darwin and Sir John F. W. Herschel: Nineteenth-Century Science and its Methodology*

CHARLES PENCE

Louisiana State University

charles@charlespence.net

---

James Lennox has argued that if it is indeed possible to say that Darwin was an innovator in his field, “it is as a philosopher and methodologist.” It would be clearly wrong, however, to argue that Darwin did not work within the context of nineteenth-century philosophy of science. Indeed, studies of Darwin’s work and development have provided us with a bewildering variety of claims connecting Darwin to various philosophical influences, including to Herschel, Whewell, Lyell, German Romanticism, Comte, and others. I argue here that, whatever is to be made of the other connections, Herschel’s influence on Darwin is undeniable.

The form of this influence, however, is often misunderstood. While Jon Hodge has worked out a careful interpretation of both Darwin and Herschel over a series of some half-dozen articles, this interpretation misreads Herschel’s use of the *vera causa* principle, as well as his discussion of the role of hypotheses in scientific theory construction. Hodge’s reading overstates the role of the *vera causa* principle (declaring something a *vera causa* is only a minimal criterion that a hypothesis must meet in order to be suitable for further scientific investigation), and underestimates the extent to which Herschel will countenance the arbitrary proposal of hypotheses when useful for furthering scientific inquiry (Herschel had no qualms with proposing otherwise unsupported hypotheses, as long as they are verified carefully once proposed).

With a minor change to Hodge’s view, however, we can see clearly that Darwin learned from Herschel the way in which one should frame a scientific argument – first by proposing a speculative hypothesis, grounded on an extensive analogical basis, then by demonstrating the adequacy of that hypothesis to produce the desired effect, and lastly its ability to account for a wide variety of phenomena that it was not originally proposed to explain. Darwin read Herschel’s *Preliminary Discourse* in 1838 – the pivotal time at which he was constructing the theory of natural selection for the first time. The Herschellian form of Darwin’s argument is found even in Darwin’s early work in the *Essay* and *Sketch* which precede the writing of the *Origin*. And we can explain why it is that Darwin would have thought that his work on embryology was one of the most important parts of the *Origin* – it follows Herschel’s dictum that we should attempt to confirm theories by checking them against instances which would, at first blush, seem to provide the largest challenge.

In addition to grounding this new reading of Darwin’s relationship to Herschel in a close reading of both Darwin’s own work and Herschel’s *Preliminary Discourse*, I will consider a new archival source – Herschel’s own marginalia to Darwin’s *Origin*, drawn from the archives at the University of Texas. While Herschel’s

critique of Darwin's theory as "the law of higgeldy-piggeldy" is well known to historians and philosophers of biology, these marginalia give us the ability to understand Herschel's more sophisticated criticisms of Darwin's work – focusing, as many at the time did, on the lack of a plausible explanation for the existence of the variations that Darwin needed to make natural selection work.

In short, we can see a clear connection between Hodge's prescriptions for scientific method and the way in which Darwin constructed his argument in the *Origin*.

While it remains difficult to elucidate the full range of connections between Darwin and the philosophy of science of his day, Herschel's influence is irrefutable. The reading of that influence that I offer here goes farthest toward explaining why Darwin wrote the *Origin* in the way that he did, as well as why Herschel's criticism of his theory stung Darwin so deeply.



## Symposia &amp; Contributed Papers II

**Non-Causal Aspects of Scientific Explanation****Symposium**

Organizer: Alexander Reutlinger &amp; Mathias Frisch

Chair: Adán Sus

Room 5F, Thursday 09:30 – 11:30

*On the Supposed Incompatibility of Causal and Non-Causal Explanations*

ALISA BOKULICH

Harvard University, Boston University

abokulic@bu.edu

*A Counterfactual Account of Non-Causal and Causal Explanations*

MATHIAS FRISCH

University of Maryland

mfrisch@umd.edu

ALEXANDER REUTLINGER

LMU Munich

alexander.reutlinger@lrz.uni-muenchen.de

*Varieties of Structural Explanations and the Notions of Explanatory Pluralism*

PHILIPPE HUNEMAN

CNRS (IHPST)

philippe.huneman@gmail.com

*Explanatory Abstraction in a Counterfactual Framework*

IDA L. S. JANSSON

Nanyang Technological University

ilinasjansson@gmail.com

JUHA SAATSI

University of Leeds

j.t.saatsi@leeds.ac.uk

**General Discription**

The question “what is a scientific explanation?” has taken center stage in modern philosophy of science, from its beginnings in the early 20th century until the present day. Both philosophers and scientists typically regard explanation as a key epistemic goal in the natural and social sciences. Scientific explanations are often portrayed as answers to why-questions of the form “why did phenomenon P occur?”. Answering these

explanation-seeking questions is primarily a scientific task. The philosophers' task is to provide a general and informative characterization of what a scientific explanation *is*.

The most widely accepted account of scientific explanation is the causal account. According to causal accounts of explanation, the sciences explain by identifying the causes of (or the causal mechanisms for) the phenomenon to be explained (see Cartwright 1983, 1989; Salmon 1984, 1998; Lewis 1986; Machamer, Darden and Craver 2000; Woodward 2003; Craver 2007; Strevens 2008). Two reasons clearly speak in favor of the causal account: first, many paradigmatic explanations in science are causal explanations; second, the causal account successfully meets desiderata that previously proposed accounts of scientific explanation, such as the covering-law account (Hempel 1965) and the unificationist account (Friedman 1974, Kitcher 1989), failed to satisfy (Salmon 1989: 46-51),

However, are *all* scientific explanations *causal* explanations? The answer to this question seems to be negative, because scientists give non-causal answers to *why*- questions. Since the mid 2000s, a growing number of philosophers of science has argued that the explanatory practices in the sciences are richer than the causal model of explanation makes us believe, pointing to several types of explanations in the sciences whose explanatory power does not derive from identifying causes and mechanisms. Consider two examples of non-causal explanations from the recent literature:

- **Example 1: mathematical explanations.** A broad class of non-causal explanations is grouped under the concept of 'mathematical' explanations. Mathematical explanations of empirical facts proceed by drawing on 'purely' mathematical facts (Baker 2009, Lange 2013a). For instance, the fact that it is impossible to cross all of the famous bridges of Königsberg without crossing one bridge twice is explained by the fact that the arrangement of bridges is isomorphic to a non-Eulerian graph (Pincock 2012, Lange 2013a). If referring to genuinely mathematical facts is explanatory, then there are explanations that do not work by identifying causes.
- **Example 2: renormalization group explanations.** So-called renormalization group explanations constitute another kind of non-causal explanation. Microscopically different physical systems (such as various gases and metals) display the same macro-behavior when undergoing phase-transitions (e.g. from a liquid to a vaporous phase). Renormalization group explanations do not explain this surprising phenomenon by identifying causes and mechanisms. Instead the explanations draw on limit theorems (e.g. the thermodynamic limit), mathematically sophisticated coarse-graining procedures (renormalization group transformations), and the determination of fixed points. None of these explanatory assumptions identifies causes (Batterman 2002, Morrison 2012, Reutlinger 2014, Hüttemann et al. forthcoming). Additional kinds of non-causal explanations to include 'purely' statistical explanations (Lipton 2004, Lange 2013b), geometric explanations (Nerlich 1979, Balashov and Janssen 2003), topological explanations (Huneman 2010, Lange 2013a), and explanations that are based on symmetry principles (Van Fraassen 1989; Lange 2011) and inter-theoretic relations (Batterman 2002, Weatherall 2011).

The philosophical lesson from the existence of non-causal explanations is that the causal account cannot be the whole story about scientific explanation, as there are scientifically legitimate non-causal and causal explanations.

We believe that the philosophy of explanation cannot but highly profit from a discussion of the examples of non-causal explanation. However, in the recent literature, the primary goal of discussing examples of non-causal explanations has been a negative one: showing that the received causal account cannot accommodate certain scientific explanations. The current debate has been largely silent on a more positive and constructive approach to non-causal explanations and to explanatory liberalism. The goal of this symposium is to advance such a constructive approach. To achieve this goal, all of the contributions to this symposium address the following core questions:

- What distinguishes causal and non-causal explanations?
- By virtue of what are non-causal explanations explanatory?
- Are there different kinds of non-causal explanations?
- Is there hope for a unified account of causal and non-causal scientific explanations? Or do non-causal explanations require a pluralist approach to explanations?

The four individual talks of the symposium assess the following distinct issues:

- **Bokulich's talk:** is it possible to accept a causal and a non-causal explanation of the same phenomenon? If this is so, what is the distinctive virtue of non-causal explanations? Are non-causal explanations sometimes superior to causal ways of explaining?
- **Frisch and Reutlinger's talk:** can a counterfactual account of explanation provide a unified account of both causal and non-causal explanations? Does the counterfactual account apply to all types of non-causal explanation?
- **Huneman's talk:** what kinds of non-causal structural explanations are there (e.g. topological and purely statistical explanations)? Can one develop and defend a convincing pluralist view regarding different types of non-causal structural explanation?
- **Jansson and Saatsi's talk:** what role does 'abstraction' play in a scientific explanation? Are abstract explanations distinctively mathematical and non-causal *because of* being abstract?

## Abstracts

### 1. Alisa Bokulich: *On the Supposed Incompatibility of Causal and Non-Causal Explanations*

According to the so-called causal imperialists, all scientific explanations are causal explanations – to explain a phenomenon is just to cite the causes of that phenomenon. Defenders of non-causal explanation have traditionally challenged this imperialism by trying to find an example of a phenomenon for which no causal explanation is available. If the imperialist, in turn, can find a causal explanation of that phenomenon, then it is believed that the defender of non-causal explanation has been defeated. Implicit in this dialectic is the

assumption that finding an example of a non-causal explanation requires finding something like an uncaused event, and more generally that causal and non-causal explanations of a phenomenon are incompatible. This has left non-causal explanations as relatively few and far between, relegating them to fields such as fundamental physics or mathematics.

In what follows I challenge this assumption that non-causal explanations require finding a phenomenon for which no causal story can be told. Instead I argue that one can have a non-causal explanation of a phenomenon even in cases where a complete causal account of the phenomenon is available. Moreover, in some cases these non-causal explanations will be superior to the detailed causal ones.

I illustrate these points by drawing on an example from the Earth sciences: namely, explanations of the shapes of sand dunes. Sand dunes come in different morphology classes, such as barchan (crescent) dunes, linear dunes, and star dunes. Although there is no doubt that there is a complete causal story that in principle could be told about all the forces acting on the grains of sand that resulted in a particular barchan dune, for example, that is not incompatible with there being a non-causal explanation of the dune morphology as well. Indeed it is often these non-causal explanations of the phenomenon that are the ones scientists are particularly interested in. I use this case to clarify one important sense of non-causal explanation, and conclude that they are in fact more widespread than the traditional examples have led us to believe.

## **2. Mathias Frisch and Alexander Reutlinger: *A Counterfactual Account of Non-Causal and Causal Explanations***

Are all scientific explanations causal explanations? David Lewis (1986), Wesley Salmon (1984) and Brad Skow (2014) argue the answer is 'yes', but in the last decade there have been several papers (by Batterman 2002; Bokulich 2008; Colyvan 2001; Lange 2013a,b, among others) that argue 'no'. Our aim in this paper is to advance our understanding of this question. Our goals are threefold:

First, we propose a general account of scientific explanations, a counterfactual account, which can accommodate paradigmatic examples of both causal and non-causal explanations (see Bokulich 2008, Saatsi and Pexton 2013). According to the counterfactual account, both causal and non-causal explanations are explanatory in revealing that the explanandum counterfactually depends on the explanans. We argue for this account by applying it to paradigm cases of non-causal explanations (mathematical explanations, purely statistical explanations, renormalization group explanations).

Second, we show that this framework allows us to capture one central source of the disagreement between those who affirm and those who deny that all explanations are causal: the disagreement, we argue, consists in the different demands placed on causal explanations by the different sides in the debate. For instance, Lewis and Skow characterize causal explanation as providing “causal information” while others think that explanations are causal because of being “abstract” (see Batterman 2002, Pincock 2012). Our positive proposal is to distinguish between causal and non-causal explanations by drawing on essential

features of causal relations such as locality and distinctness of the relata, asymmetry, and ‘intervenability’. We argue that if one adopts the view on which all explanations are causal, then this has the consequence that we cannot capture what the distinctive function of paradigmatically causal explanations is in helping us to make our way about in the world.

Finally, we will focus on the problem of explanatory asymmetries. We argue that it is possible to capture the explanatory asymmetry for non-causal explanation in the framework of the counterfactual account (by appeal to pragmatic considerations, asymmetric non-causal dependence relations, and the modal strength of non-causal explanatory assumptions).

### **3. Philippe Huneman: *Varieties of Structural Explanations and the Notions of Explanatory Pluralism***

Since a decade philosophers of biology embraced the view that explaining is to unraveling a mechanism; neurosciences (Craver 2007) and molecular biology (Darden 2006) provide many evidences for this view. However, it has been increasingly argued that some explanations are not mechanistic (Batterman 2010), because the fact that some mathematical properties are instantiated is playing the crucial explanatory role. I label “structural explanations” those explanations in which mathematics are not representative, but explanatory. Families of mathematical properties thereby define several types structural explanations: e.g. optimality explanations common in economics or behavioural ecology (Krebs and Davies 1991), purely statistical/probabilistic explanations usual in population genetics (Lange 2013b) or when the central limit theorems is appealed to, topological explanations, common in ecology (Huneman 2010). I intend to account for the specificity of such explanations, their variety, and question the proper explanatory pluralism required to make sense of their coexistence.

I start by clarifying the senses of “mechanistic explanation”, in order to show what sets structural explanations apart from mechanisms. Then I distinguish kinds of structural explanations in evolutionary biology and ecology according to the types of mathematical properties playing an explanatory role: I focus on optimality explanations used by behavioural ecologists and topological explanations, instantiated by network analysis in ecology and fitness landscapes in population geneticists.

Finally I contrast two *pluralisms* – both concerning mechanistic vs. structural explanations as well as various kinds of structural explanations (the case examined here): the *consilient* one, where for any phenomenon two mathematically intertranslatable distinct explanations coexist, and the *divergent* one, in which a set of phenomena calls for two mutually exclusive explanations. While robustness in ecology illustrates the latter pluralism, I will study how in evolutionary biology a consilient pluralism brings together optimality style explanations and topological explanations – referring to the recent notion of Formal Darwinism elaborated by Grafen to unify seemingly exclusive research traditions in evolutionary biology (Grafen 2002, 2009, Huneman 2014)

**4. Lina Jansson and Juha Saatsi: *Explanatory Abstraction in a Counterfactual Framework***

Scientific explanations vary in their degree of abstractness. Some explanations turn on very concrete explanatory features, while others appeal to highly abstract explanantia. Yet others fall between the two extremes. But what does abstraction in a scientific explanation amount to? We will propose an answer to this question, and consider the ramifications to two live issues:

1. Are there sui generis, non-causal, 'abstract explanations' in science? (Pincock forthcoming)
2. Are some scientific explanations 'distinctly mathematical' by virtue of being abstract? (Lange 2013a)

One natural intuition is that an explanation's abstractness has to do with a lack of specificity: more abstract explanations have more abstract explanantia, which in turn can be (partially and comparatively) understood in terms of the possible cases to which the explanation applies. (Weslake 2010) We will analyze this idea of abstraction-qua-lack-of-specificity in the context of a counterfactual account of explanation in the spirit of Woodward (2003). This counterfactual framework has natural conceptual resources for analyzing abstraction in terms of the notion of 'same-object counterfactual': the more abstract the 'same object', the more abstract the explanation. We will refine the crucial notion of 'same-object counterfactual' by analyzing the way in which a scientific theory fixes what counts as the 'same object' in explanatory counterfactual reasoning.

Armed with this analysis of explanatory abstraction, we will criticize Pincock's idea that some 'highly abstract' explanations – such as Königsberg's bridges and Plateau's laws for soap bubbles – should be understood as sui generis abstract explanations. We will argue that a modified 'same-object counterfactual' analysis can naturally accommodate these very abstract explanations. We will also briefly note how our analysis of explanatory abstraction relates to Lange's recent analysis of 'distinctly mathematical' scientific explanations.

**References**

- Baker, A. (2009): "Mathematical Explanations in Science" *British Journal for the Philosophy of Science* 60: 611–33.
- Balashov, Y. and M. Janssen (2003): "Critical Notice: Presentism and relativity" *British Journal for the Philosophy of Science* 33: 327-346.
- Batterman, R. (2002): *The Devil in the Details. Asymptotic Reasoning in Explanation, Reduction and Emergence*, New York: Oxford University Press.
- Batterman, R. (2010): "On the Explanatory Role of Mathematics in Empirical Science", *British Journal for Philosophy of Science* 61: 1-25.
- Bokulich, A. (2008) "Can Classical Structures Explain Quantum Phenomena?" *British Journal for the Philosophy of Science* 59(2): 217–235.
- Cartwright, N. (1983): *How the Laws of Physics Lie*, Oxford: Oxford University Press. Cartwright, N. (1989): *Nature's Capacities and Their Measurement*, Oxford: Clarendon Press.
- Colyvan, M. (2001): *The Indispensability of Mathematics*, Oxford: Oxford University Press.
- Craver, C. (2007) *Explaining the Brain*, New York: Oxford University Press.
- Darden, L. (2006), *Reasoning in Biological Discoveries: Mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge: Cambridge UP.
- Friedman, M. (1974) "Explanation and Scientific Understanding", *Journal of Philosophy* 71: 5- 19.
- Grafen, A. (2002): "A first formal link between the Price equation and an optimisation program." *J. Theor.*

*Biol.* 217: 75–91.

- Grafen A (2007): “The formal Darwinism project: a mid-term report”. *J Evol Biol* 20: 1243–1254.
- Hempel, C. G. (1965) *Aspects of Scientific Explanation*, New York: Free Press. Huneman, P. (2010): “Topological Explanations and Robustness in Biological Sciences”, *Synthese* 177: 213-245.
- Huneman, P. (2014): “A Pluralist Framework to Address Challenges to the Modern Synthesis in Evolutionary Theory”, *Biological Theory* 9(2): 163-77.
- Hüttemann, A., R. Kuhn and O. Terzidis (forthcoming): “Stability, Emergence and Part-Whole Reduction”, In *Why More Is Different Philosophical Issues in Condensed Matter Physics and Complex Systems*, B. Falkenburg and M. Morrison (eds.), New York: Springer.
- Kitcher, P. (1989) “Explanatory Unification and the Causal Structure of the World”, In P. Kitcher and W. Salmon (eds.) (1989), 410–505.
- Krebs J.R. and Davies N.B. (eds.) (1984): *Behavioural Ecology*, Oxford: Blackwell. Lange, M. (2011): “Conservation Laws in Scientific Explanations: Constraints or Coincidences?” *Philosophy of Science* 78: 333-352.
- Lange, M. (2013a): “What Makes a Scientific Explanation Distinctively Mathematical?”, *British Journal for the Philosophy of Science* 64: 485-511.
- Lange, M. (2013b): “Really Statistical Explanations and Genetic Drift”, *Philosophy of Science* 80: 169-88.
- Lewis, D. (1986): “Causal Explanation”, In D. Lewis (1986) *Philosophical Papers Vol. II*, New York: Oxford University Press, 214-240.
- Lipton, P. (1991/2004) *Inference to the Best Explanation*, London: Routledge. Machamer, P., L. Darden, and C. Craver (2000): “Thinking about Mechanisms” *Philosophy of Science* 67: 1-25
- Morrison, M. (2012): “Emergent Physics and Micro-Ontology” *Philosophy of Science* 79: 141-166.
- Nerlich, G. (1979): “What Can Geometry Explain?” *British Journal for the Philosophy of Science* 30: 69-83.
- Pincock, C. (2012) *Mathematics and Scientific Representation*, New York: Oxford University Press.
- Pincock, C. (forthcoming): “Abstract Explanations in Science”, *British Journal for the Philosophy of Science*.
- Reutlinger, A. (2014): “Why is there universal macro-behavior? Renormalization group explanation as non-causal explanation” *Philosophy of Science* 81: 1157-1170.
- Salmon, W. (1984) *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.
- Salmon, W. (1989). “Four Decades of Scientific Explanation”, In Philip Kitcher and Wesley Salmon (eds.), *Scientific Explanation: Minnesota Studies in the Philosophy of Science, Volume XIII*. Minneapolis: University of Minnesota Press, pp. 3-219.
- Saatsi, J. and M. Pexton (2013): “Reassessing Woodward’s account of explanation: regularities, counterfactuals, and non-causal explanations”, *Philosophy of Science* 80: 613-624.
- Skow, B. (2014): “Are there non-causal explanations (of particular events)?” *British Journal for the Philosophy of Science* 65: 445-467.
- Strevens, M. (2008) *Depth*, Cambridge, MA: Harvard University Press. Van Fraassen, B. (1989): *Laws and Symmetry*. Oxford: Clarendon Press.
- Weatherall, J. (2011): “On (Some) Explanations in Physics”, *Philosophy of Science* 78: 421-447.
- Weslake, B. (2010): “Explanatory Depth” *Philosophy of Science* 77: 273-294. Woodward, J. (2003): *Making Things Happen*, New York: Oxford University Press.

**Symposium on Approaches in Philosophy of  
Science in Practice**

**Symposium**

Organizer: Marcel Boumans

Chair: Alexander Christian

Room 5H, Thursday 09:30 – 11:30

---

*Symposium on Approaches in Philosophy of Science in Practice*

MARCEL BOUMANS

University of Amsterdam

m.j.boumans@uva.nl

---

*An Argument for Local Critique in Philosophy of the Social Sciences: The Case of Rational Choice Theory*

CATHERINE HERFELD

Munich Center

catherine.herfeld@lrz.uni-muenchen.de

---

*Modeling Multi-level Disorders: Overcoming the Mechanistic-systemic Dichotomy*

MARTA BERTOLASO

UCBM

m.bertolaso@unicampus.it

RAFFAELLA CAMPANER

University of Bologna

raffaella.campaner@unibo.it

---

*Science in the Flesh: The Epistemological Role of Bodily Sensations and Operations in 20<sup>th</sup> Century*

*Oceanography*

LINO CAMPRUBI

Max Planck Institut

lcamprubi@mpiwg-berlin.mpg.de



---

*Making Sense of Theoretical Practices: Scripts, Scruples, and the Mass of the Universe*

JACO DE SWART

University of Amsterdam  
jacobusgdeswart@gmail.com

---

**1. Marcel Boumans: *Symposium on Approaches in Philosophy of Science in Practice***

The Society for Philosophy of Science in Practice grew out of a recognition of the need to promote the philosophical study of “science in practice”, by which the organizers of the Society meant both scientific practice and the functioning of science in practical realms of life. Despite occasional exceptions such as some recent literature on models, experimentation, and measurement which have engaged in detailed consideration of scientific practices in pursuit of their philosophical points, concern with practice has tended to fall outside the mainstream of analytic philosophy of science. SPSP was founded with the aim of changing this situation, through the promotion of conscious, detailed, and systematic study of scientific practice that nevertheless does not dispense with concerns about truth and rationality.

In many of its traditional forms, philosophy of science has focused on the relation between scientific theories and the world, often to the neglect of scientific practice. In contrast, in social studies of science and technology the dominant trend has been to examine scientific practice as a human creation, sometimes willfully disregarding the world except as a product of social construction. Both of these approaches have their merits, but they each offer only a limited view neglecting some important perspectives or approaches which are necessary for the development of a fuller picture of science. If we are interested in exploring the assumptions and methods underlying the sciences, it is essential not only to explore the theories and results produced by scientists, but the processes by which they came to these conclusions. And what we learn from history of science is that scientific practices should be evaluated in their historical contexts reaching up to the present moment.

Without excavating underneath the tidy surface of published papers or finalized theories, it is extremely difficult to identify these processes. SPSP is dedicated to fostering the pursuit of a philosophy of science that considers theory, practice and the world simultaneously, and never in isolation from each other.

The direction of philosophy of science we advocate is not entirely new and clearly has prestigious forerunners. For example, naturalistic philosophy of science has often emphasized the need to study scientific practices. In integrated approaches to the history, philosophy and sociology of science, much philosophical attention has been paid to actual scientific practices. Outside of the philosophy of science, pragmatists, ordinary-language philosophers and followers of the later Wittgenstein have attempted to ground truth and meaning in practices.

And those in the Continental philosophical traditions often have also emphasized the need to consider experience and practice, as well as rejecting the positivist traditions which they view as privileging science unduly and taking scientific progress for granted. Scholars participating in SPSP activities have been making conscious efforts to inject these lines of thought into analytic philosophy of science.

In order to understand the implications of the SPSP approach, it is necessary to consider what is meant by 'practice.' Practice consists of organized or regulated activities aimed at the achievement of certain goals. Therefore, any useful investigations of a particular form of practice must elucidate what kinds of activities are associated with and required for the generation of knowledge in that domain. Traditional debates in the philosophy of science concerning epistemological concepts such as truth, fact, belief, certainty, observation, explanation, justification, evidence, and so on may be usefully re-framed in terms of activities. Rather than asking abstract or theoretical questions about the appropriate scientific standards for evidence, recasting the questions of interest in terms of activities allows us to explore various (and often competing) approaches to the generation and weighing of evidence. Examining the goals underlying the activities associated with science also forces us to focus not only on epistemological considerations but also on the values, norms, and ideals inherent in the pursuit of scientific knowledge. Further, it encourages us to question the metaphysical and ontological assumptions underlying these practices rather than taking them as obvious or as unquestionable 'givens.' In short, focusing on practice allows philosophy of science to return to fundamental issues which have increasingly become neglected in favor of a relatively narrow preferred approach to the field which is largely epistemic, highly theoretical, and often overlooks the implications of the sciences as practiced.

The purpose of this session is to present some of its approaches in studying scientific practices that have begun to emerge as SPSP has taken shape and grown. On the whole, the papers presented in this session are pioneering forays into new directions. Our wish is that they will stimulate further work by illustrating the kind of philosophy of science that we are attempting to promote in SPSP.

## **2. Catherine Herfeld: *An Argument for Local Critique in Philosophy of the Social Sciences: The Case of Rational Choice Theory***

This paper presents a meta-methodological undertaking in philosophy of the social sciences. The question addressed is how philosophers of the social sciences can fruitfully appraise theoretical frameworks while taking scientific practices seriously. The point of departure is the persistent criticism that appraisal in philosophy of the social sciences is fruitless for the improvement of science, if a theoretical framework is assessed in isolation.

Taking scientific practices into account faces several dangers. One danger is that if philosophers get too close to actual practices, then philosophical analysis lapses into a purely descriptive enterprise with its primary focus on reconstructing the scientist's undertaking. In this case, philosophy gives up its normative

function. A further difficulty is that notions of ‘contextualization’ and ‘scientific practice’ have multiple, often vague, meanings, yet require precise specification in order to make them useful for appraisal. Finally, even if philosophical accounts are based upon a detailed analysis of scientific practices, their formulation is often inspired by the natural sciences. As such, the usefulness of those accounts for approaching the social sciences is limited.

To cope with these challenges, I make a case for applying what I call ‘the method of local critique’ in philosophy of science. This method to appraisal is inspired by Philip Kitcher’s approach to biology. It draws upon the idea of directly engaging with cases representative of a specific practice and/or with the arguments of one’s opponents, while allowing for general conclusions about the epistemic potentials and limitations of these practices. By restricting the scope of my argument to philosophy of the social sciences, I claim that Kitcher’s philosophical approach to biology opens up a possible way for philosophers of the social sciences to take into account the specifics of the social world and thereby meet the complaints of their critics.

### **3. Marta Bertolaso and Raffaella Campaner: *Modeling Multi-level Disorders: Overcoming the Mechanistic-systemic Dichotomy***

Remarkable progresses in the biomedical sciences in the last decades have led to increased awareness of the complex character of most pathologies, as well as greater appreciation of the difficulties in representing and explaining their features and behaviors. A pressing question concerns the adequacy of molecular biology and systems biology to account for complex diseases. In the philosophical scenario, the discussion has been largely focusing on whether the features attributed to mechanistic models by the “new mechanistic philosophy” can adequately account for explanatory models in biomedical and cognitive sciences or on multilevel and integrated approaches in biological sciences.

This paper aims to clarify the philosophical relevance of convergences and complementarity in scientific explanatory accounts when dealing with multilevel complex diseases. Two case studies will be discussed from cancer research and neuropsychiatry showing how cutting through traditional disciplinary barriers novel approaches are currently developing in scientific practice. More specifically, we will:

- 1) show how the elaboration of the explanatory model rather starts from the choice of a minimum set of characterizing features of the target system, which can be regarded as an array of inter-regulatory subsystems. In the process, any progress in mechanistic understanding of some level further constrains the space of possible mechanisms underpinning the disorder, with descriptive and explanatory processes co-evolving, and correcting each other;
- 2) elaborate on this, highlighting some distance between the philosophical debate on mechanistic models and how disorders are actually – and always tentatively – modeled, and the need of further conceptual tools to give justice to the dynamics of modeling disorders at the crossroad of known and unknown systems.

These reflections will be accompanied by some rethinking of explanatory dependence and independence in the light of the actual adoption of explanatory models in scientific practice, and, more specifically, their impact on clinical contexts.

**4. Lino Camprubí: *Science in the Flesh: the Epistemological Role of Bodily Sensations and Operations in 20th Century Oceanography.***

Although the role of materiality in the generation of scientific knowledge is acknowledged, subject-object relationships do not provide a sufficient framework for understanding science beyond theoretical propositions. Each science is a rich structure or body made of a wealth of percipient subjects working with a wide variety of objects and apparatuses and transforming important parts of the world as much as making sense of them. The human body seems to fade in the face of these humongous structures of interrelated *things*. When it comes to the modern earth sciences, this contrast in scale is even more acute, since interconnectedness and global circulation would seem to render individual and local perception irrelevant.

This paper deals with the role of the human body in modern oceanography. It does so through exploring a specific sensory organ: hearing. Transducers, hydrophones and sonar equipment transformed the ocean, once the epitome of a silent immensity, into a sonically rich environment. Making sense of sound required developing a precise knowledge of underwater acoustics. In the 20th century some of the largest research projects in oceanography attempted to model this relationship and produced detailed local and global maps of current circulation. This research would be extended in the 1980s to tracing temperature variations across the globe. The goal this time was not to monitor submarines but a changing climate.

This paper will explore the role of sonar operators and other listening subjects in the development of global oceanography. I will discuss how listening practices and devices changed over time transforming the oceans themselves and what we know about them. In its turn, new oceanographic knowledge informed new technologies and also new experiences of underwater listening – new *phenomenotechniques*.

**5. Jaco de Swart: *Making Sense of Theoretical Practices: Scripts, Scruples, and the Mass of the Universe***

The scientific practices we could signify as “theoretical” – involving formalisms, equations and calculations – have enjoyed relatively little attention. As Bruno Latour has put it: “almost no one has had the courage to do a careful anthropological study of formalism”. In this paper I will take this observation seriously, and elaborate on some recent ideas of Latour to show that a new and more performative terminology will hand tools to better approach theoretical practices.

Latour analyses acts of calculation as they appear in economic activity, and are used to “express preferences, to establish *quittances*, to trace *ends* [and] to *settle* accounts”. He deploys the notion of *scripts* – constraining narratives – and Frank Cochoy’s notion of *qualculation* – quality-based judgements – to make

sense of what he refers to as the *scruples* of organisational, moral and economic action. I will show that the notions of scripts, scruples and qualculations are very suitable to study theory as a scientific practice.

To explore this, I make use of an example from early twentieth century physical cosmology to understand the mass density of the universe. The arguments, assumptions and the calculation of a mean density of the universe in this work exemplifies the piling up of scripts and coping with scruples as activities in producing theoretical knowledge. It will become clear that extending Latour's new work to a context of theoretical science can indeed offer a valuable set of tools that helps to shift attention towards a more performative assessment of theory as practice. I will emphasize this by showing that the activity of making objects adequate, the process of *adequation*, plays a central role in the performance of theory. Contrasting this perspective with Latour's earlier focus on centres of calculation and their bookkeeping, I hope to create room for the practices of theoretical sciences to be followed more closely.

---

*Scientific Realism and Fundamental Physics*

CARL HOEFER

ICREA, Universitat de Barcelona

carl.hoefer@uab.es

---

Scientific realism (SR) holds that we have reason to believe that theories in the "mature" sciences are approximately true, and that the theoretical and non-observable entities posited by those theories really exist. In discussions of SR, physical theories get more attention than those of all the other sciences combined, perhaps due to the sociological phenomenon that physics is considered the "Queen" of the sciences.

In this paper I will argue that this is a mistake. In fact our current fundamental - or possibly-fundamental physics theories are by no means things which we can regard as true or approximately true, nor are their posited entities clearly things that we should (or in some cases, *can*) believe to exist. In this respect fundamental physics theories – despite their enormous empirical successes – are quite unlike the majority of the rest of what we take to be our best current theories in the mature sciences. I will explore why these other theories and sciences are secure in a way in which fundamental physics cannot be, now or in the foreseeable future.

I will center my attention on two areas of fundamental physics. The first, to be discussed only briefly, is General Relativity theory. About GR, I will point out that physicists themselves do not seem to take GR to be a candidate for full-stop truth.

The descriptions of both matter-energy and spacetime structure in GR are such that physicists do not take them to be serious candidates for truth except in an approximate sense that amounts to nothing more than mere empirical adequacy.

The second area I will discuss is quantum mechanics, both non-relativistic standard quantum theory and quantum field theories such as those of the Standard Model. I will discuss how the well-known measurement problem of QM impedes our taking the theory as in any serious sense "approximately true". Current debates about whether we should take the wave-function (and the configuration space in which it is defined) to represent anything physically real are symptomatic of the deep difficulties QM makes for a realistic attitude. I will also discuss the Unruh effect and other features of quantum field theories that make the ontological fundamentals of the theories extremely unclear in a different way.

Mathias Egg (2012) discusses some of the special problems for SR raised by fundamental physics theories. He defends a version of entity realism or "causal" realism to overcome the problems. I will argue that his

solution is insufficient, being based on a realistic understanding of both quantum particles and the things they do (e.g. moving through space from *A* to *B*, carrying energy) that is untenable if one takes the standard accounts of quantum theories (especially field theories) seriously.

In consequence, I will urge that scientific realists should remove fundamental physics from their lists of things to take realistically.

By contrast, the theories and more generally the accumulated lore of other areas of science, since the mid-20th century at least, are secure and *believable* in a way that is radically different from how things stand in fundamental physics. In these other areas there is no reason to think that we may have to radically revise our beliefs in such a way that we no longer think the earlier theories and beliefs were *approximately* true. In fact, I will argue, it is difficult to imagine a scenario where such radical revision takes place without the scenario verging towards radical skeptical scenarios of the Cartesian Demon/brain-in-a-vat type. Whereas in fundamental physics *unconceived alternatives* are easy to imagine (and are produced regularly), in the bulk of the rest of our best sciences they are just not a threat.

If time permits I will offer some further reflections on the realism-antirealism debates and how SR can best be defended.

---

*Approximate Truth and Scientific Realism*

ROBERT NORTHCOTT

Birkbeck College, University of London

r.northcott@bbk.ac.uk

---

Historically, the motivation for defining a scientific theory's approximate truth has mainly come from the scientific realism debate. Indeed, finding such a definition has been seen by some as essential for buttressing the realist position. As anti-realists often point out, philosophers have had great difficulty in giving a plausible and consistent account of approximate truth. Yet a good and useful definition of it can be found nevertheless – but only once we cast off this inherited entanglement with scientific realism. It turns out that influential recent work in the causation literature is a much more fertile inspiration, as approximate truth can be well defined in causal terms. The crucial move is to change our focus from theories as a whole instead to application-specific models.

Why reject the attempt to define approximate truth for theories as a whole? The biggest reason is a fundamental difficulty facing any such attempt: namely that a theory's errors can be very empirically costly in one application but not at all costly in another, thus leaving it ill-defined how serious those errors are in any context-independent or absolute sense. For example, in dynamical systems theories, should we prefer a theory whose dynamic equations are almost correct but whose empirical predictions quickly become wildly wrong, or a theory with the opposite pattern? Moreover, there seems to be no good way of making sense of 'ontological' approximate truth independent of empirical success in particular applications.

Another common difficulty is the attempt to capture in one measure both accuracy and comprehensiveness, so that, for instance, we are not forced to rank a trivial tautology above the false but widely useful Newtonian theory. But again this difficulty melts away once we relativise approximate truth to specific applications, for then the Newtonian theory's much wider scope will immediately be reflected by it scoring well in many more applications.

Even if we accept these general reasons for preferring an application-specific approach, how exactly should that be carried out? Drawing on previous work, I show how a definition can be framed in terms of causes, in a manner natural to the special sciences. Roughly speaking, according to it a model is approximately true if it captures accurately the strengths of the causes actually present in a given situation. Accordingly, getting closer to the truth consists in capturing these causal strengths more and more accurately. In order to make this idea precise, the notion of degree of causal importance, or causal strength, has to be defined, and then also a measure of closeness between a model's allocation of causal strengths and the true allocation.



Completing these tasks leads naturally to a definition with just the application- specificity needed to solve the problems above that confound a more generalist approach.

The definition brings with it other advantages. One is that it is not an abstract logical measure but rather is couched in the causal language that actual scientists use. Moreover, a high score for approximate truth now guarantees empirical success. It also carries another easily interpreted implication, namely that it guarantees accurate quantitative predictions of the impact of *interventions* – here, the recent extensive literature connecting causation with interventions pays dividends. Moreover, the counterfactual element of causation allows us to avoid rewarding ‘fluke’ empirical successes. (The definition also has several other attractive technical features.)

There is another way too in which the approach ties neatly into scientific practice. Much recent philosophical work has focused on the many cases in which progress does not come via development of new theory but rather via a lot of case-specific extra-theoretical investigation. Often, we end up with an empirical model tailored very closely to a unique event or task, but which cannot be derived from theory or even piecemeal from a group of theories or by trial-and-error tinkering with a theory’s parameter values. Progress towards the truth in such cases is well represented by an application-specific causal definition – but is inevitably invisible if we define approximate truth in terms of theories as a whole.

This, finally, is where the inheritance from the scientific realism debate reveals itself to be unhelpful. In particular, that debate has usually concerned itself with whether we should be realist about *theories*. ‘Convergent realism’, for instance, postulates that our best theories are over time gradually getting closer and closer to the truth. Yet one implication of viewing approximate truth application-specifically is that progress towards the truth is only ever a local not a global phenomenon – quite contrary to convergent realism. Historically, from Popper on, the approximate truth literature has overwhelmingly been focused on theories, not application-specific models. Yet, I argue, it is the latter that is the true route to success.

---

*Who is Afraid of Multiple Realisability?*

FOAD DIZADJI-BAHMANI

California State University Los Angeles

foad.dizadji.bahmani@gmail.com

---

**Overview**

Multiple realizability is an important issue in the context of reduction. Putative multiply realizable properties (MRPs) have been used in a variety of ways to argue both against reductionism and against specific reductions. One class of such arguments pertains to ontological simplification: it is widely claimed that ontological simplification is required for successful reduction. However, MRPs undermine the ontological simplification that a reduction is to afford, or so it has been argued. (See, for example, Bickle (1996); Kim (1992, 2000); Endicott (2005); Esfeld and Sachse (2007); Lyre (2009)). I proffer a novel route to ontological simplification, one which is not undermined by MRPs. I then preempt one important potential criticism of this approach, and argue against it.

**Context and Motivation**

Whether reductionism is true or whether a specific reduction is successful turns on what reduction is. The model of reduction that underpins much of the literature on reduction is Nagel's. The Nagelian model has been developed by Schaffner (1967), and subsequently by Dizadji-Bahmani et al. (2010), and I shall adopt this last version here.<sup>1</sup> Granting that reduction better had afford ontological simplification, the salient question is whether Nagelian reduction does. The standard answer is: yes, via what I call 'simplistic ontological simplification' (SOS), according to which, one ought to construe bridge-laws as identity statements. Thus, what were taken to be two distinct properties are, via the reduction, shown to be two different names for the same property. How do MRPs frustrate SOS? To answer this, consider the definition of a MRP: A 'higher-level' property is a MRP iff it is realizable by two (or more) 'lower-level' properties. However, given that identity is a transitive relation, and that by hypothesis the 'lower-level' properties are *not* identical, the 'higher-level' property cannot be identical to both. The argument from MRPs against SOS goes back as least as far as Putnam (1967), and has generated a *huge* volume of work. There are arguments for the domain specificity of identities: the 'higher-level' property is to be identified with 'lower-level' property in a particular domain. (See (Lewis, 1969), Sklar (1993) Sober (1999), and Marras (2002).) The problem that the anti-SOSist

---

<sup>1</sup>Both Richardson (2008) and Dizadji-Bahmani et al. (2010) have pointed out that Nagelian reduction does not require ontological simplification – it is at most a desideratum – but I do not pursue this exegetical point here.

fi with this move is that it forgoes the unity of the ‘higher-level’ property: in what sense can one still say that the ‘higher-level’ property is a single property? (See Kim (1992).) Moreover, SOS is still undermined by so-called ‘radical’ multiple realizability, properties which are multiply realized in the same domain. (See Horgan (1993) and (Bickle, 1996), and for counterarguments cf. Shapiro (2000) and Lyre (2009).) It is hard to overstate just quite what a morass the issue of multiple realizability and ontological simplification is! Rather than wading through it, I proffer a different route to ontological simplification.

### Central Argument

My central argument is that it is reduction *as a whole*, not bridge-laws *qua* identities, which affords ontological simplification. And it does so in a way that is unaffected by MRPs.

I appropriate Quine’s (Quine, 1948) well-known meta-ontological position as per the slogan that “to be, is to be the value of a bound variable *in our best conceptual scheme*.” The pertinent question is it takes to be the ‘best conceptual scheme’. Quine himself does not set out in much detail what this amounts to but he does point us in the right direction, namely that we adopt “*the simplest conceptual scheme into which the disordered fragments of raw experience can be fitted and arranged*.” (*ibid* )

I suggest the following development of Quine’s position: the best conceptual scheme is one which, in order of priority, best balances empirical adequacy and simplicity. Sort conceptual schemes by empirical adequacy *fi* and then select the simplest amongst those. This is, of course, a context specific procedure but I do think that the idea here is intuitively appealing. Now consider two theories, a ‘higher-level’ theory T1 and a ‘lower-level’ one theory T2. Suppose that a successful reduction of T1 to T2 is carried. The reduction ensures that the empirical content of T1 is captured by T2, so, *ceteris paribus*, any conceptual scheme which includes T2 is at least as strong as one which contains T1. Given the ‘priority clause’ above this rules out conceptual schemes containing T1 but not T2. Moreover, the reduction of T1 to T2 ensure that conceptual schemes containing both T1 and T2 are ruled out on grounds of simplicity: including T1 as well as would be superfluous as its empirical content is accounted for, or so I argue. The route to ontological simplification is now straightforward: if T1 is not in our best conceptual scheme then there is no ontological commitment to its ontology. If we suppose that prior to a reduction both T1 and T2 are in our best conceptual scheme with the ontological baggage that that entails, then after a reduction our ontological load will be lighter.

In the paper I illustrate this idea by using the example of the reduction of thermodynamics to statistical mechanics, and by considering the property of temperature. This *fi* well with the above discourse because temperature is often cited as a MRP. (See, for example, Sklar (1993) and Lyre (2009).)

### Responding to Objections

I preempt the following important potential criticism of this approach: the empirical content of T1 is not captured by T2 alone, but by T2 supplemented with bridge laws. Thus the best conceptual scheme includes

those bridge principles, and those bridge principles refer to the ontology of T1. We thereby will keep the commitment to those entities in T1, and so there's no ontological simplification after all! My response to this that the use of bridge laws does not commit one to the ontology of T1. I argue that bridge laws are, in fact, theoretical constructs, contra to the prevailing views which considers them to be contingent or nomic correlations or identities. To do so I return to and draw on the aforementioned example of temperature and its 'lower-level' realizer, mean kinetic energy.

### Conclusions

A successful reduction is good reason to not include the reduced theory in our best conceptual scheme. Having adopted Quine's meta-ontological position this entails ontological simplification. The properties of reduced theories may well be kept as useful but when the dust of our scientific and philosophical work has settled, we are not bound to be committed to these properties. It is reduction as a whole, not bridge-laws *qua* identities in particular, which affords ontological simplification, and this is unaffected by MRPs.

### References

- Bickle, J. (1996). New wave psychophysical reductionism and the methodological caveats. *Philosophy and Phenomenological Research* 56 (1), 57–78.
- Dizadji-Bahmani, F., S. Hartmann, and R. Frigg (2010). Whos afraid of Nagelian reduction. *Erkenntnis* 73 (2), 393–412.
- Endicott, R. P. (2005). Multiple realizability. In *Encyclopedia of Philosophy, 2nd edition*. Thomson Gale, Macmillan Reference.
- Esfeld, M. and C. Sachse (2007). Theory reduction by means of functional sub-types. *International Studies in the Philosophy of Science* 21 (1), 1 – 17.
- Horgan, T. E. (1993). Nonreductive materialism and the explanatory autonomy of psychology. In S. Wagner and R. Wagner (Eds.), *Naturalism: A Critical Appraisal*. University of Notre Dame Press.
- Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52 (1), 1–26.
- Kim, J. (2000). *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. MIT Press.
- Lewis, D. (1969). Review of "art, mind, and religion". *Journal of Philosophy* 66.
- Lyre, H. (2009). The "multirealization" of multiple realizability. In A. Hieke and H. Leitgeb (Eds.), *Reduction – Abstraction - Analysis*. Proceedings of the 31th International Ludwig Wittgenstein-Symposium in Kirchberg. Ontos Verlag.
- Marras, A. (2002). Kim on reduction. *Erkenntnis* 57 (2), 231–57.
- Putnam, H. (1967). Psychological predicates. In W. H. Capitan and D. D. Merrill (Eds.), *Art, Mind and Religion*. Pittsburgh University Press, Pittsburgh.
- Quine, W. V. (1948). On what there is. *Review of Metaphysics* 2, 21–38.
- Richardson, R. C. (2008). Autonomy and Multiple realization. *Philosophy of Science* 75 (5), 526–536.
- Schaff K. F. (1967). Approaches to reduction. *Philosophy of Science* 34 (2), 137–147.
- Shapiro, L. A. (2000). Multiple realizations. *Journal of Philosophy* 97 (12), 635–654.
- Sklar, L. (1993). *Physics and Chance: Philosophical Issues in the Foundations of Statistical Mechanics*. Cambridge University Press.
- Sober, E. (1999). The multiple realizability argument against reductionism. *Philosophy of Science* 66 (4), 542–564.

---

*Representation, Models and Structure: A Reconceptualization*

FRANCESCA PERO

University of Florence  
francesca.pero@unifi.it

ELENA CASTELLANI

University of Florence elena.castellani@unifi.it

TARJA KNUUTTILA

University of South Carolina  
tarja.knuuttila@helsinki.fi

---

In this paper, we focus on the notion of structure as employed when considering the issue of scientific representation, in particular with regard to the functions of models in science. In the case of models, representation is usually cashed out in terms of the relationship between a model and its target. How to conceive such a relationship is particularly challenging when, as is very often the case, the model is an abstract mathematical structure and its target is an empirical phenomenon.

Representation is usually described either as a dyadic relationship, holding between the model and its target only, or as involving the pragmatics of model construction and models' users as well -- hence as a triadic relation. In both cases structures play a crucial role. On the one hand, structures are commonly employed to characterize models. On the other hand, structural relationships are used to connect models to the target according to users' intended scope.

Structuralist approaches to scientific theories have a long and respectable tradition in the philosophy of science. In particular, the semantic view of scientific theories and recent versions of structural realism have notoriously contributed to the philosophical interest in the role of structures and their connection to models. Which kind of structure to consider with respect to models, and how this structure is used and related to a target system in order for the model to “represent”, is a crucial point in the relevant literature. In the paper, we focus on this very point and argue that a source of confusion in current debates has to do precisely with a misleading use of structures.

More precisely, we find this use misleading in a twofold sense. First, in the literature the two levels at which the use of models (and related structures) takes place are seldom distinguished. Drawing on French's terminology (French 2012), we call these two levels the “object-level” and the “meta-level” of analysis. The object-level is that of working scientists, where scientific theories are elaborated and tested. At the “meta-level” of analysis, on the other hand, the results presented at the object-level are reconceptualized in terms of abstract structures such as sets or categories. The second sense in which the use of structures is misleading concerns the kinds of structures considered and their supposed linkage to the world. We argue for this point

by using examples from physics, biology and economics. One particularly interesting case study is provided by the Ising model, because of its wide and interdisciplinary range of applications, from physics to sociology and genetics (Knuuttila and Loettgers 2014). With respect to physics, for example, recent works have questioned the explanation of the success of the model in accounting for the macrolevel phenomena that are exhibited by very different kinds of systems undergoing phase transitions by appealing to some shared features or some common relevant causes (see Batterman and Rice 2014, Reutlinger 2014). The situation becomes even more complicated when the Ising models is transferred e.g. to economics, where methods used in physics do not apply, and the traditional paradigm has been one of giving microfoundations to macrophenomena.

### References

- Batterman, R.W and Rice, C.C. 2014. "Minimal Model Explanations." *Philosophy of Science* 81: 349-376.
- French, S. (2012), The Presentation of Object and the Representation of Structure, in Landry, E. and Rickles, D. (eds.), *Structural Realism*, The Western Ontario Series in Philosophy of Science, vol. 77, pp. 3-28, Springer, the Netherlands.
- Knuuttila, T and Loettgers, A. 2014. "Magnets, Spins, and Neurons: The Dissemination of Model Templates Across Disciplines." *The Monist* 97: 280-300.
- Reutlinger, A. 2014. "Why Is There Universal Macrobehavior? Renormalization Group Explanation as Noncausal Explanation." *Philosophy of Science* 81: 349-376.

---

*Significance Testing, P-Values and the Principle of Total Evidence*

BENGT AUTZEN

University of Bristol

b.autzen@bristol.ac.uk

---

Significance testing is widely used across the natural and social sciences. Given its popularity in scientific practice, it might come as a surprise that significance testing has attracted severe criticism in both the statistical and the philosophical literature. This paper is concerned with a particular objection made by Sober (2008): the claim that significance testing violates the Principle of Total Evidence (PTE). Sober argues that using a logically weaker rather than a logically stronger description of the data in a significance test violates PTE. More specifically, when calculating p-values in a significance test one does not consider the observations in all their detail but rather the fact that they fall in a certain region. In order to examine Sober's claim I will proceed in two steps. First, I will clarify the interpretation of PTE. Second, I will apply my reading of the principle to the question of whether the use of p-values for inductive inference violates PTE.

While PTE is invoked in a number of philosophical arguments, references to the principle are typically rather vague in nature. This is unfortunate as the merits of Sober's objection depend on a clear understanding of PTE. I will therefore spend considerable effort to state the principle in a precise way. By drawing on Carnap's *Logical Foundations of Probability* (Carnap, 1962) I will argue that the application of PTE requires the prior specification of a criterion for evidential assessment (or 'theory of evidence' for short). Given such a theory of evidence, I propose the following reading of the principle:

Suppose you are given a hypothesis  $H$  as well as data  $d_1$  and  $d_2$  with  $d_1 \Rightarrow d_2$  and  $d_2 \not\Rightarrow d_1$ , then an inference about  $H$  should be based on  $d_1$  if changing between  $d_1$  and  $d_2$  affects the evidential assessment.

Based on this reading of PTE, I will argue two points with regard to Sober's claim that significance testing violates PTE. First, I will argue that as it stands the claim is incomplete. In order to assess what data should be used for inductive inference, PTE requires the prior specification of a theory of evidence. Sober, however, does not presuppose a particular theory of evidence in his argument. Without such a specification PTE can neither be satisfied nor be violated by the use of p-values for inductive inference. Second, I will turn to what might be considered as the best case for Sober's argument: I will adopt the law of likelihood (LL) as a theory of evidence given the central role of LL in Sober's writings. I will argue that even under this assumption there is no universal conflict between the use of p-values for inductive inference and PTE. It depends on the particular inference at hand, whether the significance tester violates PTE from a likelihoodist perspective. In

particular, I will argue that in a one-sided significance test with a sufficient test statistic, using p-values for inductive inference does not violate PTE. Matters are different, however, in a two-sided significance test. Here, I will argue that the use of p-values indeed violates PTE. The discussion also reveals that the statistical notion of sufficiency and PTE can part ways and, hence, refer to two distinct ideas.

**References**

- Carnap, R. (1962). *The Logical Foundations of Probability*. Chicago: Chicago University Press.
- Sober, E. (2008). *Evidence and Evolution: The Logic Behind the Science*. Cambridge: Cambridge University Press.



---

*A Measure for Partial Knowledge*

DAVID ATKINSON

University of Groningen D.Atkinson@rug.nl

JEANNE PEIJNENBURG

University of Groningen jeanne.peijnenburg@rug.nl

---

It is a platitude that belief comes in degrees, but the same cannot at all be said of knowledge. Ever since the beginning of epistemology, knowledge has been seen as a yes/no notion that does not allow grades, and today the received view still is that knowledge is a categorical concept. During the past two decades, however, dissident voices have been heard, claiming that knowledge might after all allow a more or less. According to these dissidents, sentences like ‘Annie does not really know that  $p$ ’ and ‘Boris knows  $p$  better than Chris’ constitute linguistic evidence for the gradability of knowledge.

The dissident movement started when attributor contextualists such as Stewart Cohen and Keith DeRose argued that knowledge attributions are context sensitive: whether a proposition like ‘John knows that  $p$ ’ is true or false depends on the context, in particular on the standards for knowledge that are assumed (Cohen 1999, DeRose 1992). The ideas of contextualists have been criticized, notably by Jason Stanley, who denies that knowledge is context sensitive, and argues that the gradability of knowledge is an illusion (Stanley 2005). In Stanley’s view, contextualists have valuable intuitions, but these are better accounted for by what he calls ‘interest-relative invariantism’, according to which knowledge is partly determined by a person’s practical interests. The greater a person’s actual investment in the truth of a proposition is, the stronger must be that person’s evidence if it is to be said that he *knows* the proposition. Rather than pointing to a gradability of knowledge, expressions like ‘knowing  $p$  better than  $q$ ’ serve as *pragmatic indicators*.

Recently the discussion about the gradability of knowledge has been given a new twist by Sarah Moss (2013). According to Moss, knowledge is indeed gradable for the simple reason that we can speak of *probabilistic knowledge*. She makes a case for what she calls ‘the radical thesis’, which involves the claim that degrees of probabilistic knowledge should be equated with degrees of belief, so that both degrees of knowledge and degrees of belief are measured in the same Bayesian way.

Like Moss we defend the gradability of knowledge. Unlike Moss, however, we are interested in a theory in which partial knowledge is *not* the same as partial belief. In order to distinguish our account from that of Moss, we speak of *partial knowledge* rather than of *probabilistic knowledge*, which is the term that Moss uses.

In this paper we develop the concept of partial knowledge and we explain how it can be measured in a way that is different from the measurement of partial belief. We take as our point of departure the work of Timothy Williamson (2014, 2000, 1990). This may sound ironical, since Williamson is one of the philosophers who are strongly opposed to the notion of graded knowledge. Yet we argue that our measure of partial

knowledge can be regarded as an extension of his approach, turning the latter into a limiting case. Moreover, we show that our measure has the advantage of avoiding certain counter-intuitive consequences that follow from Williamson's approach.

The paper consists of two parts. In the first part we summarize Williamson's argument. For expository reasons we concentrate on a certain example, viz. a particular clock that Williamson describes in Williamson 2014, but neither Williamson's reasoning nor our generalization of it is in any way specific to this particular template: both approaches are general and robust. We explain that Williamson's argument is a much-needed combination of possible world semantics and classical probability theory, and we stress the formal link that Williamson makes between knowledge and visual discriminability. Williamson defines knowledge in terms of possible worlds, and he takes the accessibility relation between worlds to be a relation of indiscriminability. We describe how Williamson derives from his argument a very surprising conclusion, namely that one might fully and categorically know a proposition  $p$ , while the probability that we know  $p$  on our evidence is arbitrarily close to zero.

Williamson stands by this conclusion, but in the second part of our paper we explain why we find it rather counterintuitive. We further argue that the conclusion arises from a rather artificial assumption, namely that visual discriminability is an all-or-nothing affair. The moment we give up this assumption the counterintuitive result disappears.

Like Williamson's method, ours too is a combination of possible world semantics and probability theory, but we will give a greater role to probability theory than Williamson does. This enables us to gradualize the accessibility relation in Williamson's possible world semantics, and thereby to introduce grades of visual indiscriminability and discriminability. We then explain how this notion of graded discriminability leads to a concept of graded knowledge: where Williamson says that full discrimination between worlds  $w$  and  $w'$  implies knowing that  $w$  and  $w'$  are distinct, we say that discrimination between  $w$  and  $w'$  to degree  $x$  implies knowing to degree  $x$  that  $w$  and  $w'$  are distinct. We call the concept of graded knowledge 'partial knowledge at a world'. The next step is to relate the concept of 'partial knowledge at a world' to that of 'partial knowledge on one's evidence'. We show that the probability of the latter is not arbitrarily close to zero, contrary to Williamson's finding, and that under reasonable conditions it is not even small.

The result of our efforts is a thoroughly probabilistic epistemology, in which 'knowledge' is no longer synonymous with 'categorical knowledge', and in which degrees belong as much to knowledge as they belong to belief.

## References

- Cohen, S. 1999. Contextualism, Skepticism, and the Structure of Reasons. In: J. Tomberlin (ed.), *Epistemology*. Philosophical Perspectives 13. Oxford: Blackwell, pp. 57-89.
- DeRose, K. 1992. Contextualism and Knowledge Attributions. *Philosophy and Phenomenological Research* 52, 913-929.
- Moss, S. 2013. Epistemology Formalized. *The Philosophical Review* 122, 1-43. Stanley, J. 2005. *Knowledge*

*and Practical Interests*. Oxford: Clarendon Press. Williamson, T., 1990. *Identity and Discrimination*. Oxford: Blackwell.

Williamson, T., 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press. Williamson, T. 2014. Very Improbable Knowing. *Erkenntnis* 79, 971-999.

---

*Confirmational Holism and Theory Choice: Arrow meets Duhem*

ELEONORA CRESTO

CONICET - UBA

eleonora.cresto@gmail.com

MIRANDA DEL CORRAL

CONICET

miranda.delcorral@gmail.com

DIEGO TAJER

CONICET - University of Buenos Aires

diegotajer@gmail.com

JUAN NASCIBENE

Torcuato Di Tella University

JNASCIBENE@GMAIL.COM

ALEJANDRO CASSINI

CONICET - UBA

ALEPAFRAC@YAHOO.COM.AR

---

In a recent paper Samir Okasha has suggested an interesting application of Arrow's impossibility theorem to theory choice: when epistemic virtues are interpreted as 'voters' in charge of ranking several competing theories, the final ordering is bound to coincide with the one proposed by one of the voters (the dictator), provided a number of seemingly reasonable conditions are in place ["Theory Choice and Social Choice: Kuhn versus Arrow". *Mind*, 2011]. Although Okasha seeks to offer a way out, it remains unclear whether the proposed solution is indeed feasible. In a similar spirit, Jacob Stegenga has shown an application of Arrow's theorem to the amalgamation of evidence; the voters here are the different sources of evidence ["An Impossibility Theorem for Amalgamating Evidence." *Synthese*, 2011]. As with Okasha's proposal, it is not clear how to avoid Arrow's pessimistic conclusion.

In this paper we develop a novel argument that purports to show that, in typical examples, Arrow's result does not obtain when dealing with the amalgamation of evidence. The reason is that, for most interesting cases, we cannot escape the well-known Duhem-Quine problem: we are seldom, if ever, able to assess isolated statements; rather, the evidence actually confirms (or disconfirms) a complex conjunction that includes various auxiliary hypotheses. We argue that confirmational holism forces us to restrict the domain of a reasonable choice function, thus violating one of Arrow's conditions. The upshot is that we are now able to see the Duhem-Quine problem under a different, positive light: to wit, we are able to interpret it as a phenomenon that makes theory choice possible in the first place, when there are at least three options on the table.

**Background considerations**

Notice that we will not be dealing with the more general problem of combining theoretical virtues, but with the problem of amalgamating evidence from different sources, so as to obtain a unified ranking of hypotheses by level of confirmation. In other words, our current enterprise can be thought of as aiming at the prior goal of finding the right input for just *one* of the traditional virtues – empirical adequacy.

There have been many suggestions concerning how to escape from Arrow within the realm of theory choice. For example, Okasha has pointed out that virtues (among them, hypothesis confirmation) should be measured by the same cardinal scale; there is no consensus, however, as to whether this is indeed a reasonable demand [see J. Stegenga, “Theory Choice and Social Choice: Okasha versus Sen”, *Mind*, forthcoming]; here we adopt a conservative stance and assume no more than a comparative notion of confirmation. On a different line, Michael Morreau has argued that many virtues provide a *rigid* way of ranking hypotheses, in the sense that their rankings could not have been different from what they actually are [“Theory Choice and Social Choice: Kuhn Vindicated”, *Mind*, forthcoming]. Hence the domain of the Arrowian function is not universal; Arrow’s result might not obtain once the domain is so restricted. It is not clear whether rigidity, in Morreau’s sense, also applies to empirical adequacy – he thinks it does not. In any case, when dealing with the amalgamation of evidence, there are independent reasons to believe the domain is actually restricted. We show that (i) the Duhem-Quine thesis forces a particular domain restriction, and (ii) the domain restriction so induced is Arrow-consistent.

**The main argument**

Let  $H_1$ ,  $H_2$  and  $H_3$  be rival hypotheses. Consider two experimental sources  $S_\alpha$  and  $S_\beta$  – our “voters”. In light of the Duhem-Quine problem, strictly speaking each  $S_i$  does not choose among  $H_1$ ,  $H_2$  and  $H_3$ , but among more complex packages containing auxiliary hypotheses (such as the conjunction of  $H_1$  and  $A_{1,\alpha}$ ). Say  $S_\alpha$  deals explicitly with

- (a)  $H_1 \& A_{1,\alpha}$
- (b)  $H_2 \& A_{2,\alpha}$ , and
- (c)  $H_3 \& A_{3,\alpha}$ .

whereas  $S_\beta$  deals with

- (d)  $H_1 \& A_{1,\beta}$ ,
- (e)  $H_2 \& A_{2,\beta}$ , and
- (f)  $H_3 \& A_{3,\beta}$ .

Let us grant that all six items are on the table for both voters. This is not a concession we make just for the sake of the argument. Suppose for a moment that  $S_\alpha$  is not an abstract entity, but a real person who actually performed experiments of a certain kind– say, type- $\alpha$  experiments. Moreover, she has heard that her

colleague (i.e.,  $S_\beta$ ) is running type- $\beta$  experiments. Although she ignores the results obtained by  $S_\beta$ , she does know that type- $\beta$  experiments presuppose in each case – for each rival  $H_i$  – certain definite auxiliary hypotheses. She is then asked to assess and rank  $H_1$ ,  $H_2$  and  $H_3$ , to the best of her knowledge. Given that both  $S_\alpha$  and  $S_\beta$  are very much aware of each other's research project  $S_\alpha$  feels that any serious assessment of the three rival hypotheses should take the *six* items into account simultaneously; she rightly expects  $S_\beta$  to do the same.

Notice, however, that  $S_\alpha$  has no information about the outcome of  $S_\beta$ 's experiments, and vice-versa. In the absence of such vital information, not all logically possible orderings make sense as inputs. From  $S_\alpha$ 's point of view, she should apply to (d), (e) and (f) what she knows about the way  $H_1$ ,  $H_2$  and  $H_3$  behaved in her own experimental setting (that is, (a), (b) and (c)).

There are other natural restrictions to consider. For example, it is reasonable for a source of evidence not to place packages involving auxiliary hypotheses from different sources higher than their own packages. Furthermore, we will also argue that it is reasonable to require voters to assemble hypotheses in clusters, where each cluster contains the packages corresponding to a single source of evidence.

Once these constraints are adopted, it can be proved that the resulting domain is *non-saturating*, in the sense of Kalai, Muller and Satterthwaite [“Social welfare functions when preferences are convex, strictly monotonic, and continuous”, 1979; see also M. Le Breton and J. A. Weymark, “Arrovian Social Choice Theory on Economic Domains”, 2011]. Moreover, we also show how to build aggregation rules satisfying Weak Pareto, Independence of Irrelevant Alternatives and Non- Dictatorship, thus proving that the domain is indeed Arrow consistent.

---

*A General Model of Diversity in Science*

RICO HAUSWALD

Dresden University of Technology

ricohauswald@gmx.de

---

Diversity is among the central issues in today's philosophy of science; it is at the core of debates about pluralism, dissent, justice, or the division of cognitive labor in science. However, illustrious as it may be, the concept is not yet well understood and remains fairly unclear. This paper aims to remedy this shortcoming by addressing conceptual questions like: "How can diversity be defined precisely?", "Which relations hold between its various scientifically relevant subtypes (social, methodological, theoretical, subject matter diversity, etc.)?", "(How) is it possible to quantify these subtypes?" Using the ecological notion of biodiversity as a comparison case, in the first part I develop a model of diversity in general. Here is the outline of this model (see figure 1). There is some field  $F$  (e.g., an ecosystem or a scientific community), which somehow contains a number of typical elements  $A_1 \dots A_n$  (e.g., species; or theories, subject matters, methods, social categories, propositions). Every  $A$ -element stands in some characteristic relation  $R$  to elements  $B_1 \dots B_n$  (species are instantiated by individual organisms; theories are pursued, subject matters studied, methods applied, social categories exemplified, and propositions believed or valued by scientists). One can now define, on the one hand, various types of diversity, depending on which  $A$ -elements and relations  $R$  are considered (one gets species, theoretical, subject matter, methodological, social, doxastic, and axiological diversity, respectively). On the other hand, one can define a couple of dimensions of diversity, the most important of which are richness (the number of  $A$ s in  $F$ ), evenness (the degree to which the  $B$ s are equally distributed among the  $A$ s), and dissimilarity (the average disparity between any two  $A$ s in  $F$ ). Applying mathematical tools such as (variants of) the Shannon-Wiener, the Simpson or the Gini-Simpson index, these dimensions can then be integrated into an overall measure of diversity. An important point to note is that  $R$  can be an injective or a non-injective relation, which has significant consequences on how the dimensions (esp. evenness) are to be applied. For example, if we consider species and individual organisms,  $R$  is injective, i.e. every individual instantiates exactly one species. By contrast, if we consider, for example, social categories (e.g., being female, being a Finn, being atheist, etc.),  $R$  is non-injective, i.e. a scientist can simultaneously be a member of multiple social categories (see  $B_5$  in figure 1).

In ecology, the application of this general model of diversity is relatively straightforward. But what about applying the model to scientific communities? The main part of the paper is devoted to answering this question. I proceed as follows: I apply the mentioned dimensions to the relevant types of diversity in science and explore how these diversity-types can be quantified. In particular, I correlate each type with each

dimension and discuss what it would mean to maximize and to minimize diversity on this dimension. In doing so, the crucial challenge is how to individuate the elements, since measuring the diversity of theories, methods, subject matters, etc. presupposes a clarification of how these entities can be counted (which is a necessary condition for applying the dimensions richness and evenness) and what it means that one of them is more similar to a second than to a third (which is a necessary condition for applying the dimension dissimilarity). The individuation of theories, methods, propositions, etc., are issues dealt with in various fields of philosophy, such as the philosophy of science, ontology, and semantics, so my strategy will be to briefly discuss the respective standard accounts and to examine the prospects for the desired quantitative approach. Regarding doxastic and axiological diversity (i.e., diversity of believed and valued propositions, respectively), I take it that propositions are sets of possible worlds (Lewis 1986), so that dissimilarity between them can be measured by extending the metric between worlds to a metric between sets of worlds (Ruspini 2008). Mathematically, this is possible by applying distance measures between subsets of metric spaces, such as the Hausdorff distance. Regarding subject matter diversity, the most promising strategy is to reduce subject matters to questions and, in accordance with the standard account in interrogative semantics (tracing back to Hamblin 1958), to analyze questions as sets of propositions, i.e. sets of sets of possible worlds. Dissimilarities between subject matters can then be quantified by using distance measures between sets of sets of a metric space. Regarding theoretical diversity, I draw on the structuralist account (Balzer/Moulines/Sneed 1987), according to which theories are hierarchical structures. One can then say that two theories are more similar the more they agree not only in their central, but also in their more peripheral parts. Regarding diversity of methods, the most promising strategy is to analyze a method as a certain type of action that aims at achieving knowledge. Ontologically, actions qua events are occurrents, i.e. concrete individuals. A type, in turn, can be analyzed as a set of individuals (Lewis 1986). Hence, similarity between methods (types of actions) can be analyzed in terms of distances between sets of individuals. Finally, in order to analyze social diversity, I draw on recent approaches in social ontology, according to which social categories can belong to either of three ontological types: plural individuals (qua social collectives), sets, and classes (Author 2014). In a final section, I sum up the account and give a brief outlook to possible applications to the debates about dissent, pluralism, objectivity, and justice in science. I emphasize the difference between maximizing and optimizing diversity. Whatever function or value one may attribute to diversity in science, in most cases achieving an optimal level is not tantamount to maximizing diversity.



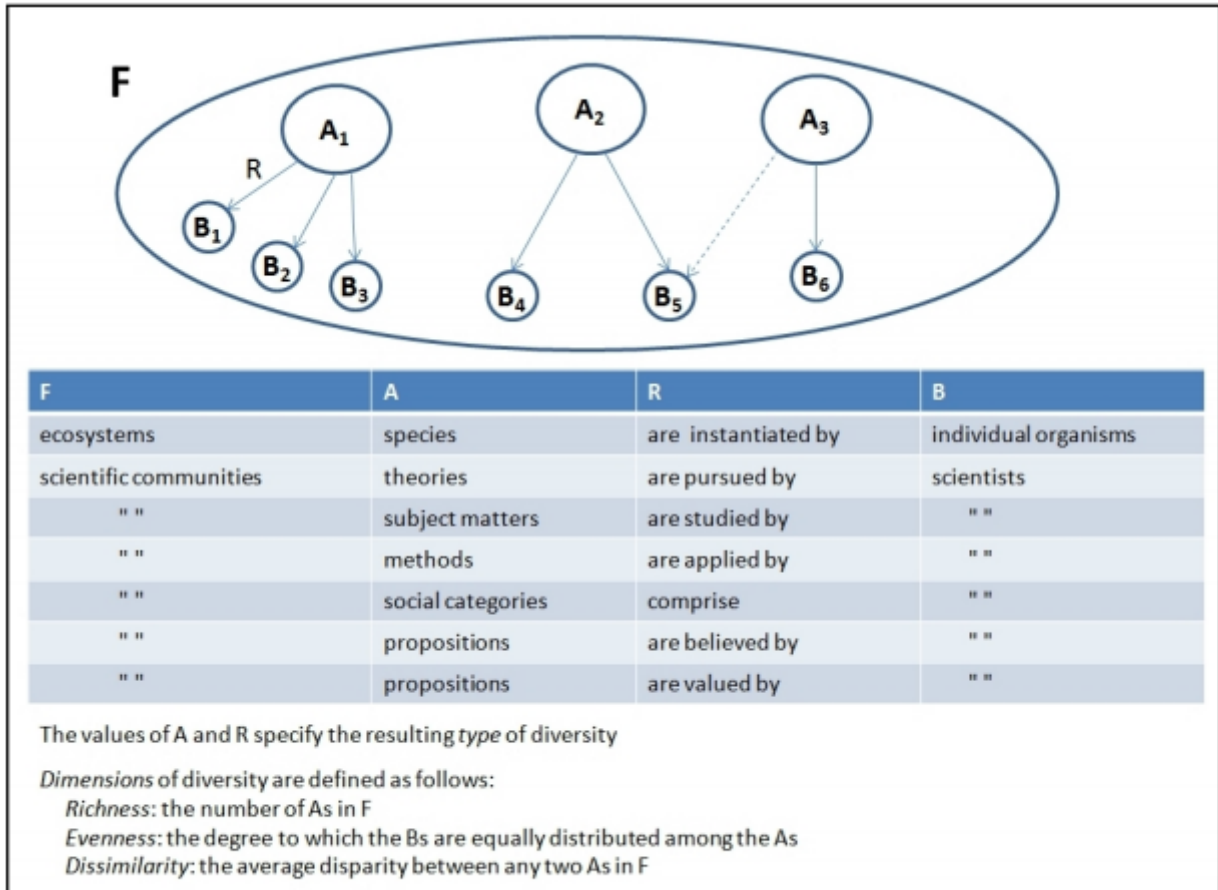


Figure 1

---

*Psychiatric Classification between Science and Practice*

ANKE BUETER

Leibniz University Hannover

anke.bueter@philos.uni-hannover.de

---

The current psychiatric classification (as exemplified by the DSM<sup>2</sup>) exhibits severe problems, and its recent revision, culminating in the DSM-5 (2013), has left many disappointed. On the one hand, there are controversial debates on the criteria for individual diagnoses and the question whether they pathologize normal feelings and behavior, for example, in the cases of ADHD or depression. On the other hand, there are also more general criticisms that question the overall system of psychiatric classification.

Currently, this classification is based on (statistically co-varying) observable symptoms of mental disorders. This “phenomenological” or “atheoretical” approach was introduced in 1980 with the 3rd revision of the DSM in order to improve the lack of interrater-reliability that characterized the DSM-I and –II diagnoses. A second goal was to eliminate the psychoanalytical theoretical background informing the first editions of the manual and to create a taxonomy that would be commonly acceptable and usable by all psychiatrists, regardless of their various theoretical orientations. While this “first revolution” managed to increase interrater-reliability and to establish one shared taxonomy, critiques argue that it has sacrificed validity by adopting a symptoms-based approach instead of aiming for a classification based on causal information. To be more specific, it is often argued that the heterogeneity of groups picked out by the DSM’s polythetic criteria, the excessive rates of comorbidity, and the lack of predictive success of the DSM diagnoses indicate a severe lack of validity. The main proposal for improving the situation is to change the classification from a phenomenology-based one to an etiology-based one that groups disorders according to our best scientific theories about their underlying causes. In my talk, I will discuss the question whether it is time for such an etiological “revolution” in psychiatric classification.

Advocates of this revolution often present it as a move forward towards a more scientific, evidence-based nosology. Even more cautious criticisms often seem to assume that the change towards a more valid etiological classification is only a matter of time, awaiting further research results. What I want to show is, first, that the question of the classificatory basis is not one that can be answered by empirical evidence alone.

---

<sup>2</sup> The DSM is the authoritative system in the US; most European countries use the ICD-classification of the WHO. Since both systems strive for compatibility and share the symptoms-based approach to classifying mental disorders, the same problems and arguments apply.

Instead, it requires judgments on what level of evidence is needed to justify changes as well as judgments on what kind of evidence is most important. Second, in making these judgments we need to weigh the needs of clinical practice and scientific research.

Regarding the question of how much evidence is enough to legitimate a more radical revision, it is important to note the DSM's multiple purposes. While it aims to be a suitable basis for research, it also thoroughly shapes psychiatric practice. Changes can be very consequential in that they affect patient's diagnoses and possibly treatment, might impact questions of reimbursement, and even change public views of mental disorder and normality. Therefore, before one starts a revolution, there should be solid evidence that this will improve the situation in terms of science as well as health care. What exactly that means (i.e. how much evidence is enough) is moreover not a purely scientific question but calls for value-judgments on the weighing of inductive risks and consequences of possible errors. Since for most mental disorders etiology is still unknown or highly contentious, it seems inappropriate to risk damage to patients by a premature change of the classification at this point of time.

A further problem is that criticizing the DSM-diagnoses as lacking validity often presumes that there was one common, clear concept of validity, which is actually not the case. Validity is supposed to mean something like "matching the reality" – but how one can measure that is not an easy question. Accordingly, psychiatry has to deal with different validators – for example, antecedent validators (e.g., diagnoses correlating with the exposure to certain risk factors or familial aggregation) or predictive validators (e.g., diagnoses correlating with treatment success or course of the disorder). This is a problem because empirical research shows that these validators do not always match up – one way of classifying mental disorders might perform better on predictive items, while an alternative outperforms it with regard to antecedent validators (one example here are different ways of subtyping schizophrenia). Thus, it becomes necessary to decide what kind of evidence is most important in such cases. The current DSM gives preference to predictive validators due to their importance for the use of the DSM in clinical practice. However, from the perspective of etiological research, antecedent validators seem more important.

In consequence, the needs of clinical practice and of scientific research do at present stand in conflict with each other. First, the importance of various validators can be judged differently from the perspective of science than from that of healthcare. Second, while clinical practice calls for a conservative approach and high standards of evidence before every radical change, the DSM does pose problems as a basis of scientific research (such as the heterogeneity of study populations or the comorbidity of different diagnoses). These are indeed severe and call for pluralistic explorations of possible alternatives and causal explanations. It is highly problematic that the DSM-categories are often mandatory in research (e.g., to get clinical studies approved, funded, and/or published), which presents a heuristic obstacle to ever achieve etiological knowledge that also allows for successful predictions – and that could serve as a solid enough basis for an etiological reorganization of the overall system.

The central difficulty in psychiatric classification is, accordingly, not just a lack of validity or a lack of evidence, but lies in integrating the different demands of research and practice. Therefore, I will argue, it is not time for an overall etiological revolution of psychiatric classification (yet) – instead it is time to distinguish between taxonomy and theory, which call for different strategies: conservatism versus pluralism.

---

*Getting Real about Words*

JASPER VAN DEN HERIK

Erasmus University Rotterdam

jasperherik@gmail.com

---

Many philosophers and cognitive scientists assume that *contentful* mental states are necessary for explaining *cognition*, which is then an umbrella term for all our activities that involve knowledge. Contentful states are understood as having truth-conditions or conditions of satisfaction and therefore *represent something in the world*. Call this *representationalism*. Over the past 25 years, there is a resurgence of the idea that representationalism is fundamentally misguided. Proponents of Enactivism, part of the embodied and embedded cognition paradigm, argue that cognition is best understood in terms of dynamic interaction between an organism and its environment, without any appeal to contentful mental states. Cognition is re-conceptualised as behaviour (activity with a purpose) of an organism in the world. The general idea is that recurring interactions shape an organism's structure, as well as the structure of the organism's surroundings, which in turn determines how the organism can and will interact with this world. The 'mentality-constituting' interactions are to be explained solely in terms of an organism's history of interactions.

At the same time, there is a wide-spread conviction that these non-representationalist explanations will break down in the face of typically *human* cognition, such as our linguistic abilities; to explain those, one still needs content. This is apparent even in the work of the self-proclaimed radical kinsmen of the Enactivist family, Hutto and Myin, who state that 'some cognitive activity – plausibly, that associated with and dependent upon the mastery of language – surely involves content' (Hutto & Myin 2013: xviii). This leads further to the view that 'there are *a variety of types of minds* (some are contentful, some not)' (our emphasis, Hutto, Kirchhof & Myin 2014: 1).

In this paper I argue that this distinction between contentless and contentful cognitive activities is problematic. My thesis is that *if* we can indeed explain basic, e.g. non-linguistic, cognition in contentless terms, we should also explain linguistic cognition in contentless terms, and give up the distinction between these two. Instead of representing the world, the primary function of language becomes the coordination of behaviour. Note that this does not entail that we cannot use language to *talk about the world* (as we often do), but rather that an explanation of these abilities should not be given in representational terms.

I focus on two closely related arguments. *First*, I argue that, although it might at first seem like our language consists of readily identifiable recurring units (i.e., words) which determine the contents expressed, this reification of linguistic symbols is indefensible (Love 2004). For in order to substantiate this claim, we need a method for deciding when a token-utterance is a repetition of an earlier one, that is, falls under the

same abstract type. Simply relying on the acoustic or written form of a word alone is clearly inadequate, as is shown for instance by homonyms: context is needed.

Although our language does consist of recurring forms, these forms can be *always* be used in novel ways (think for instance of malapropisms), in which case there can be no principled way to answer the question whether the word, as opposed to the form, is repeated, and thus which contents are expressed. A traditional answer to this problem of individuating words could be given by relying on mental representations: a repetition of the same linguistic form can than be seen as a repetition of the same word if it is an encoding of the same mental representation. However, this option is not open to the Enactivist, who denies these representations.

This leads me to my *second* argument: once we relinquish representations at the basic level of cognition, we lack the resources to explain *how* a person can come to know *that* a word stands for a certain worldly offering (Kravchenko 2007). The notion of a proposition is unfit for explaining linguistic cognition, because the naturalistically inclined Enactivist can also not rely on public *abstract objects*, which is what a proposition is. Fortunately, we can do without content: rather than claiming that knowing a word is knowing-that the word stands for a particular worldly offering, we can claim that knowing a word is knowing *how* to use it in concrete situations (Van Elk et. al 2010).

Together, these two considerations motivate a re-conceptualisation of linguistic symbols. In line with the Enactivist focus on cognition as activity with a purpose, we conceive of learning a language as learning to *do* things with words in a social context, with the aim of co-ordinating behaviour through the constraining of cognitive and interpersonal dynamics. Words are not representations, but mind-guiding pieces of voicing (Bottineau 2010). This entails that we cannot just look at the symbols by themselves, as we also have to take into account the dynamics that is constrained by linguistic activity. This means that linguistic activity becomes fully embodied, and that the strict dichotomy between basic and linguistic minds that Hutto and Myin propose has to be relinquished.

One might object that although this might explain the origins of language, our linguistic utterances are nevertheless representational in nature once we can speak, for we often talk *about* things. Here we have to make a distinction between the unfolding of linguistic behaviour and its products. An observer that focusses only on the products of linguistic behaviour, that is, the utterances, might conceive of these as instantiations of abstract types that refer to worldly objects. But this *observed representational function* need not have an *explanatory representational status* (Harvey 2015). If the two arguments given here are convincing, there are no abstract types in play in linguistic behaviour to which this explanatory representational status can be ascribed.

From an Enactivist point of view, linguistic cognition is therefore best explained without relying on the notion of content, which only comes into play in the descriptions of observers who abstract away from actual linguistic behaviour and focus merely on its products. Just as for more basic forms of cognition, these

representational presuppositions turn out to be untenable for explaining linguistic cognition within an Enactivist account.

### References

- Bottineau, D. (2010) Language and enaction. In J. Stewart, O. Gapenne, & E. Di Paolo (Eds.) *Enaction: Towards a New Paradigm for Cognitive Science*. Cambridge, MA: The MIT Press: 267–306.
- Elk, M. van, Slors, M., & Bekkering, H. (2010) Embodied language comprehension requires an enactivist paradigm of cognition. *Frontiers in Psychology*, 1(234): 1–9.
- Harvey, M. (2015) Content in languaging: why radical enactivism is incompatible with representational theories of language. *Language Sciences*, 48: 90–129
- Hutto, D.D., & Myin, E. (2013) *Radicalizing Enactivism – Basic Minds without Content*. Cambridge, Ma.: The MIT Press.
- Hutto, D.D., Kirchhof, M.D., & Myin, E. (2014). Extensive enactivism: why keep it all in? *Frontiers in Human Neuroscience*, 8(706).
- Kravchenko, A. V. (2007). Essential properties of language, or, why language is not a code. *Language Sciences*, 29: 650-671.
- Love, N. (2004) Cognition and the language myth. *Language Sciences*, 26: 525-544.

---

*Are Causal Accounts of Explanation always Useful? In the Case of Personality Trait Explanation they are Probably Not*

LILIA GUROVA

New Bulgarian University

[lilia.gurova@gmail.com](mailto:lilia.gurova@gmail.com)

---

One reason why people prefer causal explanations, both in everyday life and in science, is that causal explanations are useful as they provide information how to eventually get control over the explained phenomena (Woodward, 2014). But are causal explanations always useful? And aren't there non-causal accounts which might do a better job in some cases? In this paper I argue that in the case of personality trait explanations (a) the causal accounts face serious conceptual problems, and (b) a non-causal account is possible, which better grasps the explanatory virtues of traits and which is more in tune with the way trait explanations are used in practice.

The attacks on personality trait explanations in psychology have a famous history (Boag, 2011) where the critiques vary from claims that traits are a merely descriptive tool to blames that trait explanations are circular. Although these critiques are at odds with the increasing uses of trait explanations both in psychology and beyond, the controversy over personality traits persists because of the conceptual confusions surrounding the most popular account of traits as tendencies or dispositions which are causally connected to particular patterns of behavior (McCrae & Costa, 1995). The confusions come to light when one faces the question in what sense a tendency or a disposition 'causes' events which are instances (or parts) of the same tendency or disposition. On the standard construal of the cause-effect relation, the cause should precede its effect. The whole-parts relation, however, is different: the whole cannot precede its parts. Once this difference is taken seriously, it becomes clear why any attempt to construe tendencies as causes of the events which are part of them is conceptually confusing. The proponents of the causal account of personality trait explanations try to avoid this confusion by distinguishing between traits as patterns of observed behavior and traits as underlying tendencies that cause the observed patterns. Those who support this distinction (e.g. McCrae and Costa) insist that insofar as the observed patterns are not anymore instances or parts of the underlying tendencies, there is no problem to say that the latter causally explain the former. This conceptual maneuver, however, is not applicable to all trait explanations. There are examples such as the hypothesis explaining the association between higher paternal age and the risk to have an autistic child by reference to specific personality traits of the fathers (Puleo et al., 2008), where it is the observed phenotypic expression of the personality traits that does the explanatory job rather than any assumed underlying causes of this phenotypic expression.



What might stand as a viable alternative to the causal construal of trait explanations? To answer this question, we should find out first what personality trait explanations are good for. A glance through the literature reporting such explanations reveals that they are appreciated for two main reasons: (1) for ruling out alternative explanations; and (2) for allowing to draw additional inferences about the explained phenomenon. Thus on the example of the association between higher paternal age and the risk to have an autistic child we see that: (1) the personality trait explanation of this association rules out the hypothesis that the accumulation of *de novo* mutations is responsible for the higher risk of having a child with autism in later age, and (2) the personality trait explanation implies the testable prediction that the risk level to have an autistic child for men exhibiting personality traits which are part of the so-called ‘broader autism phenotype’ does not correlate with their paternal age. It can be shown on this and similar examples as well that the elimination of alternative explanations is a consequence of the increased number of predictions which are allowed by a good explanation. Personality trait explanations are appreciated exactly for these reasons by most authors referring to such explanations including McCrae and Costa (1995) who admitted that trait explanations are useful because they allow “inferences which go beyond the observed data” (McCrae & Costa, 1995, p.243).

The view that a mark of a good explanation is the increased number of inferences about the explained phenomenon which it provides is part of the following two-statement account of scientific explanation:

A: A good explanation increases our understanding of the explained phenomenon.

B: A mark of increased understanding is the increased number of inferences that could be drawn about the explained phenomenon on the basis of the proposed explanation.

This account does not directly confront with the causal account of explanation. It rather subsumes it, insofar as on the proposed account, causal explanations increase our understanding of the explained phenomena insofar as they allow for inferences “going beyond the observed data” such as “if the cause were not present, the effect would not be present, as well” (on the counterfactual account of causation), or “any time the cause is present, the effect will follow it” (on the classical ‘constant conjunction’ account of causation), or “it is more probable the effect to appear if the cause is present” (on the probabilistic account of causation), etc.

The main virtue of the proposed inferentialist account of explanation is that it makes possible to assess the explanatory value of a particular explanation without reducing it to a causal explanation. Thus the inferentialist account helps to avoid any hopeless discussions like those about whether personality traits causally affect behavior and how they eventually do that.

---

*A Naturalistic Theory of Perceptual Content*

MARC ARTIGA

MCMP, Ludwig-Maximilians-Universität München

marc.artiga@gmail.com

---

Perceptual states represent the world as being in a certain way. Although not completely uncontroversial, this is a common assumption underlying many contemporary discussions in philosophy of perception and philosophy of mind (Siegel, 2010). However, a standing difficulty in defending or criticizing this claim is that we currently lack a general theory of what makes certain states representational or how their content is determined. If such an account could be provided, theoretical debates on the representational nature and content of perceptual states would stand on a much firmer ground.

Similarly, cognitive science is pervaded with representational talk (Eliasmith, 2000; Kandel et al., 2000). Among neuroscientists, for example, it is extremely common to claim that neuronal activity in a certain region of the brain represents particular features. Indeed, some people argue that an appeal to representations might indeed be inevitable in the explanation of behavior and off-line capacities (Sprevak, 2013). Nonetheless, we do not have yet a satisfactory theory explaining what distinguishes representations from non-representations. This account would help us very much understanding and justifying the scientific use of representational notions.

The goal of this paper is to sketch such a theory. In particular, I will argue that a particular approach to mental content (SR-Teleosemantics) can explain why many states produced in the perceptual system are representations and might provide the resources for determining their representational content. As we will see, however, the application of these theories to perceptual states is not straightforward. In particular, two pressing questions need to be addressed: whether sender-receiver systems can be identified in the brain and whether the simple sender-receiver framework assumed by teleosemantics can model the complex structures found in perceptual systems. Showing how these challenges can be met will not only support representational theories of perception, but it will also help teleosemantic theories to move forward.

Accordingly, the paper is structured in two main sections. First of all, I briefly outline the teleosemantic framework I rely on. Secondly, I show how it can be used to naturalize perceptual representation. To do that, I first address certain problems concerning the application of teleosemantic ideas to neural structures, and afterwards exemplify this model with some examples. Let me briefly go over them.

As I said, the teleosemantic framework relies on the biological notions of function and sender-receiver structure. On the one hand, they adopt an etiological theory of functions, according to which functions are selected effects 1 (Neander, 1991). More precisely, a trait's function is the effect that explains why past traits

of the same kind have been selected for by natural selection (Wright, 1973; Millikan, 1984). For instance, the function of kidneys is to filter wastes from blood because this is the effect that explains why kidneys were selected for. The second important notion is that of a sender-receiver structure. In short, a sender-receiver structure is constituted by two systems: a sender, which takes some input and produces a state, and the receiver, which takes this state as input and produces an effect as output. This sender receiver structure is widely used in game-theory and abstract models of signaling.

The key move of teleosemantics is to add the notion of etiological function to the sender-receiver framework in order to provide a naturalistic account of representations in the biological world. In a nutshell, the idea is that a representational system is composed of two different kinds of mechanisms (sender and receiver), which are endowed with certain etiological functions. The function of the sender (or 'producer system') is to produce a representation R when certain state of affairs obtains. The function of the receiver (or 'consumer system') is to act in a certain way (e.g. produce some behavior) when representation R is produced. Crucially, representations are states that stand between a sender and the receiver that possess these functions (Millikan, 1984; Godfrey-Smith, 1996, 2006; Shea, 2007. Representational content is also determined by the function of producer and consumer systems.

This teleosemantic framework has been mainly applied to certain areas of biology (e.g. animal communication) and psychology (e.g. concepts). However, in this essay I will concentrate on perceptual systems for three main reasons. First of all, there is a widespread intuition that perceptual representations are somehow basic, in the sense that many other representational capacities are grounded on perceptual states. To the extent that this claim is true, a naturalization of perceptual representations will have important consequences for many other projects. Secondly, in contrast to other cognitive capacities (e.g. Millikan, 2000), few attempts have been made to naturalize this sort of content. Finally, as I will show, perceptual systems seem to satisfy many of the conditions established by teleosemantics. So if one is interested in a naturalistic theory of mental representation, that seems to be one of the best places to start with.

Yet there are at least two preliminary problems with an application of this framework to perceptual states. First of all, are there sender-receiver structures in the brain? A quick look at current neuroscientific research suggests that sender-receiver structures are not missing but pervasive. Thus, a potential worry is that there might just be too many systems in the brain that could qualify as sender-receiver structures (Cao, 2012; Godfrey-Smith, 2013, p. 52). With respect to this worry, I concede there is a huge amount of sender-receiver structures in the brain, but I reply that there seems to be an equally large amount of neural representations. Furthermore, I argue that there are specific areas of neuroscience devoted to discover them.

There is however a second and more pressing difficulty in applying the teleosemantic framework to perceptual systems. The simple sender-receiver model involves a single sender and a receiver, in which the former sends a signal to the latter. Yet there are at least three ways in which perceptual systems fail to fit this simplistic model. First, in many cases a single sender sends signals to multiple receivers and multiple senders

are linked to the same receiver. Secondly, one can easily find sender-receiver structures forming a sequence, in which the very same mechanism plays the role of receiver with respect to some mechanism downstream and the role of sender with respect to another upstream. Thirdly, connections often go in both directions. For instance, it is well known that many areas in early vision receive information from high-order visual areas (Goldstein, 2013, p.24). Prima facie, it is not obvious that these three structures can be accommodated by the unidirectional and simple model suggested by teleosemantics. As a response, I argue that recent developments in game-theory show that these different structures can be modeled within the sender-receiver framework (Skyrms, 2009, 2010, 2014). Finally, I illustrate my proposal with some examples from vision science.

## Symposia & Contributed Papers III

### Life as Process: Reconceptualizing the Organism

Symposium

Organizer: John Dupré

Chair: Marie I. Kaiser

Room 5D, Thursday 13:00 – 15:00

---

#### *Introduction*

JOHN DUPRÉ

University of Exeter

j.a.dupre@ex.ac.uk

---

#### *Metabolic Identity: Approaches to the Particularity of Life from a Processual Perspective*

ANNE SOPHIE MEINCKE

University of Exeter

a.s.spann@exeter.ac.uk

---

#### *A Process-Based Understanding of Biological Boundaries*

STEPHAN GUTTINGER

University of Exeter

s.guttinger@exeter.ac.uk

---

#### *Reconceptualizing the Organism: From Complex Machine to Flowing Stream*

DANIEL J. NICHOLSON

University of Exeter

dan.j.nicholson@gmail.com

---

### General Description

This symposium proposal grows out of the ERC-funded project, A Process Ontology for Contemporary Biology, directed by John Dupré at the University of Exeter. This project explores the implications of the intuition that life is better understood in terms of an ontology of processes than in terms of the ontology of things, or material substances, that has predominated in most of Western metaphysics and philosophy of

science. The symposium will focus more specifically on processual approaches to the concept of an organism, and on the role of metabolism in defining the identity and boundaries of living systems.

Reflection on the last hundred years of physics might naturally lead one to suppose that the ancient debate as to whether the world was ultimately composed of things or processes had been resolved in favour of the latter. Quantum physics seems to constitute a decisive rejection of the atomism at the root of traditional thing, or substance, ontologies. However for macroscopic entities, specifically those that are the subject matter of biology, the basic parameters for most subsequent thought were laid down by Aristotle, and the Aristotelian tradition is firmly on the side of things rather than processes.

Organisms were for Aristotle the standard exemplars of what he called primary substances. A primary substance was a thing of a particular kind, for example a cat. Particular cats, such as Tabitha, are assumed to have sharp boundaries: everything is either part of Tabitha or not. Subsequent substance-based thought about the living world has seen substances as forming a hierarchy. Cats are composed of organs, organs are composed of cells, etc., and cats might themselves make up larger entities such as species.

Against the tradition of substance thinking an important minority of philosophers have defended more processual views of reality. These include, arguably, such major figures as Leibniz and Hegel, but the definitive exposition of a processual metaphysics is in the work of Alfred North Whitehead, for example in his classic, *Process and Reality*. Nevertheless, in the metaphysics and philosophy of science this tradition has had very little impact in recent years.

One common statement of the difference between a substance- and a process-centred metaphysics is that the former sees being as basic, and the latter prioritises becoming. For Aristotle substance was closely connected to essence: to be a cat was to possess the essence of catness, the property or properties that made that entity a cat. For a processualist, on the other hand, a cat is a pathway from zygote to kitten to mature animal to death. Any time slice of this sequence can only be properly understood as part of this longer process, and in terms of both its preceding history and its possible future trajectory. No property need be common to every stage of a cat life cycle; it is rather the relations between stages in its history that constitute it as a cat.

For a substance theorist, what require explanation are the changes that occur to an entity, and the conditions under which an entity can remain the same thing through change. For a process theorist, on the contrary, the central questions concern rather how a combination of processes can maintain the appearance of stability and persistence in an entity that is fundamentally only a temporary eddy in a flux of change.

I noted that Aristotelian substance metaphysics assumes that things have determinate boundaries. It is no doubt possible to modify this assumption and allow for some fuzziness about exactly where an entity begins and ends. But—and this is one of the central reasons for preferring to think of life in processual terms—it increasingly appears that the attempt to identify even fuzzy boundaries as objective features of the living world may be fundamentally misguided. The omnipresence of symbiosis begins to suggest that isolating

traditional organisms from the complex set of relations that are required for their persistence may be at best a pragmatic exercise for particular purposes. 90% of the cells in a human body, for example, are microbial symbionts, huge numbers of which are required for healthy functioning, and others of which have more or less cooperative relations with the whole system or important subsystems. And humans are probably as close as one could get to a traditional biological object with sharp boundaries. In general, it seems increasingly plausible that there is no unique way of distinguishing biological things from the multiple processes by which they are sustained, which supports the hypothesis that the ontology of biological things is less fundamental than that of living processes.

The shift to a processual perspective has potentially profound implications for our understanding of life and the life sciences. One set of issues that seems much better suited to a processual understanding is development. The neo-preformationist view of development as a determinate sequence of events driven by the genome, an object with almost magical capacities to direct ontogeny has been widely rejected, at least by philosophers. The alternative epigenesist interpretation is best understood as fully processual, and requires a better understanding of the nature of processes with reliable dynamic trajectories.

Process ontologists have also questioned the very basic dichotomy between structure and function that underlies much of the investigation of living systems. It is common to think of biological objects having particular structures that enable them to perform particular functions. But if these 'objects' are in fact constantly fluid and evolving processes, this perspective can be misleading. Structure and function are intertwined aspects of process. A good example of a phenomenon that invites such a view is plant development. The growing meristem of a plant is typically an opportunistic growth process capable of producing a variety of structures—leaves, flowers, roots—in response to the environment it encounters. Even the attempt to distinguish sharply between these traditional structures is often problematic. As a committed pluralist, I don't want to assert dogmatically that the world is composed of processes not things. However, I am confident that a process ontology provides a generally more illuminating view of the living world, and certainly the same seems true of contemporary physics. At the very least the question whether things or processes provide a better framework for interpreting science is one that should be a central concern for everyone interested in the metaphysics of science. And perhaps it is a question that matters beyond just getting to the truth. As Whitehead wrote in *Process and Reality*: "There is urgency in coming to see the world as a web of interrelated processes of which we are integral parts, so that all of our choices and actions have consequences for the world around us."

This symposium will focus specifically on the nature of the organism as process. An earlier paper by Dupré and Maureen O'Malley ("The Varieties of Living Things: Life at the Intersection of Linage and Metabolism", *Philosophy & Theory in Biology*, 1, 2009) stressed the centrality of metabolism as encompassing the processes that connect the various entities that interact in sustaining living systems generally, and more specifically the entities traditionally considered as organisms. In his Introductory talk, Dupré will revisit the central idea of

that paper, interpreting it in a more explicitly processual way. (Had it been written today the word “things” in the title would have been replaced by “systems” or “processes”.)

In the following presentations Spann will sketch a process philosophical account of the identity of organisms; Guttinger will explore the resources of a process-based account for identifying the boundaries between living entities: and Nicholson will conclude by contrasting the processual view of the organism with the currently popular mechanistic perspective.

## **Abstracts**

### **1. Anne Sophie Meincke: *Metabolic Identity: Approaches to the Particularity of Life from a Processual Perspective.***

Organisms are material beings just like waves or wardrobes. They consist of matter. Yet, they are different. Organisms are alive; they are living material beings as distinguished from non-living ones.

What this difference amounts to, what it implies and physically requires, has been as extensively debated amongst philosophers as well as biologists. In particular, more recent attempts to overcome traditional Cartesian mind/body dualism have made organisms attract increasing philosophical interest.

In this paper I will present a philosophical approach to the particularity of living material beings which focuses on the special kind of identity they exhibit. In this regard, metabolism will turn out to be crucial both by revealing the organism as to some extent independent from the matter that composes it, and at the same time by showing the organism’s diachronic identity to be irreducibly processual. Instead of the organism being a function of the changing matter, metabolism rather is a function of the organism, namely such that the organism never actually coincides with its material constitution at any particular point in time. An organism’s identity is maintained by that organism itself and is real only as the process of its maintenance.

I shall discuss to what extent a processual notion of biological identity of the kind proposed might provide the resources for defending the view that organisms as living beings are indeed distinguished from waves or wardrobes while nevertheless being fundamentally material, and what this might tell us about the particular character of life in general. I will also briefly touch on related issues including emergence and top-down causality.

### **2. Stephan Guttinger: *A process-based understanding of biological boundaries.***

Metabolic activity is often seen as one of the defining features of living entities. However, as Dupré and O'Malley (2009) point out, metabolism is a collaborative activity, which makes the idea of autonomous metabolic units (and any definition of living entities based on it) difficult to defend.

Their focus on the collaborative nature of metabolic processes leads Dupré and O'Malley to propose a definition of living entities as the intersection of lineage-forming entities that collaborate in metabolism. Under this definition viruses or even plasmids can be seen as living entities, at least for as long as they



intersect and collaborate with other lineage-forming (and metabolically active) systems. This account, however, threatens to undermine the boundaries we intuitively draw around living systems unless we simultaneously adjust our idea of what a boundary is.

I will begin by discussing in more detail Dupré's and O'Malley's idea of metabolism as collaborative activity and the definition of living entities they base on it. I will show that their approach leads to counterintuitive examples of what counts as a living entity, considering for example the elaborate in vitro systems used in current research. The key issue for their account is not the idea of collaboration itself but the concept of a boundary it builds on.

I will then argue that this problem can be overcome if we switch from a static to a more dynamic and process-based understanding of what a boundary is. To illustrate the benefits of such a switch, I will analyse the example of endosymbiont-to-organelle transformation in microorganisms, in particular the case of *Paulinella Chromatophora*. This processual understanding of a boundary will allow the collaborative view of living entities to deal with problematic and counter-intuitive examples.

### **3. Daniel J. Nicholson: *Reconceptualizing the Organism: From Complex Machine to Flowing Stream.***

Since Descartes, organisms have been conceptualized as ingenious pieces of machinery, different from artefacts only by their superior complexity. The machine metaphor has proven so heuristically useful that it is often simply taken for granted. However, it is becoming increasingly apparent that organisms differ from machines in fundamental ways, and that the machine metaphor is impeding rather than enabling further understanding of living systems.

This paper examines an alternative metaphorical conception of organisms that attempts to capture their thermodynamic nature as open systems far from equilibrium. This conception, first proposed by Heraclitus, was advocated by Locke, Cuvier, Hegel, and several early twentieth-century organicists, like Bertalanffy, who dubbed it 'the stream of life'.

Heraclitus asserted that one cannot step into the same river twice because water is continuously flowing. Like the river, ever changing in its waves yet persisting in its flow, an organism only appears to be constant and invariable, but in reality it is the manifestation of an enduring current. Living beings are never truly in being as they are always becoming—they are the expression of an ongoing stream of matter and energy which passes through them and at the same time constitutes them.

This paper explores the meta-theoretical consequences of the stream of life. For example, the traditional divide between anatomy and physiology was based on the mechanical assumption that structure and function are distinct properties that can be studied independently. Indeed, a mechanic can examine the structure of a machine at rest, or study its function while operating. However, organisms are not objects—like machines—but manifestations of processes. What this implies is that considerations pertaining to

structure and function are actually epistemic abstractions of a processual reality. The paper considers this and other implications of the stream of life metaphor for biological ontology.

**Levels, Computation, and Causation in****Symposium****Cognitive Neuroscience**

Organizer: Jens Harbecke, Vera Hoffmann-Kolss, Marcin Miłkowski &amp; Oron Shagrir

Chair: Alexander Gebharter

Room 5E, Thursday 13:00 – 15:00

---

*Level Distinctions and Methods for Constitutive Inference in Cognitive Neuroscience*

JENS HARBECKE

Witten/Herdecke University

jens.harbecke@uni-wh.de

---

*Computations, Mechanisms, and the Role of the Environment*

ORON SHAGRIR

Hebrew University Jerusalem

oron.shagrir@gmail.com

---

*Causal Relations in Mechanistic Explanations*

VERA HOFFMANN-KOLSS

University of Cologne

vera.hoffmann-kolss@uni-koeln.de

---

*The False Dichotomy between Causal Realization and Semantic Computation*

MARCIN MIŁKOWSKI

Institute of Philosophy and Sociology of the Polish Academy of Sciences

marcin.milkowski@gmail.com

---

**General Description**

In recent years, the philosophy of cognitive neuroscience has become a widely debated and increasingly important area of contemporary philosophy of science. A question that has attracted much attention is how we should understand the various kinds of level distinctions occurring in theories of neuroscientific explanation (Kaplan forthcoming; Egan 2014; Shagrir and Bechtel 2014). One famous example is the 3-level framework proposed by Marr (1982); this framework has recently been re-introduced to explicate the

Bayesian computational approach to cognition (Chater et al. 2006; Griffiths et al. 2008; Tenenbaum et al. 2006). Another and very different framework is the one identified by Churchland and Sejnowski (1992), who talk about levels of organization. Yet a third proposal, closer to the one put forward by Churchland and Sejnowski, has been advanced more recently by Craver (2007) who suggests that much research in cognitive neuroscience is about levels of mechanism. Further proposals about the role and nature of levels in the sciences have been discussed, for instance, by Lycan (1987), Bechtel (1994) and Wimsatt (1994).

The aim of this symposium is to explore the various level distinctions in cognitive neuroscience with a particular focus on computation and causality within mechanistic frameworks. A comprehensive theory of the various levels taxonomies and a general clarification of the different senses in which the term “level” is used within these are essential for a comprehensive understanding of the conceptual underpinnings of contemporary cognitive neuroscience. Furthermore, while computation and causation present somewhat different difficulties to the mechanistic framework, they are also linked through the question of whether computational properties understood as higher-level properties are causally relevant to the cognitive task. More specifically, the symposium will tackle the following issues:

(1) It is not clear whether the notions of level employed by cognitive scientists and philosophers in the different frameworks are compatible. For instance, it is questionable whether what Churchland and Sejnowski (1992) have described as “levels of organization” are identical to Craver’s (2007) levels of mechanism. Moreover, it is an open question how these level distinctions relate to levels of processing also described by Churchland and Sejnowski, as well as to the mereological and nomological level distinctions by Lycan, Bechtel, and Wimsatt. A general question concerning all of these frameworks targets the inference rules on the basis of which levels are distinguished in the first place.

(2) Even within one framework, it is not always clear what the relations between the phenomena at different levels are. For instance, it is unclear whether computational levels as proposed by Marr can be described in terms of part-whole relations or in terms of mutual manipulability. For one thing, it is dubious what a concrete manipulation of a computational function understood as an abstract mathematical entity could be. Moreover, it is an open question whether the inter-level relations occurring in computational explanations can be analyzed in terms of notions developed in traditional philosophy of mind, i.e. supervenience, realization, integration, and reduction (see e.g. Beckermann 1992; Esfeld 2005; Kim 1998; Marras 1993). Resolving this lack of clarity is crucial for making sense of the role of computational explanations and their relations to mechanistic explanations.

(3) There is a puzzle about the role that causal relations play in inter-level explanations. The issue of causality is important because many well-established neuroscientific explanations of cognitive phenomena presuppose that higher-level phenomena can have distinct causal profiles. For instance, the behavior of a rat navigating through a Morris water maze can be causally explained in terms of higher-level phenomena, such as environmental cues. In certain contexts, such an explanation will be adequate without having to go down

to the molecular level. However, this view is challenged by the consideration that higher-level entities are causally inert at least to the extent that they add nothing above and beyond the causal powers of lower-level entities. If that were the case, higher-level explanations would turn out to be meaningless, which would be a devastating result for almost all hierarchical level explanations. Recently, several philosophers have claimed that this so-called *causal-exclusion problem* can be avoided if causation is understood in terms of Woodward's interventionist approach (Woodward 2003) or causal modeling approaches as developed by Pearl, Halpern and Hitchcock (Halpern and Pearl 2005; Hitchcock 2001, 2007; Pearl 2000). However, it is still an open question whether causal modeling approaches are indeed capable of dispelling causal exclusion worries (Baumgartner 2009; Hoffmann-Kolss 2014; Raatikainen 2010; Woodward 2008).

It has been argued that there is a tension between the semantic characterization of neural computation and the causal account of computation. Just because the heuristic role of semantic entities in cognitive science is hard to deny (Bechtel 2014), it might be thought that the causal account is descriptively inadequate for current scientific practice. However, the causal role of semantics and the causal account of neural computation are not mutually exclusive, and they both have important explanatory, descriptive, and heuristic roles. In particular, semantic notions usually require rich interactions with the environment and appropriate internal orchestration of the mechanism; purely computational modeling, on the other hand, is usually limited to the internal functioning of a mechanism, and there are complex inter-level and intra-level relationships between computational, semantic, and, more broadly speaking, causal posits in explanatory models in neuroscience.

## Abstracts

### 1. *Jens Harbecke: Level Distinctions and Methods for Constitutive Inference in Cognitive Neuroscience*

According to the mechanistic approach, explanation in neuroscience essentially requires the identification of the mechanisms underlying a to-be-explained phenomenon (cf. Bechtel and Richardson 1993; Machamer et al. 2000; Craver 2002, 2007). In the debate about this general model, philosophers of neuroscience have often distinguished between what Machamer et al. (2000, 3) have called "being productive of regular changes" and what Bechtel and Abrahamsen (2005, 423) have characterized as being "responsible for one or more phenomena" from causing a phenomenon or event. The non-causal and synchronous relation between the phenomena and their mechanisms is now usually referred to as "constitution", "composition", or "constitutive relevance". As mechanisms can sometimes become explanantia themselves in the sense that their occurrence is explained by other lower-level mechanisms constituting them, the idea of a distinction of mechanistic levels has entered the picture.

## **2. Oron Shagrir: *Computations, Mechanisms, and the Role of the Environment***

In *Vision*, Marr (1982) famously advances a three-level approach to the study of visual processes. The computational level specifies what is being computed and why. The algorithmic level characterizes the system of representations that is being used, and the algorithm for the transformation from input to output. The implementation level specifies how the representations and the algorithm are physically realized. Marr's conceptual framework has led to extensive debates within philosophy about the nature of the top computational-level theories, the relations between the three levels (including whether they should be kept distinct and whether we need them all), and about the direction of investigation (e.g., top-down, from the computational to the implementation, or bottom-up); see, e.g., Egan (1992, 1995), Polger (2004), Shapiro (1993, 1997).

In this talk, Shagrir's aim is to develop an account of the relations between the computational level and levels of mechanisms that takes seriously the role of the environment in Marr's notion of computational analysis. On this more "ecological approach" (Sterelny 1990; Shagrir 2010; Shagrir and Bechtel 2014) the computational analysis has two tasks. One is to characterize the mathematical function(s) that the cognitive system computes (Marr calls it the "what" aspect of the analysis). Thus in the case of edge-detection the early visual processes (roughly) compute the zero-crossing of second-derivatives of the retinal images. Another task is to relate the computed function with the physical environment, by showing that the computed function mirrors a pertinent relation in the visual field (Marr calls it the "why" aspect of the analysis). In the case of edge-detection the pertinent (mirrored) relation is sharp changes in the reflection function (occurring along physical edges such as object boundaries).

More, specifically, the aim will be to examine the extent to which this ecological understanding of the computational level is applicable in computational cognitive neuroscience in general, including Bayesian models of cognition; initial results indicate that it is applicable (Shagrir 2012; Bechtel and Shagrir 2014). This hypothesis will be explored systematically by focusing on Bayesian models of cognition, especially those pertaining to inferences about hidden causes (e.g. Pearl 2000; Griffiths and Tenenbaum 2007a, 2007b).

## **3. Vera Hoffmann-Kolss: *Causal Relations in Mechanistic Explanations***

The mechanistic account of explanation presupposes that mechanisms can contain higher-level and even inter-level causal relations. In this context, causal relations are usually analyzed in terms of an interventionist account. Hoffmann-Kolss argues in her paper that applying interventionism to causal relations at different levels raises a hitherto unnoticed problem.

Woodward's interventionist account of causation (Woodward 2003), is currently considered one of the most promising approaches to describing causal relations occurring at different levels. However, one consequence of this approach is that there may be several causal explanations of the same phenomenon. Therefore, Woodward suggests that the interventionist account should be supplemented with the

proportionality constraint, according to which a causal explanation should provide as much relevant detail as necessary, but no irrelevant detail (Woodward 2008, 2010). If there are several causal explanations of the same phenomenon, the proportionality constraint should be used to decide which of them is most appropriate.

The aim of this paper is to show that if complex causal structures come into play, the proportionality constraint is inapplicable in a number of cases, and is, hence, not generally applicable within the interventionist framework. It follows that causal relations at different levels raise more serious problems within the interventionist framework than is commonly assumed and that it is at least questionable whether the causal relations occurring in mechanistic explanations can be analyzed in interventionist terms.

#### **4. Marcin Miłkowski: *The False Dichotomy between Causal Realization and Semantic Computation***

One of the important questions in philosophy of cognitive neuroscience is whether computations in cognitive neuroscience are best understood in terms of semantic relationships, as argued by Cantwell Smith (2002) and recently by Rescorla (2012a, 2012b, 2013), or causally, as defended by Chalmers (2011), Piccinini (2007) or Miłkowski (2013). It can hardly be denied that semantic relationships have an important role in neuroscientific research, and they obviously have had important heuristic value, for example in research on cognitive maps in rodents (Bechtel, 2014). This, however, need not mean that the notion of representation can be reduced to the one of computation. Purely computational characterizations of phenomena do not allow one to talk of causal interactions with the environment and satisfaction conditions of content, contra Rescorla, as there are clear counter- examples to his claims that the role of syntactic tokens such as '10' depends on the social interpretation of these tokens as decimal, binary, or hexadecimal. It would be premature, however, to conclude that satisfaction conditions are explanatorily irrelevant for computational explanations, contra (Hutto, 2013), as normally, computational syntax has systematic relationship to semantics.

By looking at some recent work on cognitive maps (Pfeiffer & Foster, 2013), Miłkowski will argue that one need not deny an important causal role of semantic entities and still account for computation in causal, mechanistic terms. From the mechanistic perspective, the notion of computation can be characterized in terms of the organization of one level of the mechanism, which is then 'bottomed-out' in lower levels; but one of the most important reasons to posit computation in the first place is a hypothesis that a mechanism has representational capacities, which also requires one to think of interrelations between the contextual level of the mechanism and the isolated level. Just because our best representational explanations in cognitive neuroscience are at the same time computational and causal, the dichotomy between the semantic explanations and computational explanations turns out to be false.

**References**

- Baumgartner, M. (2009), 'Interventionist Causal Exclusion and Non-Reductive Physicalism', *International Studies in the Philosophy of Science* 23(2): 161-178.
- Bechtel, W. (1994), 'Levels of Description and Explanation in Cognitive Science', *Minds and Machines* 4(1): 1-25.
- Bechtel, W. (2014). *Investigating neural representations: the tale of place cells*. Synthese.  
doi:10.1007/s11229-014-0480-8
- Beckermann, A. (1992), 'Supervenience, Emergence, and Reduction', in A. Beckermann, H. Flohr and J. Kim (eds.), *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, de Gruyter: 94-118.
- Cantwell Smith, B. (2002). *The foundations of computing*. In M. Scheutz (Ed.), *Computationalism: New Directions*. MIT Press.
- Chalmers, D. J. (2011). *A Computational Foundation for the Study of Cognition*. *Journal of Cognitive Science*, (12), 325–359.
- Chater, N., Tenenbaum, J.B. and Yuille, A. (2006), 'Probabilistic Models of Cognition: Conceptual Foundations', *Trends in Cognitive Sciences* 10(7): 287-291.
- Churchland, P.S. and Sejnowski, T.J. (1992), *The Computational Brain*, MIT Press.
- Bechtel, W. and Abrahamsen, A. (2005), 'Explanation: A mechanist alternative', *Studies in History and Philosophy of Biological and Biomedical Sciences*, Elsevier, 36, 421-441
- Bechtel, W. and Richardson, R.C. (1993), *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton: Princeton University Press.
- Craver, C. (2002), 'Interlevel experiments and multilevel mechanisms in the neuroscience of memory', *Philosophy of Science* 69, 83-97.
- Craver, C. (2007), *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Clarendon Press.
- Egan, F. (1992), 'Individualism, computation, and perceptual content', *Mind* 101(403): 443-459. Egan, F. (1995), 'Computation and Content', *Philosophical Review* 104: 181-204.
- Egan, F. (2014), 'Function-theoretic cognitive explanation and neural mechanisms', in D. Kaplan (ed.), *Integrating Mind and Brain Science: Mechanistic Perspectives and Beyond*, Oxford University Press.
- Esfeld, M. (2005), 'Mental Causation and Mental Properties', *Dialectica* 58(1): 5-18.
- Griffiths, T.L. and Tenenbaum, J.B. (2007a), 'From mere coincidences to meaningful discoveries', *Cognition* 103: 180-226.
- Griffiths, T.L. and Tenenbaum, J.B. (2007b), 'Two proposals for causal grammars', in A. Gopnik and L. Schulz (eds.), *Causal learning: Psychology, philosophy, and computation*, Oxford: Oxford University Press.
- Griffiths, T.L., Kemp, C. and Tenenbaum, J.B. (2008), 'Bayesian Models of Cognition', in R. Sun (ed.), *The Cambridge Handbook of Computational Psychology*: Cambridge University Press: 59-100.
- Halpern, J.Y. and Pearl, J. (2005), 'Causes and Explanations: A Structural-Model Approach. Part I: Causes', *British Journal for the Philosophy of Science* 56(4): 843-887.
- Hitchcock, C. (2001), 'The Intransitivity of Causation Revealed in Equations and Graphs', *Journal of Philosophy* 98(6): 273-299.
- Hitchcock, C. (2007), 'Prevention, Preemption, and the Principle of Sufficient Reason', *Philosophical Review* 116(4): 495-532.
- Hoffmann-Kolss, V. (2014), 'Interventionism and Higher-Level Causation', *International Studies in the Philosophy of Science* 28(1): 49-64.
- Hutto, D. D. (2013). *Exorcising action oriented representations: ridding cognitive science of its Nazgul*. *Adaptive Behavior*, 21(3), 142–150. doi:10.1177/1059712313482684
- Kaplan, D., (ed.) (forthcoming), *Integrating Psychology and Neuroscience: Prospects and Problems*. Oxford: Oxford University Press.
- Kim, J. (1998), *Mind in a Physical World*, MIT Press. Lycan, W.G. (1987), *Consciousness*, MIT Press.
- Machamer, P.K., Darden, L. and Craver, C.F. (2000), 'Thinking about Mechanisms', *Philosophy of Science* 67(1): 1-25.
- Mackie, J. (1974), *The Cement of the Universe*, Clarendon Press.
- Marr, D. (1982), *Vision: A computational investigation into the human representation and processing of visual*



- information, Freeman Press.
- Marras, A. (1993), 'Psychophysical Supervenience and Nonreductive Materialism', *Synthese* 95(2): 275-304.
- Miłkowski, M. (2013). *Explaining the Computational Mind*. Cambridge, Mass.: MIT Press.
- Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*, Cambridge University Press.
- Raatikainen, P. (2010), 'Causation, Exclusion, and the Special Sciences', *Erkenntnis* 73: 349-363.
- Pfeiffer, B. E., & Foster, D. J. (2013). Hippocampal place-cell sequences depict future paths to remembered goals. *Nature*, 497(7447), 74–9. doi:10.1038/nature12112
- Piccinini, G. (2007). Computing Mechanisms. *Philosophy of Science*, 74(4), 501–526. doi:10.1086/522851
- Polger, T.W. (2004), 'Neural Machinery and Realization', *Philosophy of Science* 7: 997-1006.
- Rescorla, M. (2012a). Against Structuralist Theories of Computational Implementation. *The British Journal for the Philosophy of Science*. doi:10.1093/bjps/axs017
- Rescorla, M. (2012b). How to Integrate Representation into Computational Modeling, and Why We Should. *Journal of Cognitive Science*, 13(1), 1–38.
- Rescorla, M. (2013). A theory of computational implementation. *Synthese*, 191(6), 1277–1307. doi:10.1007/s11229-013-0324-y
- Shagrir, O. (2010), 'Marr on Computational-Level Theories', *Philosophy of Science* 77(4): 477-500.
- Shagrir, O. (2012), 'Structural Representations and the Brain', *British Journal for the Philosophy of Science* 63(3): 519-545.
- Shagrir, O. and Bechtel, W. (2014), 'Marr's Computational-Level Theories and Delineating Phenomena', in D. Kaplan (ed.), *Integrating Psychology and Neuroscience: Prospects and Problems.*, Oxford University Press.
- Shapiro, L.A. (1993), 'Content, Kinds, and Individualism in Marr's Theory of Vision', *Philosophical Review* 102: 489-513.
- Shapiro, L.A. (1997), 'A Clearer Vision', *Philosophy of Science* 64: 131-153.
- Sterelny, K. (1990), *The Representational Theory of Mind: An Introduction*: Blackwell.
- Tenenbaum, J.B., Griffiths, T.L. and Kemp, C. (2006), 'Theory-Based Bayesian Models of Inductive Learning and Reasoning', *Trends in Cognitive Sciences* 10(7): 309-318.
- Wimsatt, W.C. (1994), 'The Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets', *Canadian Journal of Philosophy* supp. vol. 20: 207-274.
- Woodward, J. (2003), *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press.
- Woodward, J. (2008), 'Mental Causation and Neural Mechanisms', in J. Hohwy and J. Kallestrup (eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation*, Oxford University Press: 218-262.
- Woodward, J. (2010), 'Causation in Biology: Stability, Specificity, and the Choice of Levels of Explanation', *Biology and Philosophy* 25(3): 287-318.

**Historical, Social and Cultural Studies in****Contributed Papers****Philosophy of Science I**

Chair: Vasso Kindi

Room 5G, Thursday 13:00 – 15:00

---

Social Exclusion despite Methodological Criteria: On Biases in Scientific Quality Evaluation

ANNA LEUSCHNER

Karlsruhe Institute of Technology

anna.leuschner@kit.edu

---

Certain social groups are systematically excluded from the academic establishment. For example, there is a systematic exclusion of the contributions<sup>1</sup> of women philosophers from academia which is not solely caused through stereotypical behavior but also through biases that work deeply inside the scientific practice, particularly in the processes of scientific quality evaluation (e.g., Haslanger 2008; Hutchison & Jenkins 2013). I'll argue that such scientific exclusion is possible despite methodological criteria that are meant to come to an objective (i.e. purely epistemic) justification in the process of theory evaluation.

As Kuhn (1977) pointed out methodological criteria, such as accuracy, external consistency, simplicity, breadth of scope, and fruitfulness are underdetermined, i.e. they must be interpreted and weighed in a given context. This is, as Kuhn emphasizes, epistemically beneficial as it produces theoretical diversity.

Yet, for being epistemically beneficial in Kuhn's sense the criteria must be applied by a socially heterogeneous scientific community. In a socially homogenous community it is likely that the range of interpretations is too narrow to sufficiently warrant diversity. This becomes clear in light of feminist philosophy of science, particularly in light of the work of Longino (1990, 2002): as non-epistemic preferences influence the judgments of scientists via background assumptions, social plurality in scientific communities is required to keep biases in check. Only critical discourse makes biases visible; but for doing so (1) venues for criticism, (2) uptake of criticism, (3) shared standards (i.e. methodological criteria) to evaluate criticism, and (4) tempered equality in the communities are required (Longino 1990: 76–81, 2002: 128–135).

However, I'll argue that through biased interpretation and weighing of methodological criteria prejudices that are related to gender, race, social background, etc. can be tacitly and unconsciously reproduced, even in communities that fulfill Longino's four requirements.

Roughly speaking, a simple theory, for instance, can be evaluated as a brilliant clarification of complex relations but can also be dismissed as a naïve simplification—depending on the submitter's sex. A highly

---

<sup>1</sup> Note that contributions in this sense include not only submissions to conferences and journals but also essays, comments, or questions from female students in classes, colloquia, etc.

accurate theory can be regarded as perfectly adequate or as not innovative. External consistency can be interpreted as a quality feature or as boring and unoriginal.

Longino suggests adding three alternative methodological values to the traditional list: novelty, ontological heterogeneity, and complexity of interaction. These alternative values “(do or could) serve feminist cognitive goals” (Longino 2008: 72 & 77). However, there is no reason to assume that these values aren’t underdetermined. Hence, they are likewise open to interpretation and weighing in the context of their application. E.g., with respect to ontological heterogeneity a contribution can be evaluated as extraordinarily adequate or as too complex—a scientific non-finding.

I’ll argue that biased quality evaluation can work in two ways. First, *local* disqualification means that a scientific community tends to interpret and weigh methodological criteria *ad hoc*, so that they serve to disqualify the contribution of a woman. Consider, e.g., that the representation of female authors in scientific journals increases dramatically when review procedures become anonymous (Hutchison & Jenkins 2013). Second, *global* disqualification means that the respective scientific community tends to interpret methodological criteria in a way that *habitually* disqualifies the contributions of women. While it might sound unlikely at first glance that the contributions of women differ methodologically from those of men, empirical research reveals that the work of female scientists bears specific methodological characteristics: “As a group, women, as relative newcomers to science, adopted—or were taught to adhere—to an extra-high measure of conformity to the formal norms of conducting research.” (Sonnert & Holton 1996: 152) Women are “using a greater degree of caution, carefulness, attention to detail, and perfectionism.” (Fehr 2011: 151) It is likely that this characteristically influences the technical, interpretative, conceptual, and stylistic choices that women make: choices within the construction of experiments, the selection of instruments, the interpretation of data and other evidence, the weighing of results, etc.

In light of empirical studies, I’ll demonstrate how such biased methodological evaluation on both the local and global level leads to an exclusion of women from the philosophical establishment.

Finally, I’ll emphasize that methodological criteria are epistemic values, i.e. they are important scientific tools to distinguish signal from noise. But it is important to raise awareness of biases that can—and currently do—influence their application. Because social biases in science are so deeply embedded in the practices of scientific quality evaluation, we have good reason to be afraid that well-intentioned commitments won’t suffice to substantially change the situation. I’ll thus conclude that often mechanical solutions (e.g., strict adherence to alphabetical order in co-authored publications, triple-anonymous review procedures, quotas, etc.) are required to break the circle of exclusion and a seeming justification of that very exclusion.

## References

- Fehr, C. (2011): What Is in It for Me? The Benefits of Diversity in Scientific Communities. *Feminist Epistemology and Philosophy of Science: Power in Knowledge*. H. Grass- wick (ed.). Springer, 133–155.  
Haslanger, S. (2008): Changing the Ideology and Culture of Philosophy: Not by Reason (Alone). *Hypathia* 23

(2), 210–223.

Hutchison, K. & F. Jenkins (eds.) (2013): *Women in Philosophy. What Needs to Change?* Oxford University Press.

Kuhn, T. (1977): Objectivity, Value Judgment, and Theory Choice. *The Essential Tension.*

Selected Studies in Scientific Tradition and Change. The University of Chicago Press, 320–339.

Longino, H. (1990): *Science as Social Knowledge. Values and Objectivity in Scientific Inquiry.* Princeton University Press.

Longino, H. (2002): *The Fate of Knowledge.* Princeton: Princeton University Press. Longino, H. (2008): Values, Heuristics, and the Politics of Knowledge. *The Challenge of the Social and the Pressure of Practice. Science and Values Revisited.* M. Carrier, D. Howard, and J. Kourany (eds.). University of Pittsburgh Press, 68–86.

Sonnert, G. & G. Holton (1996): Career Patterns of Men and Women in the Sciences. *American Scientist* 84, 63–71.

---

*Self-Evidence in Scientific Practice*

ANTONIOS BASOUKOS

University of Exeter

antonis.basoukos@gmail.com

---

Ian Hacking's *Representing and Intervening* (1983) has been seminal in the development of the philosophy of science for helping put experimentation firmly on the map. Hacking's emphasis on experimentation culminates in his defence of entity realism. Entity realism is an unusual variety of scientific realism; the entity realist maintains we can know that postulated entities, like electrons, really exist, on the basis of their being routinely used by scientists as tools in the practice of science.

David Resnik (1994, 407), who has provided one of the most thorough criticisms of entity realism, has argued that Hacking's realism suffers from a serious shortcoming: manipulating electrons in experiments can yield belief – even true belief, Resnik adds- in the existence of electrons, but not justified true belief.

In this presentation I reconstruct Hacking's argument for realism. He has labelled this argument the 'experimental argument for realism'. My purpose is to demonstrate that the experimental argument for realism is a transcendental argument. In doing so, I will conclude that Hacking's famous dictum "if you can spray electrons, then they are real" is a self-evident proposition. My purpose is to show that Resnik is wrong in his evaluation about entity realism and epistemic justification.

Before proceeding on to the reconstruction of the experimental argument for realism, I must provide some background information on transcendental arguments. Transcendental arguments are a special kind of argument, which aim to show the legitimacy of certain partly non-empirical concepts like causation, the existence of other minds, and, in our case, the existence of unobservable entities postulated by science. By 'partly non-empirical' I mean that concepts like the aforementioned are familiar through application in everyday experience, but their application is not rooted in experience alone. This signifies that the validity of a transcendental argument is the result not only of experience, but also of the very content of the propositions involved. The content of these propositions is assessed by reason (using philosophical analysis) or by the meaning of the concepts making up these propositions. It is due to the *partly* non-empirical character of a transcendental argument that the self-evidence of propositions enters the scene.

Often, the general form of a transcendental argument is as follows:

P

P presupposes Q

Q

Therefore, if P then Q

You will notice that Hacking's dictum about spraying electrons has the same form as the conclusion in the outline above. It is a conditional.

Now let us move on to the reconstruction of Hacking's argument.

In a chapter entitled "Break" (Hacking 1983, 130-146), situated after the section on "Representing" and before the section on "Intervening", Hacking narrates an anthropological myth. Our ancestors of time immemorial used to gather around the camp fire to tell jokes. In those evening gatherings the concept 'real' emerged as a way to compare the likeness of artifacts and ideas with what they were supposed to represent. Soon, people started to expand their styles of making likenesses, saying "if this real, then that real". They even started talking about likenesses being "not real". For Hacking, the concept 'real' is a partly non-empirical concept; it exists as a manifestation of the discursive and the representing activities of people.

Of course, people's representing styles can give rise to mere speculation, which might range from the heuristically fruitful to the entirely misguided. We must return to the world, Hacking asserts, for the assessment of our speculative endeavours. It is by the return to the world that theoretical schemata become science. For instance (Hacking 1983, 141, 145), Democritus' atomism is perfectly sensible as a theoretical schema even if one disagrees with it, but it does not yield knowledge of how the world is; it is just speculation, albeit of the philosophical sort. To achieve knowledge we have to turn to experimentation.

The whole part on "Intervening" is a long argument for the autonomy of a scientific style of representing, namely the deployment of experiment both to control postulation and to explore by observation and measurement (Hacking 2002a, 182). Now we are in the domain of a clearly scientific style of representing, a style which encompasses visiting and revisiting the world. The experimental style is one of six scientific representing styles (Hacking 2002b, 161). All six of them, and each one on its own, are "reason itself" (Hacking 2002b, 176). They offer the background, i.e. spaces where intelligibility is possible, against which the spraying of electrons presupposes that they are real.

During the deployment of the experimental argument for realism Hacking makes clear that he is not talking about possible psychological states of experimenters who might objectify any entity they might be using (Hacking 1983, 265). Also, Hacking understands scientific anti-realism as scepticism (1983, 23, 263, 275). Transcendental arguments are usually deployed against sceptical positions on the knowability of a proposition. In this case, the sceptic doubts we can know if unobservable entities postulated by science exist. We can establish definitively that Hacking is deploying a transcendental argument when we consider that he argues for his own brand of verificationism. Barry Stroud, an influential critic of transcendental arguments, has argued persuasively that for such an argument to work we need to assume a verification principle (Stroud 1968, 255-256). Hacking (2002b, 165; 2002c, 4) does exactly that, adding only the clarification that meanings change in time, along with the methods of verification -which are the scientific styles of representation.

So, according to Hacking, we can *know* that if we can spray electrons then they are real, on the basis of the practices of the scientific enterprise.

**References**

- Hacking I. 1983. *Representing and Intervening*. Cambridge: Cambridge University Press.
- Hacking I. 2002a. "'Style' for Historians and Philosophers". In *Historical Ontology*, Cambridge: Harvard University Press, 178-199. Hacking I. 2002b. "Language, Truth and Reason". In *Historical Ontology*, Cambridge: Harvard University Press, 159-176.
- Hacking I. 2002c. "Historical Ontology". In *Historical Ontology*, Cambridge: Harvard University Press, 1-26.
- Resnik D. 1994. "Hacking's Experimental Realism". *Canadian Journal of Philosophy*, 24:3, 395-411.
- Stroud B. 1968. "Transcendental Arguments". *Journal of Philosophy* 65:9, 241–256.

---

Measurement Theory from the Viewpoint of Practical Realism on the Example of the Periodic Table of  
Chemical Elements

AVE METS  
University of Tartu  
avemets@ut.ee

---

Measurement as a procedure of assigning numbers to matter serves to order objects in the world into a system according to (magnitudes) of some attribute (Suppes and Zinnes), and thus to “reduce the problems of nature to the determination of quantities by operations with numbers” (Maxwell). The set of objects makes up, in terms of the attributes in question, an empirical relational system (ERS), and numbers, or more precisely – the arithmetic with a specific axiomatics – assigned to this set makes up a numerical relational system (NRS). David Hand distinguishes representational and pragmatic aspects of measurement. Representational approach of measurement first determines an ERS – observed attributes of objects and their relations – and then seeks to construct a NRS which preserves the structure of the ERS. The pragmatic approach of measurement chooses the objects to be measured on grounds external to the empirical system being represented without requiring that the variable defined represents a property in the real world.

Using the periodic table of chemical elements as example and practical realism as philosophical framework, I will show that 1) the objects making up ERS are abstract objects, not real concrete objects in the world, 2) the material procedural basis of measurement is of fundamental importance for the possibility of systematisation, 3) a measurement system is made up not only of ERS and NRS, but also of the material basis and theoretical and metaphysical presuppositions about it and 4) therefore discerning representational from pragmatic measurement in physical sciences is not straightforward.

The periodic system of chemical elements can be regarded as a measurement system: The contemporary shape of the periodic table orders chemical elements and assigns them numbers according to their atomic number  $Z$ , or the number of protons in their nuclei, engendering an integral scale with arithmetic summation as concatenation (compounds of elements) defined on it; and valence, or the number of electrons free for bonding with other elements, with closed integral scale and a more complicated arithmetic. The periodic law can be regarded as the ordering principle for chemical elements: chemical and physical properties vary periodically over to the atomic number of elements, that answers to Suppes' and Zinnes' contention that “to construct an instrument which will provide direct or at least quick measurement of some fundamental or derived scale it is generally necessary to utilize some established empirical law or theory involving the fundamental or derived scale in question,”.



However, the original ordering of elements on the basis of their atomic weights was much less straightforward and complicated by technical and conceptual difficulties in chemical research (Scerri, Bensaude-Vincent, Brooks). Firstly, discerning element as abstract notion from simple substance was fundamental for defining something to have a stable essence or properties, unlike the material simple substances that are often difficult or impossible to obtain and vary according to their external physical and chemical circumstances (e.g. in different compounds they have different properties, and also different from the pure substance). In order to have some entity that hypothetically stays the same over all those varying circumstances, element as abstract entity had to be postulated. Secondly, chemical reactions were the empirical, experimental basis for determining both atomic (at first equivalent) weights and valences of elements, that thereby depend on each other and require some additional hypothesis. Whereas hydrogen was used as the unit of mass, not all elements readily react with it, so oxygen as an intermediary element had to be used, thus engendering more uncertainties. If atomic weight were taken as the ultimate basis for ordering elements, their scale were rather in disorderly, even though additional assumptions enabled determining approximate arithmetic similarities between chemically similar elements. However, x-ray bombing as the material method for determining the structure of atom, thus the number of protons in it, indeed provides an arithmetically elegant and unambiguous scale in contrast to chemical reactions. Thirdly, additional hypothesis, mentioned above, that enabled conceptualising the chemical reactions as measurement situations, were the atomic hypothesis about the essence of chemical elements and thus underlying the notion of atomic weight (in contrast to equivalent weight) and the further hypotheses of the number of particles in a unit of volume or mass, and of polyatomicity of simple substances, particularly of gases. The latter ones united in Avogadro hypothesis were an essential framework for interpreting reactions of gaseous substances – determining the ratios of atoms and thus atomic weights and valences. Underlying those are thus the metaphysical assumption of atomicity of matter and identity of elements across compounds. But also Mendeleev's metaphysical assumption of individuality of elements – that each element has its unique physical properties not shared with any other elements, was an important background belief in constructing the table. Fourthly, according to atomic weights, elements did not yield neither a periodicity meaningful for chemical purposes, as the weights of several elements had to be "corrected" to save other chemically and physically important properties, nor a simple elegant arithmetic of assigned numbers, as the relations found were rough and statistical. Nonetheless, the idea of systemicity from elegant numerical relations guided the formation of the periodic law, the corrections and predictions. This drive toward simplicity, neglecting empirical discrepancies, witnesses of the pragmatism, in Hand's sense, of the measuring system that the periodic law is.

---

*Local Ontologies and the Integration of Indigenous Knowledge*

DAVID LUDWIG

VU Amsterdam

davidundludwig@gmail.com

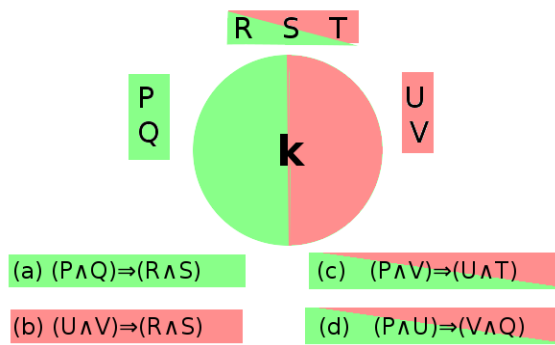
---

The integration of indigenous and scientific knowledge is a widely discussed topic in ethnobiology that has not received sufficient attention in philosophy of science. I propose a model of knowledge integration and of integration limits that reflect the local stability of clustered properties. I argue that the limits of knowledge integration are best understood as creating normative and not metaphysical problems. Two knowledge systems that refer to different property clusters may be metaphysically integrated in a broader framework but this does not solve the normative question what type of knowledge should guide actions in local environments.

**1. A simple model of knowledge integration**

“Local knowledge” is a vague and rather controversial term. On the one hand, the very idea of locality is often used as a tool of relativization that contrasts “subjugated knowledges” (Foucault 1976, 7) with a dominant scientific discourse. On the other hand, anthropologists are also using the term “local knowledge” in efforts that aim at integration with modern science. The most prominent examples come from conservation biology and integration efforts in the development of sustainable conservation strategies (Berkes et al. 2000).

In the following, I propose a simple model of knowledge integration that is based on stable clustering of properties. Although this model does not capture all shades of current integration debates, a substantial part of successful integration arguably reflects joint reference to property clusters. Consider a simple example such knowledge about jaguars in an indigenous and modern scientific context. Jaguars typically share a large number of (e.g. anatomical, behavioral, ecological, genetic, phylogenetic, physiological) properties and at least some them (e.g. overt morphological properties and certain types of behavior) will be familiar in both knowledge systems. Other properties may be known only in a modern scientific (e.g. genetic properties) or indigenous (e.g. hunting habits on a local Tapir population) context. More abstractly, we end up with a scenario along figure 1. A kind  $k$  has a number of properties  $P, Q, R, S, T, U, V$ . Some properties are known in both knowledge systems  $S1$  and  $S2$  while other only known in  $S1$  or  $S2$ .



*Figure 1:* A kind *k* comes with a cluster of properties P-V that are partly recognized by knowledge system S1 (green), partly by S2 (pink), and partly by both knowledge systems (green-pink). (a) - (d) illustrate the inferential productivity of integrating S1 and S2 by not only allowing addition of probabilistic inferences such as (a) and (b) but enabling genuinely novel inferences such as (c) and (d).

**2. A simple model integration failures**

While the proposed model works surprisingly well in a many ethnobiological case studies, knowledge integration can fail for several reasons. In this section, I will present a model of integration failure that is based on the phenomenon of merely local stability. Property clusters are usually only stable to a certain degree (Slater 2014). For example, typical behavioral properties of jaguars can disappear in captivity while typical morphological or genetic properties could disappear through breeding. Probabilistic inferences such as (a) - (d) therefore presuppose normalized circumstances. The proposed model of knowledge integration fails if a property cluster is stable under normalized circumstances of S1 but fails to be stable under normalized circumstances of S2.

There are at least two important classes of this merely local stability. First, local stability may reflect focus on a specific ecosystem. For example, Ludwig (forthcoming) discusses the case of mushroom taxa among the Tzeltal Maya that appear to refer to stable property clusters within the highlands of Chiapas (they grow during specific times at specific locations, are not edible, have similar size and appearance). However, this cluster is not stable in other ecosystems. Second, property clusters are often “hybrid” in the sense that they involve biological and cultural (e.g. aesthetic, economic, nutritional, medical, spiritual) properties. Insofar as the cultural properties vary across societies, different knowledge systems will also involve slightly different property clusters. While this variation does not make much of a difference in some contexts (S1 & S2 may both refer to jaguars, even if only S1 associates them with certain aesthetic properties), there are arguably many kinds that recognized as relevant kinds *only* because they have both biological and certain cultural properties.

Both cases of merely local stability set limits to integration and leads to a more complex picture as represented in figure 2.

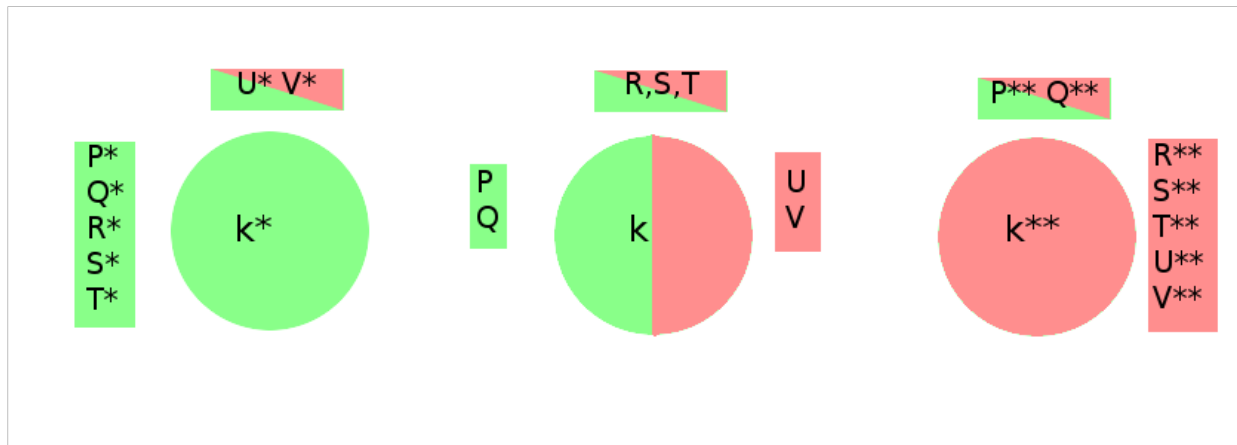


Figure 2:  $k$  is recognized in both knowledge systems as  $[P, Q, R, S, T]$  constitutes a relevant property cluster for S1 while  $[R, S, T, U, V]$  constitutes a relevant property cluster for S2. In contrast,  $k^*$  and  $k^{**}$  only appear in the ontology of S1 or S2 because only one knowledge system recognizes a relevant property cluster.

### 3. From Metaphysics to Politics

Despite cases of successful knowledge integration (figure 1), local knowledge systems arguably set limits to knowledge integration (figure 2) and involve at least partial ontological incommensurability: some entities in S1 have no place in S2 and vice versa. While ontological incommensurability may lead to metaphysical worries, one may respond by integrating different kinds in a broader ontology. Even if  $k^*$  only appears in S1 while  $k^{**}$  appears in only in S2, we may endorse a permissive ontology that incorporates  $k, k^*,$  and  $k^{**}$ . Such an ontology will allow at least a weak form of integration of S1 and S2 in a broader framework.

However, this solution does not solve the normative problems that motivate much of the postcolonial discourse about local knowledges. Debates about the integration of indigenous knowledge systems often focus on aspects that allow seamless integration with western science along figure 1. However, the limits of integration raise the question what knowledge should be action-guiding in the interaction of different knowledge systems. As Nasady (1998, 12) puts it in the context of conservation management efforts: As long as scientists focus on collecting and documenting local knowledge “as an intellectual product to be integrated with science, they will be helping to extend [networks that] cannot help but concentrate power in the hands of scientists and resource managers”. Integration that does not involve this kind of subjugation requires a look at the limits of integration and the action-guiding potential of non-integrated knowledge.

---

*Reaction Mechanisms in Chemistry: A Comparison Case for Accounts of Scientific Explanation*

ANDREA WOODY

University of Washington

awoody@u.washington.edu

---

Philosophy of science offers a rich lineage of analysis concerning the nature of scientific explanation. In recent years, considerable attention has been directed toward the notion of mechanistic explanation, especially in the biological sciences (see, for example, Bechtel (2011), Bechtel and Abrahamsen (2005), Craver (2007), Fagan (2012), Glennan (1996, 2006), Machamer, Darden and Craver (2000) and Woodward (2013)). Much of this work aims to characterize mechanisms or determine how mechanisms need to be described in order to be explanatory. Some of it aims to make explicit contact with contemporary analyses of causation, especially those of Salmon (1984, 1998) or Woodward (2005). This paper examines another scientific context in which appeal to mechanisms is arguably as widespread and central as it is in biological contexts but which has received much less attention: explanatory patterns involving reaction mechanisms in organic chemistry. There are two fundamental aims: (1) to develop a characterization of mechanisms in chemistry as a comparison case for existing analyses of mechanism in the biological sciences, and (2) to use this comparison to highlight certain aspects of explanatory practice across the sciences.

Drawing on recent work by Goodwin (2011, 2012), and in response to Ramsey (2008), the paper begins with a general characterization of reaction mechanisms and their role in explanations in organic chemistry. From this characterization, I will argue that mechanistic explanations in chemistry seem different in important respects from their counterparts in biology. While mechanistic explanations in both contexts often stress spatial relations, mechanistic explanations in chemistry typically focus on information often lacking in biological cases, specifically information concerning the rate of operation of (some, but only some) various processes that compose a given mechanism. In this regard they resemble the dynamical explanations some consider a serious challenge to existing accounts of mechanistic explanation in biology. They also are frequently embedded in a more general analysis of the energetics of reaction pathways. At the same time, mechanistic explanations in chemistry typically omit or suppress information included in biological mechanistic explanations and do not exhibit the centrality of part-whole relations that are a hallmark of mechanistic explanations in biology.

The second step of the analysis turns more specifically to practice. I argue that the types of information included in chemical mechanisms, as well as the ways in which this information is represented in relation to

potential energy diagrams, serves to support the largely synthetic aims of organic chemistry. I will suggest that general differences in the aims of given scientific communities may influence what sort of description or information is required for a mechanism to be judged explanatory. Finally, I will return to broad issues concerning scientific explanation, arguing that an account of explanation sufficiently oriented toward explanatory practice will be best suited to make sense of the sorts of differences we observe in comparing chemical and biological mechanisms taken to be explanatory by their respective communities. Such an account of explanation stresses the methodological role of explanatory discourse in ways I have discussed elsewhere and will summarize briefly to conclude.

---

*A Dispositionalist Theory of Laws – Without Dispositions*

ANDREAS BARTELS

Universität Bonn

andreas.bartels@uni-bonn.de

---

Strong dispositionalist theories of laws as *dispositional monism* (Bird 2007) provide a promising approach for those who think that natural laws do not supervene on matters of fact, but rest on some independent metaphysical basis. According to dispositional monism, fundamental physical properties have their identity by their being specific *powers* producing their manifestations by metaphysical *necessity*. Natural laws are true by virtue of those necessary relations between powers and their manifestations.

The assumption that fundamental laws of physics are constituted by necessary links seems to be well adapted to experience. It provides an explanation of the fact that fundamental physical quantities produce their (primary) effects in some “irresistible” way, i. e. such that the effects cannot be influenced or be eliminated by other factors. The metrical properties of spacetime, for instance, produce their primary effect, the tidal forces, in a way that cannot be interfered with by any other forces. It is only the secondary, resulting effects that depend on the influences of other forces, for instance, how the tidal forces, in combination with other forces, produce the motion of a particle. But the contribution of any fundamental force within that combination is independent of the activities of the others.

On the other hand, the explanation that dispositional monism offers for the irresistibility of the primary effects of fundamental properties comes along with strong metaphysical baggage that makes it incompatible with the methodological practice of physics. If, for instance, metrical properties produced tidal forces by metaphysical necessity, then any other sorts of relations between the metric and the affine connection (e.g., independence of these structures, as it is realized in the Palatini-formalism of GR, cf. Palatini 1919) would be ruled out as presenting a metaphysically impossible world. Contrary to that verdict, those theoretical options would be treated by physicists as representing physical possibilities. Now, what counts as a physical possibility, according to some reasonable methodological practice, should not be taken to be a metaphysical impossibility. Otherwise, we would have to accept that even variants of one and the same theory – stipulating the relations between fundamental quantities in somewhat different ways – would represent ways the world *could* not be. Since such a metaphysical position would lead to a rather severe tension with the methodological practice of physics, I search for a more sparse metaphysics of laws.

The metaphysically sparse dispositionalism I plea for is called *methodological dispositionalism* (MD). According to MD, to say that laws ascribe dispositions to physical systems is to describe how physical systems would behave in the absence of other fundamental forces (cf. Hüttemann 1999). This sort of dispositionalism

accounts for the methodological practice of physics to universally apply laws which describe the behavior systems *would* display in complete isolation from other influences. What a law describes (e. g. its content) turns out to be the *stable contribution* of a fundamental property to the behavior of systems having this property, where this contribution is assumed to be displayed in some “irresistible” manner, that is, independently of the contributions of any other factors.

MD dispenses with dispositional entities, or powers, producing their effects with necessity by virtue of their essential nature. It joins with categoricalism by proposing that the identity of fundamental physical properties is given by their “categorical” characteristics (the categorical characteristics of the metric, e. g., are given by being a measure for the length of vectors of the tangential space at a point). The categorical characteristics of a property specify what property it is, while the dispositional characteristics specify, how the property is (contingently) related to other properties. The latter is what we call a (contingent) law of nature.

There is an alleged apriori objection to categoricalism: the *argument from Quidditism* (cf. Bird 2007, 70f.). According to the argument, categoricalism entails an unacceptable consequence: there could be different properties in the actual world that have identical causal profiles. Since those profiles are presumably the only way how we can get knowledge about properties, categoricalism (because of its adherence to “quiddities”) would entail an absolute limit for our knowledge of properties. I will show that the argument is inconclusive: what we get by categoricalism are at most possible cases of familiar under-determination of theoretical entities (properties) by their empirical manifestations.

Thus, there is no apriori objection to MD. But the question remains how to understand contingent dispositional characteristics of fundamental properties. Since the basic idea of MD has strong structural similarity to the DTA-approach (Armstrong 1983, 1997), I try to explain the peculiarities of MD by contrasting it to DTA:

First, MD like DTA (cf. Armstrong 1997) holds that lawful relations are contingent *causal* relations. Since the fundamental laws of physics do not specify a time direction, the time- asymmetry needed for these causal relations must be determined by the contingent asymmetric time-structure of the actual world in which the laws occur – the time-asymmetric structure of the actual world being a *typical* property of general relativistic worlds (cf. Bartels & Wohlfarth 2014).

MD specifies the nature of lawful causal relations as transfer of energy-momentum. Thus, lawful relations in our actual world are thus constituted by facts of energy-momentum transfer time-asymmetrically relating tokens of property types.

Second, according to MD, and unlike DTA, lawful relations between properties are *not necessitating* relations according to MD. Instead, MD takes facts of irresistible production by fundamental properties to be *primitive facts* about our world.



**References**

- Armstrong, David (1997): *A World of States of Affairs*, Cambridge University Press: Cambridge
- Armstrong, David (1983): *What is a Law of Nature?*, Cambridge University Press: Cambridge
- Bartels, Andreas & Daniel Wohlfarth (2014): How Fundamental Physics Represents Causality, In: Maria Carla Galavotti et al. (eds.): *New Directions in the Philosophy of Science*, Springer: New York, 485-500
- Bird, Alexander (2007): Nature's Metaphysics, Oxford University Press: Oxford
- Hüttemann, Andreas (1999): Laws and Dispositions, *Philosophy of Science* 65, 121-135
- Palatini, Attilio (1919): Deduzione invariante delle equazioni gravitazionali dal principio di Hamilton, *Rendiconti del Circolo Matematico di Palermo* 43 (1), 203-212

---

*Unitary Inequivalence in Classical Systems*

BENJAMIN FEINTZEIG

Logic and Philosophy of Science, UC, Irvine

bfeintze@uci.edu

---

Ruetsche (2011) argues that a *problem of unitarily inequivalent representations* arises in quantum theories with infinitely many degrees of freedom. When one attempts to “quantize” a classical theory, i.e. formulate a quantum theory for the same system the classical theory was meant to describe, one is not guaranteed that the resulting quantum theory is unique.<sup>1</sup> For a classical theory with infinitely many degrees of freedom, e.g. a field theory or a statistical theory in the thermodynamic limit, there are many inequivalent theories which compete to be called its “quantization”.

Work on the problem of unitarily inequivalent representations is done in the algebraic framework for quantum theories. One begins by representing physical observables as elements of an abstract C\*-algebra, which captures the structure that all of the representations of the canonical commutation or anti-commutation relations have in common. One then looks for concrete representations of that algebra in the bounded operators on some Hilbert space. The problem of unitarily inequivalent representations forces us to make a choice in interpreting the algebraic formalism. Either one takes the algebraic formulation as basic, and becomes an *Algebraic Imperialist*, or else one privileges a particular Hilbert space representation as a *Hilbert Space Conservative*.<sup>2</sup>

This paper investigates the algebraic formalism itself for the purpose of assessing these interpretive options. I will step back from the specific details of quantum theory, whose understanding is already so controversial, and shift to the simpler context of classical physics. As is already known,<sup>3</sup> one can use the very same algebraic formalism previously mentioned to describe classical theories as well as quantum ones. Since the interpretation of classical physics is at the very least better understood and better agreed upon than that of quantum mechanics, this suggests that we can use classical field theory to probe our understanding of the abstract algebraic formalism. This paper uses an algebraic reformulation of classical field theory as a concrete

---

<sup>1</sup> This assumes that a quantum theory is a concrete Hilbert space representation of canonical commutation or anti-commutation relations. This assumption is not universally shared and I will discuss precisely how it is challenged.

<sup>2</sup> I take this terminology from Ruetsche (2011, Ch. 6), who adapts it from Arageorgis (1995). Algebraic Imperialism and Hilbert Space Conservatism are just two of the most prominent stances, but they do not exhaust all possible interpretive options.

<sup>3</sup> It has for some time now been accepted that abelian algebras may be used to represent the observables of a classical system (e.g. Summers & Werner 1987, 2441), but it was not until recently that such a formulation was made explicit (Brunetti et al. 2012).

case in order to investigate whether similar interpretive and foundational issues arise in the classical case as in the quantum case, and whether natural solutions are suggested in the classical case.

In this sense, the present paper falls into a tradition in philosophy of physics of translating our previous theories into the language of our current theories for the purpose of assessing what precisely is novel about our new theories. For example, many use Newton-Cartan theory (geometrized Newtonian gravitation) as a way of translating classical Newtonian gravitation into a framework in which one can compare it with general relativity. Upon doing so, one finds that at least some of the things that have been said were distinctive features of general relativity turn out to be features of Newton-Cartan theory as well (see, e.g. Weatherall 2011). The purpose of this paper is to make a similar point about the relationship between quantum field theory and classical field theory. While many have argued that the problem of unitarily inequivalent representations is a conceptual problem for quantum theories with infinitely many degrees of freedom, I will argue that the mathematical features that lead to this problem are not distinctive of quantum theories. Furthermore, I will argue that looking at the classical case helps us understand just what the problem is and how to go about looking for solutions.

I will show that unitarily inequivalent representations arise in the algebraic formulation of classical field theories.<sup>4</sup> More specifically, I will show that in classical theories, the GNS representations for any two distinct pure states are unitarily inequivalent. In a certain sense, the problem is even more pressing in the classical case. However, there is another sense in which it is not a problem at all in the classical case because its solution is obvious. I will argue that the presence of unitarily inequivalent representations makes the position of Hilbert Space Conservatism (as extended to classical field theories) untenable. Furthermore, I will show that the standard argument against Algebraic Imperialism fails in the classical context. To the extent that this moral concerning the algebraic formalism carries over from the classical case to the quantum case, one also ought not be a Hilbert Space Conservative about algebraic quantum theory.

## References

- Arageorgis, A. 1995. *Fields, Particles, and Curvature: Foundations and Philosophical Aspects of Quantum Field Theory in Curved Spacetime*. PhD diss., University of Pittsburgh.
- Brunetti, R., K. Fredenhagen, and P. Ribeiro. 2012. "Algebraic Structure of Classical Field Theory I." arXiv:1209.2148v2.
- Ruetsche, L. 2011. *Interpreting Quantum Theories*. NY: Oxford University Press.
- Summers, S. and Werner, R. 1987. "Bell Inequalities and quantum field theory. I. General setting." *Journal of Mathematical Physics*. 28.10: 2440-2447.
- Weatherall, J. 2011. "On the status of the geodesic principle in Newtonian and relativistic physics." *Studies in the History and Philosophy of Modern Physics*. 42.4: 276-81.

---

<sup>4</sup> By this, I do not mean to assert that unitarily inequivalent representations in classical theories bear the same interpretation outlined above. In particular, I do not mean that there is some kind of non-uniqueness in our description of classical theories or that we have competing versions of classical theories. Rather, I mean that analogous technical results hold in the classical case that force us to reconsider our interpretive options.

---

*Reconceptualising Equilibrium in Boltzmannian Statistical Mechanics and Characterising its Existence*

CHARLOTTE WERNDL

University of Salzburg

c.s.werndl@lse.ac.uk

ROMAN FRIGG

London School of Economics

r.p.frigg@lse.ac.uk

---

The core posit of Boltzmannian statistical mechanics is that macro-states supervene on micro-states. This leads to a partitioning of the state space of a system into regions of macroscopically indistinguishable micro-states. These regions are called macro-regions. The largest of these macro-regions is commonly singled out as the system's equilibrium region. The two core questions about equilibrium of this paper are: 1) What justifies the association of equilibrium with the macro-state corresponding to the largest macro-region? 2) Under what conditions do systems approach equilibrium?

We start with the first question and scrutinise three answers that have been given. The first is that equilibrium corresponds to the macro-state that is compatible with the largest number of micro-states (cf. Boltzmann 1877). This justificatory strategy faces a serious problem: the absence of a conceptual connection with the thermodynamic notion of equilibrium. There is simply no conceptual connection between this notion of equilibrium and the idea that the equilibrium macro-state is the one that is compatible with the largest number of micro-states. A second answer is to define equilibrium in terms of the Maxwell-Boltzmann distribution. This is not a viable definition. The baker's gas or the Kac-ring and, in general, systems with non-negligible interactions will have equilibrium distributions that are different from the Maxwell-Boltzmann distribution (Gupta 2002). A third answer is that this notion of equilibrium can be justified by maximum entropy consideration along the following lines: we know from TD that, if left to itself, a system approaches equilibrium, and equilibrium is the maximum entropy state. Hence the Boltzmann entropy of a macro-state is maximal in equilibrium. Since the thermodynamic entropy is a monotonic function, the macro-state with the largest Boltzmann entropy is also the largest macro-state, which is the desired conclusion. This strategy also does not work. First, thermodynamics does not attribute an entropy to systems out of equilibrium at all. Second, even if this could be resolved, there would remain the question why the fact that the thermodynamic entropy reaches a maximum in equilibrium would imply that this also holds for the Boltzmann entropy. To justify this inference, the assumption would need to be made that the thermodynamic entropy reduces to the Boltzmann entropy. However, this is far from clear. A connection has been established only for ideal gases. For systems with interactions no such results are known (cf. Frigg and Werndl 2011b).

The failure of all these answers prompts the search for an alternative answer. This answer cannot be found by revising any of the received approaches, and so we propose a new definition of equilibrium. While previous approaches sought to define equilibrium in terms of micro-mechanical properties, our definition is modelled on the thermodynamic conception of equilibrium. Roughly, according to our new definition, equilibrium is the state in which almost all initial states spend most of their time in (in the long run). This new conception provides the spring-board for a general answer to our initial problem. We prove a mathematical theorem which establishes in full generality that the equilibrium macro-region is the largest macro-region (in a requisite sense). The theorem is completely general in that it makes no assumptions either about the system's dynamics or the nature of the interactions between the system's components.

We then turn to the second question of the approach to equilibrium. In our account, this question is replaced by the question: under what circumstances does an equilibrium state exist? We first review common answers to this question. The currently most influential account in physics is the typicality account. It has been developed, among others, by Goldstein and Lebowitz (2004). The leading idea behind this account is that systems approach equilibrium because equilibrium micro-states are typical. It is our considered view that this account is unsuccessful because it fails to take the system's dynamics into account. The canonical answer is given within the ergodic programme. The leading idea is that systems approach equilibrium if, and only if, they are ergodic. Even if the ergodicity account would apply to dilute gases (e.g. Frigg and Werndl 2011a), systems different from dilute gases can be expected not to be ergodic and hence this cannot be a general answer. Finally, there is a family of proposals that grounds the approach to equilibrium in different kinds of probabilistic dynamics. For instance, Boltzmann (1877) introduces the probability of a macro-state and postulates that this probability is proportional to its size. Since equilibrium is the largest state it is also the most likely state. Systems then evolve from less to more likely states, which explains the approach to equilibrium. These approaches are discussed in Frigg (2010) and found wanting both for technical and conceptual reasons.

The failure of these answers prompts the search for a new answer. We point out that for an equilibrium to exist three factors need to cooperate: the choice of macro-variables, the dynamics of the system, and the choice of the effective state space. We then prove a theorem providing fully general necessary and sufficient conditions for the existence of an equilibrium state. This theorem appeals to von Neumann's ergodic decomposition theorem, stating that every system can be broken down into ergodic components (which might be of measure zero). According to our theorem, an equilibrium state exists if, and only if, most of each ergodic component of the system is taken up by equilibrium states (except for components of measure  $\epsilon$ ). This theorem changes the way in which the problem of the approach to equilibrium should be discussed: rather than launching a search for one crucial factor (such as ergodicity or typicality), the focus should be on finding triplets of macro-variables, dynamical conditions, and effective state spaces that satisfy the conditions of the theorem. This gives the discussion of equilibrium a new direction.

**References**

- Boltzmann, L. 1877. Über die Beziehung zwischen dem zweiten Hauptsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung respektive den Sätzen über das Wärmegleichgewicht. Wiener Berichte 76: 373-435.
- Frigg, R. 2008. A Field Guide to Recent Work on the Foundations of Statistical Mechanics. In D. Rickles (ed.), *The Ashgate Companion to Contemporary Philosophy of Physics*, 99-196. London: Ashgate.
- Frigg, R. and Werndl, C. 2011a. Explaining Thermodynamic-Like Behaviour in Terms of Epsilon-Ergodicity. *Philosophy of Science* 78: 628-652.
- Frigg, R. and Werndl, C. 2011b. Entropy -- A Guide for the Perplexed. In: C. Beisbart and S. Hartmann (eds.), *Probabilities in Physics*, 115-142. Oxford: Oxford University Press.
- Goldstein, S. and Lebowitz, J.L. 2004. On the (Boltzmann) Entropy of Nonequilibrium Systems". *Physica D* 193:53-66.
- Gupta, M.C. 2003. *Statistical Thermodynamics*. New Delhi: New Age International Publishing.

**General Philosophy of Science III****Contributed Papers**

Chair: Richard David Rus

Room 5H, Thursday 13:00 – 15:00

*Causal-Possibility Explanations*

LANE DESAUTELS

University of Notre Dame  
lane.desautels@gmail.com

GRANT RAMSEY

University of Notre Dame  
grant@theramsey.org

In this paper we articulate a form of causal explanation that is modal, not ontic. We label this form of explanation *causal-possibility explanation*. By arguing for causal-possibility (C-P) explanations, we are going against the spirit (if not the letter) of the classic, long-received treatment of scientific explanation by Salmon. In challenging this received view, we suggest that not all modal explanations are non-causal. And not all causal explanations are ontic. To motivate this position, we offer examples of C-P explanations and show why they are both modal and causal. We then explore how C-P explanation squares with similar positions in the literature.

And we conclude by discussing recent accounts of non-causal modal explanation and how they relate to our account.

Salmon (1984, 1989, 1998) famously draws a tripartite distinction between kinds of explanation: *modal*, *epistemic*, and *ontic*. On the modal conception, explanation proceeds by showing of some event that “[g]iven the particular set of initial conditions, and the laws of nature, the explanandum event *had* to occur” (1998, p. 53). On the epistemic conception, explanation is “an argument to the effect that the event to be explained was to be *expected* by virtue of the explanatory facts” (1998, p. 166). And on the ontic conception, the explanation of events consists in “fitting them into the patterns that exist in *in the objective world*” (1989, p. 121). To avoid certain well-known problems with explanatory relevance and the asymmetry of explanation, Salmon and others (e.g., especially Craver 2007) argue convincingly that our explanations should be ontic and, more specifically, good explanations should locate the explanandum event in “the causal nexus of the world”; good explanations, on the received view, should be *ontic* and *causal*. In this paper, we suggest that Salmon’s long-accepted tripartite distinction appears to leave out a class of explanations important to science: explanations which are *both* modal (not ontic) *and* causal.

To begin to illustrate what we mean by C-P explanation, recall an example famously articulated by Hillary Putnam (1975) in which we are trying to explain why a one-inch square peg won’t fit into a circular hole of the same diameter. We might explain this by appeal to a detail-rich account of why the specific micro-physical properties of *this particular peg* impede its ability to pass through the molecular make-up of *this particular hole*. But the more explanatorily satisfying answer here, Putnam points out, might actually be one having to

do with the generic properties of squareness and roundness of the peg and hole as such. In giving such an explanation we can track the features of the system (e.g., the squareness of the peg and roundness of the hole) that make a *causal* difference to the behavior we're interested in explaining. But an explanation appealing to the squareness of the peg and roundness of the hole does more than explain why this particular peg fails to fit through this particular hole. It supplies *modal* information: information that tells us something about why no square peg will fit through a round hole – even when they have the same diameters. An explanation appealing to the properties of roundness and squareness is thus both *causal* (providing information about causal relations in the world) and *modal/non-ontic* (providing information relevant to more than the explanandum event in the actual world). This, we suggest, makes it a C-P explanation. Causal-possibility explanations aren't just the stuff of toy examples. We contend that C-P explanations are ubiquitous in the sciences. Physicists ascribe radioactive half-lives to molecules based on non-actual, highly-probable possible future states of decay. Evolutionary biologists explain the fitness of a particular organism (or trait) in terms of possible ways its life might go.

Ecologists make optimality ascriptions by appeal to other less optimal ways the population might have been distributed. In all of these cases, what's doing the explaining isn't just an appeal to the causal-history of the actual world. Neither is it a claim about what is *necessary* given some laws and preconditions.

An important distinguishing feature of our view of C-P explanation is that they can exhibit a variable degree of causal information and modal information. One can therefore represent scientific explanations as occupying points in two-dimensional space with causal information on one axis and modal information on the other. In one corner of the space are purely causal-historical ontic explanations: explanations in which an explanandum event is explained by merely citing the most relevant parts of the specific causal-history leading up to it. In another corner of the space are purely modal explanations: explanations in which the modal information is doing all the explaining and no causal/historical information is doing any work (perhaps distinctively mathematical explanations are an example of these). One of our central points is this: *Traditional views of scientific explanation focus only on the boundaries of the space, but we wish to draw attention to a whole range of explanations in between.*

Similar distinctions have been drawn by others. Jackson and Pettit (1990) distinguish *process explanations* from *program explanations*, Sterelny (1996) distinguishes *actual-sequence explanations* from *robust-process explanations*, and more recently, Skow (2013) defends a view of causal explanation he calls *almost-necessity explanation*. In each of these author's accounts, they have pointed to (something) like C-P explanation. However, for reasons we shall articulate, none of them are quite satisfactory. Jackson and Pettit's program explanations appeal to a notion of programming which we find overly intentional/teleological. Sterelny's account of robust- process explanation under-emphasizes the graded notion of causal and modal information articulated above. And Skow's notion of almost-necessity explanation occupies only one region of the two-dimensional space we delineate.



Furthermore, we contend that our account of C-P explanation squares nicely with recent accounts of *non-causal explanation* in the sciences. In particular, we argue that the explanations cited by proponents of non-causal explanation (e.g., distinctively mathematical explanation, identity explanation, and symmetry explanation), are aptly situated on the farthest modal corner of our two-dimensional space.

### References

- Craver, C. F. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, New York: Oxford University Press.
- Jackson, F., and Pettit, P. 1990. "Program Explanation: A General Perspective," *Analysis* 50 (2): 107-117.
- Putnam, H. 1975. "Philosophy and Our Mental Life," In *Mind, Language, and Reality*. Cambridge: Cambridge University Press, 291-303
- Salmon, W. 1984. *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.
- Salmon, W. 1989, *Four Decades of Scientific Explanation*, Minneapolis: University of Minnesota Press.
- Salmon, W. C. 1998, *Causality and Explanation*. Oxford University Press
- Skow, B. 2013, "The Role of Chance in Explanation," *Australasian Journal of Philosophy* (1): 1- 21
- Sterelny, K. 1996, "Explanatory Pluralism in Evolutionary Biology. *Biology and Philosophy* 11 (2): 193-214.

---

*Scientific Realism as a Pragmatic Attitude*

JESUS ZAMORA BONILLA

UNED

jpzb@fsof.uned.es

---

The decision whether to have a realist or an instrumentalist attitude towards scientific hypotheses is interpreted in this paper as a choice that scientists themselves have to face, rather than as a philosophical problem. This decision will be justified by pragmatic reasons, and I shall discuss it with the help of two different conceptual tools: a deflationary semantics grounded in the inferentialist approach to linguistic practices developed by some authors (e.g., Sellars, Brandom), and an epistemic utility function that tries to represent the cognitive preferences of scientists. The first tool is applied to two different questions traditionally related to the problem of scientific realism: the non-miracle argument, and the continuity of reference. The second one is applied to the problem of unconceived alternatives, and to the distinction between realism and instrumentalism.

**1. Introduction**

The main thesis defended in this paper is that, if there is something deserving to be classified under to label of ‘the problem of scientific realism’, it is *primarily* a scientific problem, rather than a philosophical one. Note that I am not claiming that realism is *not* a philosophical problem, only that its status as a scientific problem is more basic, at least in the sense that the answers given to the scientific problem of realism will be much more relevant for the possibility of answering the related philosophical questions than viceversa. By a ‘scientific problem’ I mean basically two things: First, I claim that realism is a (real) concern for scientists themselves, i.e, it is an important question *for them* to decide whether scientific theories, models, laws, etc., have to be interpreted in a realist or in a non---realist way; realism, hence, is taken here as a *pragmatic attitude* scientists may have or not have towards specific scientific claims; of course, not necessarily all scientists will have the same attitude towards the same theories, nor every single scientist will necessarily have the same attitude towards different theories. Second, the ways of answering those questions are part of what we typically call ‘scientific research’, so, they are just scientific questions like any other. Of course, nothing of this entails that the problem of realism has no interest from a philosophical point of view.

By a ‘pragmatic attitude’ I also understand two different, but related things; First, though in the rest of the paper I will have to say something about semantics (e.g., about the notions of truth and reference), I will use an approach to the philosophy of language (in particular, that of Robert Brandom, 1994) which considers that pragmatics is more fundamental than semantics, i.e., that the explanation of semantic distinctions and concepts has to do, in the end, with what speakers want to *do* with language; the semantics I am going to use is, in this sense, a ‘deflationary’ one, more concentrated in some pragmatic aspects of language than with the presumption that semantics can give us a deep answer to the ‘problem of the connection between

language –or mind– and (external) reality’. Second, I will try to illuminate some of scientists’ *decisions* about how to interpret a scientific claim, and it is important not lose sight of the fact that they are *decisions*, whose rationality will not only depend on the scientists’ beliefs, but also on their *interests, values, or preferences*. In particular, I will defend the use of ‘rational choice’ models to illuminate some of those types of decisions.

## 2. A semantic deflationist view of scientific realism.

### 2.1 *Semantic deflationism.*

A theory about semantics is deflationist if it claims that semantic notions can be sufficiently explained without committing ourselves to some or other position about metaphysical problems. One famous deflationary theory in semantics is Paul Horwich’s minimalist theory of truth, which is based on the trivial (Aristotelian, Traskian) point that proposition ‘p’ is true iff p, but adds that this is *all* that is needed to explain the role of the truth predicate.<sup>1</sup> Another interesting deflationary approach is Robert Brandom’s expressivism, which explains the role of semantic terms like ‘true’ or ‘refers’ by the *expressive* capacity they give to those languages possessing them, i.e., by the things we can say thanks to those expressions and that we could not say without them (e.g., “all logical consequences of true axioms are true”)<sup>2</sup>.

### 2.2 *What are we explaining when we explain the success of science?*

In the last decades, the most popular philosophical defence of scientific realism has been what is known as the ‘no miracle argument’ (NMA)<sup>3</sup>. Though there is a range of different interpretations of NMA<sup>4</sup>, it typically asserts that scientific realism is ‘the best explanation’ (or perhaps, the only reasonable one) of the ‘success of science’, or more particularly, of the empirical success of modern scientific theories. I admit that the argument is intuitively compelling, and my own discussion will probably not contradict it strictly speaking, but I will try to show that, when considered from the point of view of a deflationary semantics, NMA transforms itself in something close to a trivial *scientific* claim.

Let’s start by considering what is that NMA tries to be an explanation *for*. In the most compelling cases, what it tries to explain is not the ‘general success’ of modern science (i.e., how it is that *we* have managed to develop a so successful science), but, more specifically, the tremendous *empirical* success of *some* theories, especially in the natural sciences, more particularly in physics. Defenders of NMA claim that it would be almost impossible that those theories made so many and so good predictions if they were not true, or at least, very approximately true. But, what are we exactly explaining when we explain ‘the empirical success of a theory’? What kind of *fact* is our *explanandum*?

---

<sup>1</sup> Horwich (1990).

<sup>2</sup> Brandom (1994), ch. 5.

<sup>3</sup> Putnam (1981).

<sup>4</sup> See, e.g., Psillos (1999, ch. 10), Worrall (2007), and Frost-Arnold (2010).

Let  $T$  be the theory whose empirical success we want to explain, and let  $E$  be the proposition (or conjunction of propositions) about the world that constitute the empirical evidence on which  $T$  is assessed. For simplicity, let's suppose that  $T$  explains  $E$  perfectly, i.e., that  $T$  logically entails  $E$ , and that actually  $E$  has been derived from  $T$  before the truth of  $E$  has been established, so  $T$  not only explains  $E$  but *predicted* it<sup>5</sup>. The *fact* that  $T$  is 'predictively successful' seems to consist in the conjunction of the following propositions (each of them logically independent of the others):

- (1)  $T$  entails  $E$
- (2)  $E$  was deduced from  $T$  before knowing whether  $E$  is true
- (3)  $E$  is true

Hence, in order to explain the 'success' of  $T$ , we shall have to offer an explanation of these three facts. However, (1) is just a 'logical fact': there is no difference between 'explaining why'  $T$  entails  $E$  and just *proving* that  $T$  entails  $E$ ; there is nothing like a 'substantive' explanation in explaining (1), i.e., an explanation that has to do with how the world is, or anything we can conceive as related to 'the problem of scientific realism'. Regarding (2), it can be decomposed into two different claims:

- (2.a)  $E$  has been deduced from  $T$
- (2.b) That deduction was made before knowing whether  $E$  was true.

As in the case of (1), the fact (2.a) seems to have little to do with a substantive explanation in the sense necessary for assessing the NMA; at most, it is a psychological or historical fact about the specific people who carried out the deduction of  $E$  from  $T$ , and about the evolution of the mathematical or logical technics that allowed to perform it. The explanation of (2.b) (i.e., the explanation of why the fact referred to in (2.a) took place in a certain moment instead of another) seems also not to have anything to do with whether  $T$  or  $E$  are true: take into account that to explain (2.b) is *just* to explain why  $E$  was deduced from  $T$  at moment  $t$ , and possible answers to that question would be things like "because  $T$  had been invented before  $t$ , and  $E$  had not been tested before  $t$ ", or something like that. Hence, the possible explanations of (2.b) will refer to when some propositions and their truth became 'available' to the scientists, and so, something like 'the truth of  $T$ ' (which is the *explanans* favored by NMA) seems to play no role at all in the explanation of (2.b). At most, the fact (2), including its two components, would demand an explanation that belongs to the *history* of science, not to philosophy.

Lastly, we have the fact (3), i.e., the fact that  $E$  is true. What would be an explanation of *that*? Here is where our semantic deflationism reminds us that 'the fact that  $E$  is true' is just another way of expressing exactly the same fact that proposition ' $E$ ' expresses. Explaining (3) is just the same thing as explaining  $E$ . But

---

<sup>5</sup> Of course, there can be lots of discussions about what is to be a 'good explanation' and its connection to predictions, but they will distract us from the specific point I want to make, so I shall use in my argument the most naïve version of the nomologico-deductive schema of scientific explanation and prediction.

for explaining *E* we don't need any philosophical theory; what we need (and scientists pursue) is a scientific theory like *T*, one of whose goals is, obviously, to explain *E* (i.e., to explain fact nr. 3).

We find out something similar when we reflect on the usual way realists express, not the *explanandum* of NMA, but its purported *explanans*: it is, they say, the *truth* of *T*, or the fact that *T* is true (or approximately so), what 'explains' its empirical or predictive success. But from our deflationary semantics, there is no difference at all between 'the fact that *T* is true' and those facts about the world that we express in affirming *T* itself. Hence, 'that *T* is true' is just a different way of saying *that T* (when *T* is not taken as the *name* of the theory, but as the conjunction of its axioms, principles or hypotheses). Hence, by saying that 'the truth of *T* explains its empirical success', or that 'the fact that *T* is true explains why *T* is empirically successful', we are just expressing in a slightly more complicated way what we can express just by saying that *T explains E*, for '*T*' asserts exactly the same as 'the fact that *T* is true', and '*E*' asserts basically the same as 'the empirical predictions or *T* are true' (if, as it is assumed in the argument, they are). In a nutshell: *explaining the empirical success of a scientific theory is just what the theory itself does, if it is empirically successful.*

I want to insist in that I am not claiming that there is nothing like 'explaining the empirical success of a theory'. What I am saying is that, when we analyse what that 'explaining' can consist in, we find that it is just the same thing as what the theory does with the empirical data it explains, if it is empirically successful. Hence, in order to assess whether 'the truth of *T*' is a good explanation of 'the empirical success of *T*', the only thing we could do is to see whether *T* is a good, or appropriate, or acceptable explanation of *E*. But *this* is a scientific problem, i.e., a problem *for scientists* to solve, not a 'philosophical' one<sup>6</sup>. I am also not denying that there is something important to the intuition that 'it is very unlikely that *T* is not (at least approximately) true, given how empirically successful it is', nor to the idea that (novel) predictions are a better reason to accept the truth of the theory than (ex post) accommodations of already known facts; I will say something else about this in the section 3. But my point is that these intuitions are better understood as something pertaining to the process of scientific research per se, rather than to some philosophical analysis.

### **2.3 The continuity of reference.**

Another common topic in the realism debate is that about the existence of 'theoretical entities', and in particular, the 'continuity' or 'discontinuity' of the reference of theoretical terms as theories change and are replaced by others. Typically, anti---realists have tried to show that many, if not most theoretical entities hypothesised by overcome scientific theories did not exist, even if they were necessary to generate those theories' successful empirical predictions<sup>7</sup>. A common line of defence of realist philosophers has been to indicate that, even if past theories have been refuted and their theoretical terms have been shown to be

---

<sup>6</sup> Of course, philosophers can discuss about what an explanation consists in, what is its relation to prediction, etc., etc. But this is (at best) an additional clarification of the work of scientists, whereas 'explaining the success of *T*' is, if my analysis is right, exactly *that* work, not a clarification of it.

<sup>7</sup> This is the famous Laudan's 'pessimistic meta-induction'; cf. Laudan (1981).

non---denoting, or their meanings have considerably changed, there is nonetheless a strong *continuity* between successive theories<sup>8</sup>. What is understood by this continuity is that something important is ‘preserved’ in the process of passing from one theory to another.

Structural realists affirm that what is preserved is ‘structure’: things like equations, symmetries, or any other type of *abstract* forms. Entity realists, on the other hand, affirm that besides the conservation of some structural features, also the *reference* or the denotations of some theoretical terms are often preserved to some extent.

Moreover, it may be argued that many ‘overtaken’ scientific concepts can be reinterpreted as more or less erroneous (and hence, more or less right) descriptions of entities that we take as really existing<sup>9</sup>. For example, phlogiston has been interpreted as a kind of rudimentary account of electrons, and the caloric fluid has been taken as a naïve description of molecular heat. I will not enter here into historical discussions, but will try to approach the problem in a purely abstract way from the point of view of our semantic minimalism. This suggests that, instead of considering the question of whether theoretical terms ‘really’ refer or not, we should consider first of all what is what the speakers (in this case, scientists) want to express *by choosing one way of speaking or another* about those theoretical entities.

Imagine the following simplified, ‘Sesame Street’ situation. We have two scientists, Ernie and Bert, that employ several predicates P, R and S, with which they can make different assertions. Let ‘Ci(X)’ stand for ‘scientist *i* claims that *X*’ (for the sake of simplicity, I will not make any difference between ‘claiming’, ‘knowing’ or ‘believing’), and let ‘#xFx’ stand for the definite description ‘the *x* such that *Fx*’. Suppose that Ernie and Bert assert the following:

$$(4) C_e(\exists x(Px \& Rx \& \neg Sx))$$

$$(5) C_b(\exists x(Px \& \neg Rx \& Sx))$$

That is, Ernie affirms that there is something that has properties P and R, but not S, whereas Bert affirms that there is something that has properties P and S, but not R. The interesting question is, of course, whether those ‘things’ (the *x*’s) are ‘the same’ or not. From Ernie’s point of view there are two possibilities:

$$(6) C_e(\exists x(Px \& Rx \& \neg Sx) \& C_b(\exists y(Py \& \neg Ry \& Sy))) \&$$

$$\& \#xC_e(Px \& Rx \& \neg Sx) = \#yC_b(Py \& \neg Ry \& Sy))$$

$$(7) C_e(\exists x(Px \& Rx \& \neg Sx) \& C_b(\exists y(Py \& \neg Ry \& Sy)))$$

$$\& \#xC_e(Px \& Rx \& \neg Sx) \neq \#yC_b(Py \& \neg Ry \& Sy))$$

Both (6) and (7) coincide in affirming that Ernie believes the same as in (4), and furthermore, that he believes what (5) affirms about Bert. The difference between them is that, according to (6), Ernie thinks that the entity

<sup>8</sup> See, e.g., Psillos (1999, ch. 12) for a general discussion, and also Worrall (1989), Ladyman (2007), Bueno (2008) and Bartels (2010) for some different versions of the argument.

<sup>9</sup> And viceversa; cf. Chang (2012).

of which Bert believes what is expressed in (5) is the *same* entity of which Ernie believes what is expressed in (4), whereas, of course, in the case of (7) Ernie thinks that these are two *different* entities<sup>10</sup>.

One simpler way of expressing the same as in (6) would be<sup>11</sup>

$$(8) C_e(\exists x(Px \& Rx \& \neg Sx) \& C_b(Px \& \neg Rx \& Sx))$$

That is, Ernie thinks that there is something with has properties P, R, and not S, and *of which* Bert believes that has properties P and S, but not R. However, (7) does not admit such a simplification, but at most the following one

$$(9) C_e(\exists x(Px \& Rx \& \neg Sx) \& C_b(\exists y(Py \& \neg Ry \& Sy))) \& x \neq \#y C_b(Py \& \neg Ry \& Sy))$$

Let's apply these schemata to some examples. If P, R and S stand for 'is the minimal unit of a chemical element', 'is decomposable into smaller particles', and 'is perfectly spheric', a proposition like (8) and (9) would assert that Ernie (say, a contemporary scientist) believes that there are minimal units of chemical elements, that these units are decomposable into smaller particles, but that they are not perfectly spherical, whereas Ernie knows that Bert (say, a mid 19<sup>th</sup>-century chemist) thinks that there are indeed minimal units of chemical elements, but that they are perfectly spherical though not decomposable into smaller units. The difference between (8) and (9) would be that, in the case of (8), Ernie thinks that the entities accepted by Bert *are the same* entities accepted by Ernie, whereas in (9), of course, Ernie thinks that the entities Bert accepts *are not* the entities accepted by Ernie.

Probably, most contemporary scientists would opt for an interpretation like that suggested by (8): the atoms those 19<sup>th</sup> chemists that embraced atomism talked about *are* the atoms 'we' (contemporary scientists) talk about, though those chemists *attributed* to these atoms some different properties from those we attribute to them.

Let's consider a different example, in which P stands for 'is an infectious agent causing Bovine Spongiform Encephalopathy (BSE)', R stands for 'is a virus', and S stands for 'is a prion'. In this case, according to (8) and (9), Ernie would think that what causes BSE is a virus, not a prion, whereas he knows that Bert thinks that what causes BSE is a prion, not a virus. Additionally, (8) asserts that Ernie thinks that the entities Bert takes as causing BSE are the same ones that Ernie takes as causing BSE, whereas according to (9), Ernie thinks they are not the same entities. In this case, it seems more natural to interpret the disagreement between Ernie and Bert through a proposition like (9): Ernie thinks that the prions imagined by Bert *just do not exist*, for, of course, prions cannot be 'identical' to viruses.

But, what is the *real* difference between both cases? It cannot be a *formal, structural* difference, because our two examples are totally analogous from a formal point of view. However, it seems that it is not also a difference having to do with the '*causal capacities*' of the entities or systems under study, or with our capacity

<sup>10</sup> See again Brandom (1994), ch. 5, for an analysis of the pragmatics of referential terms.

<sup>11</sup> Technically, (6) logically follows from (8), but not viceversa, but this is because of some technical reasons about doxastic logic that are not relevant for my argument.

of '*manipulating*' those entities or systems: after all, it can be the case that there is a mechanical procedure to 'isolate' in a flask the infectious agents of BSE, and Ernie and Bert may agree that the flask contains such entities... only that each of them thinks that the entities the other believes that are contained in the flask do not really exist (e.g., there is not 'a BSE virus').

I think there is simply no way to solve the dilemma of choosing between (8) or (9) by applying a logical or philosophical theory of 'reference'. This, after all, is just a consequence of the 'ontological relativity' and 'inscrutability of reference' suggested by Quine, which is particularly conspicuous in the type of second-order intensional contexts we are examining (what one thinks about what other people think). But the fact is that real scientists *do* choose something like (8) in some cases and something like (9) in some other cases, and often in a pretty spontaneous way, without feeling the need of having a philosophical theory to guide their choice. My suggestion is that the decision about how much historically continuous is the reference of scientific terms in a particular context depends more on *pragmatic* reasons than on *semantic* ones. The belief of 'being talking about the same' is obviously necessary in any process of research, including conversation and con-troversy: if you and I are going to rationally debate about something, I will have to assume that *some* of the entities you are talking about *are* entities I am talking about (e.g., we both accept we are talking about the same flask). Scientific consensus will tend to enlarge the set of 'shared' entities, both as a consequence of the working of other consensus forming strategies, and as an efficient strategy in itself to promote intersubjective discussion: in cases where there are strong controversies, it is wise to make a previous effort to determine 'what are we talking about'. But these strategies will work differently in different contexts, and they will also manifest a certain degree of hysteresis: once certain criteria or methods of 'identifying' entities have been established within a field, their application will lead to accepting some entities that might have not been accepted if different methods had been established. But, on the other hand, competition may in some cases lead some researchers to emphasize the differences between their models and theories and those of their rivals more than the similarities. All these reasons explain why we may have cases like Ernie's and Bert's, where they end accepting (8) in one case and (9) in the other case, in spite of there being no *substantial* difference, neither from a formal nor from a causal point of view between both cases. This does not mean that 'ontological continuity' is not an important issue in the history of science. It only means that it is a problem that we must not think we can solve thanks to a philosophical theory, or at least, thanks to an aprioristic metaphysics, semantics or epistemology. The most similar thing to a 'philosophical' explanation of why certain continuities are accepted in the history of science and why others are not, will probably be something like an applied theory in philosophical pragmatics.



### 3. Realism and the aims of scientists.

#### 3.1 *The scientist's utility function.*

The main thesis I want to explore in the last part of this paper is that having a realist or an instrumentalist *attitude* towards a particular scientific theory, model, hypothesis, etc., is a problem that flesh and bone scientists face. To understand how scientists may decide to do one thing or the other, we need to have some insight about their preferences, or, as economists and rational choice theorists say, about their 'utility functions'. A big part of the philosophical discussion about 'the aims of science' has been carried out under the presupposition that philosophers may have the clearest ideas about what these aims or goals *should* be (after all, have not been 'knowledge', 'truth', etc., the topics of philosophical investigation for centuries?), and then perhaps some 'recommendations' about appropriate methodological practices could be derived from those philosophical theories, to be given to 'mortal' scientists as a kind of heavenly gift, if they want their practices to be really conducive to the *appropriate* epistemic goals. My point of view will be rather different: I will take for granted that scientists are society's best experts in knowledge (latin: *scientia*), as cyclists are the best experts in cycling, or fishermen in fishing, and I will try to develop a simplified model with which we may try to understand why scientists do what they do in their pursuit of knowledge, in an analogous way as to economists produce simplified models that try to understand why economic agents do what they do, and what consequences follow from that. The philosopher, hence, should not be more (nor less) patronising about the epistemic value of what scientists do as economists are about the valuations of goods and services done by consumers, entrepreneurs, voters or politicians: in epistemology, as in economics, our motto should be that *de (epistemicibus) gustibus non est disputandum*. Hence, the 'utility functions' I am going to present are not intended to stand for some aprioristic idea about, say, how epistemic evaluations *should* be, but just as a *hypothetical, empirical* reconstruction of the *real* preferences scientists have on epistemic matters. Of course, my models will be utterly simplistic (like, by the way, most economic models are), but I hope that even under the limits of their simplicity they can be useful enough to illuminate some interesting aspects of science.

I have argued elsewhere<sup>12</sup> that the behaviour of scientists can be explained to a high extent by assuming they have a utility function with two main components, a 'social' one and an 'epistemic' one. The social element would consist in *recognition*, which basically depends on how important your contributions are judged by other scientists<sup>13</sup>. The epistemic element, on the other hand, would consist in some function allowing a researcher to judge how good is a particular scientific claim on the light of the available evidence. Many such functions have been proposed in the literature, but I have defended that a notion I called

---

<sup>12</sup> See, e.g., Zamora Bonilla (2002).

<sup>13</sup> The social part of the utility function may also have some other components -power, income, social benevolence...-, but these don't seem to be very relevant in the types of cognitive contexts that are usually interesting from the point of view of philosophy of science, and, what is more important, many of them usually depend in a direct way on the recognition level attained by a scientist.

‘empirical verisimilitude’ has bigger explanatory power, i.e., seems to explain more common patterns of theory evaluation than others. Before entering into details, it is important to notice that the two main elements of the utility function are not so disparate as it may seem, for, after all, ‘recognition’ is essentially ‘recognition *for*’, i.e., for making ‘important’ contributions, and how good and important a contribution is viewed by your colleagues will depend in the end on how they evaluate it according to the *epistemic* component of *their* utility function.

The social element of a researcher’s preferences will play almost no role in the rest of my argument (besides what I said at the end of the previous section in connection to scientific competition), so I will concentrate on the epistemic part<sup>14</sup>. My suggested definition of the concept of *empirical verisimilitude of a hypothesis H on the light of evidence E* consists in the combination of a high degree of coherence or similarity between H and E, and a high degree of coherence or similarity between E and the whole truth (W). This notion of coherence between two propositions X and Y is modelled as  $p(X\&Y)/p(XvY)$ ,<sup>15</sup> where  $p$  is assumed to be a typical subjective (prior) probability function, which can be different for different scientists. Hence,

$$\begin{aligned} (10) \quad V_s(H,E) &= [p(H\&E)/p(HvE)] [p(E\&W)/p(EvW)] \\ &= [p(H\&E)/p(HvE)] [p(W)/p(E)] \quad (\text{since } E \text{ is assumed to be true}) \\ &= [p(H\&E)/p(HvE)p(E)] [p(W)] \\ &\propto p(H,E)/p(HvE) \quad (\text{since } p(W) \text{ is a constant})^{16} \end{aligned}$$

In the particular case when H correctly explains or predicts E (i.e., if H entails E), this leads to:

$$\begin{aligned} (11) \quad V_s(H,E) &= p(H,E)/p(E) \\ &= p(H)/p(E)^2 \end{aligned}$$

However, in the case when H is fully confirmed by the evidence E (i.e., if E entails H), its value is:

$$\begin{aligned} (12) \quad V_s(H,E) &= [p(E)/p(H)p(E)] \\ &= 1/p(H) \end{aligned}$$

In the case when the empirical evidence contradicts the theory (and hence  $V_s(H,E) = 0$ ), there are however some ways of using  $V_s$  to represent the epistemic preferences of scientists. For example, the epistemic value

<sup>14</sup> As I have argued elsewhere (Zamora Bonilla, 2002a), pursuit of recognition within a research community demands that its members agree on some stable norms about how to assess the epistemic valuation of their claims, and hence, scientists have an incentive to minimally agree on these norms before they know what particular problems and solutions will be studied or proposed by each one, establishing the most general rules for epistemic evaluation “under a veil of ignorance”, so to say. This justify to consider the epistemic part of their utility function as being more explanatory of their agreed epistemic norms than the pursuit of recognition. The latter will enter into the explanation of scientific norms not by helping to define ‘the scale of epistemic quality’ (so to say), but the point of the scale that will be used as the threshold separating acceptable claims from unacceptable ones, for the choice of this point determines not *what* is pursued in the game science (this will be specified by the epistemic utility function), but just *how difficult* playing the game is, and hence, what level of recognition can a player expect a priori.

<sup>15</sup> I presented this definition in Zamora Bonilla (1996). For more details, see Zamora Bonilla (2013). Independently, the formula  $p(X\&Y)/p(XvY)$  became later a standard definition of coherence, after Olsson (2002).

<sup>16</sup> For simplicity, the symbol ‘ $\propto$ ’ will be replaced by ‘=’ in the reminder of the paper.

of  $H$  can be given by  $V_s(H, E(H))$ , where  $E(H)$  can be the conjunction of established empirical laws explained by  $H$ , or the conjunction of the most approximate version of each empirical law such that  $H$  still entails it.

### 3.2. *Unconceived alternatives.*

According to our model, scientists intuitively assess their hypotheses on the basis of a function like  $V_s$ . Since usually the theories are attempted to explain the relevant empirical data, rather than being logically confirmed by them,<sup>17</sup> formula (11) will be the more useful one. Let's now consider by means of it the typical case in the discussion about the problem of *unconceived alternatives*:<sup>18</sup> although we may have one or several theories correctly explaining or predicting the known empirical facts, it can still be the case that these theories are false, and not only false, but very far from the truth. We can derive from (11) some interesting lessons in connection to this problem. In the first place, the formula entails that, of several theories rightly explaining  $E$ , the one with the highest prior probability (which in this case is equivalent to having a higher posterior probability) will be most valuable; so, a theory correctly explaining the facts will be better for a scientist, the more plausible a priori it seems to her. Note that it is the personal, subjective estimate of the probability of each theory being true what counts in our model; hence, assuming that the scientists working about the problems  $H$  is trying to solve are the *best* positioned people to judge those probabilities, there will be nothing that we philosophers may *add* to the probability judgments scientists make about this question. Suppose that  $H$  is the only *conceived* theory successfully explaining  $E$ ; this means that  $p(\neg H, E)$  is literally the probability that, according to the particular scientist whose subjective probability function is given by  $p$ , there is some *unconceived* theory which is true. To what extent will this be a problem for the *acceptability* of  $H$  from that scientist's point of view? It depends in part on the relative magnitude of  $p(H, E)$  and  $p(\neg H, E)$ . The philosophers that have criticized scientific realism on the basis of the 'pessimistic induction' argument seem to base their criticism on the idea that  $H$  is 'very' probably false, but scientists can indeed think that  $p(\neg H, E)$ , though 'high', is not *too* high to preclude their acceptance of  $H$ ; this will depend, in any case, on the scientists' judgments, not on the philosophers'.

Furthermore, (11) does not say anything in principle about how probable it is for a scientist that  $H$  is *approximately* true (i.e., that the 'world' corresponding to proposition  $W$  is 'close' to the set of worlds consistent with  $H$ ): she can accept that  $p(\neg H, E)$  is 'very' high, but also that, so to say,  $p('W$  is close to  $H', E)$  is even higher than that. So, there is nothing in the arguments about the pessimistic induction and unconceived alternatives that *forces* scientists to conclude that the posterior probability of an empirically successful

---

<sup>17</sup> Actually, as John Norton (2014) has argued, scientists often *try to prove* their theories from previously known empirical facts or laws (for example, Newton offered a 'demonstration' of the law of gravity taking Kepler's laws as premises), but if this were true at face value, it would entail that theories are not falsifiable (at least, while the empirical laws from which they are mathematically deduced are not rejected), nor can do more predictions than those derivable by the previous empirical laws alone. I will offer below a different interpretation of this kind of 'demonstrative' arguments by scientists.

<sup>18</sup> Stanford (2006).

theory is ‘too low’ to make its acceptance irrational, or to take it as a good approximation to the truth. Furthermore, and perhaps more important for our discussion: what formula (11) also allows to see is that, of those theories successfully explaining the available empirical evidence, scientists will tend to prefer those that have a higher probability, i.e, a higher probability *of being true*, and hence, *truth* must be taken as one goal of scientists (at least to the extent that Vs rightly describes, albeit in a simplified way, their epistemic preferences).

Second, (11) also let us see that the epistemic value of H given E does not only depend on H’s probability: it also depends on *how unlikely E is*. This means that, given two theories (H and H’) and two different bodies of empirical evidence (E and E’), such that H correctly explains E, and H’ correctly explains E’, it *can* be the case that H is judged better on the light of E than H’ on the light of E’, even if  $p(H,E) < p(H',E')$  and  $p(H) < p(H')$ . This is reasonable, because in order to explain an increasingly more and more exhaustive body of empirical evidence, we will tend to need theories that are stronger and stronger, and hence, more and more improbable. Hence, if E is very contentful, so that  $p(E)$  is very low, the verisimilitude of a theory H explaining E can be ‘very’ high even if both  $p(H)$  and  $p(H,E)$  are ‘very’ low. We can visually interpret this in the following way: if the set of worlds consistent with the empirical evidence E (i.e., the set of E-worlds) has become very small thanks to the addition of many strong empirical laws, then, even if an empirically successful theory H is literally false, the true state of facts will not be ‘very far from’ the H---worlds, just because there is no much space within the set of E-worlds.

In the third and last place, (11) and (12) allow to calculate something like a minimal threshold of acceptability: we may argue that the worst ‘right solution’ we might give to a scientific problem would be to answer it with a tautology, i.e., with a proposition that does not assert absolutely anything about the world.<sup>19</sup> Since any tautology Taut is entailed by any body of empirical evidence E, the verisimilitude of the former will be, according to (12):

$$(13) \text{Vs}(\text{Taut}, E) = 1/p(\text{Taut}) = 1$$

Hence, for a theory H that *entails* the evidence E, having a verisimilitude *lower* than 1 will be a reason to discard it: a non-answer, like Taut, would be epistemically preferable to H. According to (11), the condition for H having a verisimilitude *higher* than one, if H entails E, is:

$$(14) \text{Vs}(H, E) > 1 \text{ iff } p(H, E)/p(E) > 1 \\ \text{iff } p(H, E) > p(E)$$

---

<sup>19</sup> I do not deny that logico-mathematical truths play an important role in scientific argumentation and discovery, but I don’t think they can be identified with ‘theories’ in any relevant sense, at least when they are taken in isolation. They are, at most, important elements of theories or research programmes.

This result has a nice interpretation: in order to be acceptable, a necessary (but by no means sufficient)<sup>20</sup> condition a theory that successfully explains the empirical evidence must fulfil is that its own *posterior* probability must be higher than the *prior* probability of the evidence; or, stated differently, we must not accept an explanation which is *so unlikely* that, even taking into account the evidence, its truth would be less probable than the prior probability of having found that evidence.<sup>21</sup> According to this result, scientists would be interested in employing arguments that can show two things: first, that *the empirical facts their theories manage to explain or predict are very unexpected* (i.e., that  $p(E)$  is very low); and second, that *their theories are relatively plausible* (i.e., that  $p(H)$ , and hence  $p(H,E)$  is very high). And of course, each researcher can also try to do the opposite in connection with the theories of her competitors. These judgments of plausibility can take any form, from quantitative estimations of probability or improbability, to mathematical ‘proofs’ of the theories’ principles from some empirical laws (plus some more or less ‘innocent’ assumptions)<sup>22</sup>, and also to mere ‘rhetorical’ or ‘philosophical’ arguments trying to persuade the reader of the likelihood of some principles. In any case, all these arguments can be interpreted as attempts to establish that it is not very likely that the researcher’s preferred theory is false, or very far from the truth, even though some ‘unconceived alternatives’ can still be closer to the truth.

### 3.3. Why to be an instrumentalist?

From our previous discussion, it seems that the scientists’ (epistemic) utility function entails that they will necessarily have a realist attitude towards their theories: after all, what they want to prove is that the theories they defend are plausibly *true* or close to the truth, and that the theories of their competitors are false or far from the truth. Is it, hence, always irrational for them to have an instrumentalist attitude? And by the way, what would ‘an instrumentalist attitude’ consist in? A plausible answer to the latter question is that, if we have identified realism with being concerned for the probability of the truth (or approximate truth) of the competing theories, then having an instrumentalist attitude would amount to *not being concerned* by that probability; i.e., one would reveal an instrumentalist attitude towards a theory, model or hypothesis  $H$  if one is willing to give it a *high* epistemic value because it explains or predicts well the available evidence, *even if one acknowledges that the probability of  $H$  being true or approximately true is very low*. The question is, hence, are there circumstances where a researcher whose epistemic utility function is represented by  $V_s$  would assess theories in an instrumentalist way? Luckily, it is easy to show that there are.

---

<sup>20</sup> I have argued elsewhere (Zamora Bonilla, 2002a) that it is the ‘social’ part of the utility function, in addition to the epistemic part, what can be taken into account by scientists in order to establish a *sufficient* criterion of acceptability.

<sup>21</sup> This is a rewording of the idea that to accept extraordinary claims we need extraordinary evidence.

<sup>22</sup> This would be my suggested interpretation of Norton’s claim I have referred to above, according to which scientists often try to ‘prove’ their theories, not only to test them.

In the previous subsections I have assumed that the relevant empirical evidence  $E$  is given and fixed, but obviously very often this is not really so. There may be situations where scientists consider that they already have all the necessary data to decide on the acceptability of the competing theories, i.e., that it is unlikely that the possible addition of new empirical findings is going to force them to reverse the judgments they have made on their theories. However, in many other situations this is not the case, and the relevant empirical data are just being *searched*, or at least awaited. Of course, this is more common in the first stages of a new research programme (to use a Lakatosian phrase), or when competition between different programmes is very strong. In cases like these, scientists will be *uncertain* about what the future empirical discoveries will be, and it seems reasonable that theories are not directly judged according to a function like  $V_s$ , but according to the *expected value* of that function. This expected value is easy to calculate:<sup>23</sup>

$$\begin{aligned}
 (15) \text{ EV}_s(H,E) &= \sum(w \in E)p(w,E)V_s(H,w) \\
 &= \sum(w \in E \& H)p(w,E)V_s(H,w) + \sum(w \in E \& \neg H)p(w,E)V_s(H,w) \\
 &= \sum(w \in E \& H)p(w,E)p(H,w)/p(H) + 0 \\
 &= \sum(w \in E \& H)[p(w)/p(E)][1/p(H)] \\
 &= [\sum(w \in E \& H)p(w)]/[p(E)p(H)] \\
 &= p(E \& H)/[p(E)p(H)] = p(E,H)/p(E) = p(H,E)/p(H)
 \end{aligned}$$

An immediate result deriving from (15) is:

$$(16) \text{ If } H \text{ and } H' \text{ entail } E, \text{ then } \text{EV}_s(H,E) = \text{EV}_s(H',E) = 1/p(E)$$

Hence, scientists whose epistemic utility function can be represented by  $V_s$ , but that are still relatively uncertain about how the new empirical evidence will affect the acceptability of the theories under discussion, will tend not to value these theories according to how plausible they currently are, but *only according to how good their predictions have been*. This result is consistent with Lakatos' thesis that, in the first stages of a research programme, scientists only care about confirmations, and not about falsifications;<sup>24</sup> i.e., they don't discard a programme because of its failure to explain or anticipate some empirical results, but value it only in function of its empirical successes, or, in other terms, they have an instrumentalist attitude towards it. Instead, once the empirical data are considered stable enough, i.e., not forcing to change the judgements over theories, these will be evaluated according to  $V_s$ , not according to its expected value, and hence considerations about the plausible truth of the theories will become important, or, to state it in a different way, scientists will start to have a realist attitude in their epistemic judgments of those theories.

<sup>23</sup> For simplicity, I will use the expression " $w \in E$ " as an abbreviation of "a point in the logical space that satisfies  $E$ ".

<sup>24</sup> Especially if we consider the sophisticated version of our utility function we discussed a few pages above, i.e., replacing  $V_s(H,E)$  with  $V_s(H,E(H))$ . Cf. Zamora Bonilla (2002b).

Furthermore, in the same way as the function  $V_s$  may be different for scientists with different subjective probability functions, it can also be the case that different scientists have different expectations about the evolution of the empirical data, and hence some tend to apply  $V_s$  or  $EV_s$  under somehow different circumstances. But again, it is *their* decision to do one thing or the other; there is nothing like a ‘philosophical solution’ to the question of what is what they *must* decide.<sup>25</sup>

### Conclusion

In this paper I have argued for the thesis that many of the problems associated to the scientific realism debate can be seen as problems that scientists themselves have to solve, or take a decision about, in their normal research practices. I have used two different conceptual approaches to illuminate the question. First, a semantic deflationism according to which truth and reference are not so much a kind of abstract relation between language (or mind) and world, as linguistic devices that help us to express some complex thoughts that would be difficult or impossible to express if natural languages didn’t contain predicates like ‘...is true’ or ‘...refers to...’. Ideas such that the ‘truth of a theory’ can explain its empirical success, or that different theoretical terms can ‘refer to the same’ entities appear to have a radically different meaning from the point of view of this minimalist semantics, as the one it is attributed to them in most philosophical discussions about scientific realism. Second, a hypothesis about the structure of scientists’ epistemic preferences has been used to discuss the question of what is the difference between having a realist or an instrumentalist attitude towards a theory, and why it can be rational to have one attitude or the other towards the same theories depending on the circumstances. I do not claim, of course, that scientific realism is not a deeply important philosophical problem. But I think that the analysis of the pragmatic aspects of the problem can help to frame in a more illuminating way its philosophical nuances. References:

### References

- Bartels, A., 2010, “Explaining Referential Stability of Physics Concepts: The Semantic Embedding Approach”, *Synthese*, 41: 267-81.
- Brandom, R. 1994, *Making it Explicit*, Cambridge (Mass.): Harvard University Press.
- Bueno, O., 2008, “Structural Realism, Scientific Change, and Partial Structures”, *Studia Logica*, 89:213-235.
- Chang, H., 2012, *Is Water H2O?*, Dordrecht: Springer.
- Douglas, H., and P.D. Magnus, 2013, “State of the Field: Why novel prediction matters”, *Studies in History and Philosophy of Science*, 44:580–589.

---

<sup>25</sup> Though it is not as directly relevant to the discussion about realism as the other questions examined so far in this paper, I would also like to mention that the hypothesis that  $V_s$  represents more or less correctly the epistemic preferences of real scientists can be used to make sense of the fact (mentioned in section 2.2) that prediction of unknown empirical results is more valuable than mere ‘accommodation’ of those previously known. The reason is that scientists would prefer, as we have seen, to develop new theories by starting with those whose assumptions seem most likely to them. Hence, accommodating an empirical fact will demand to replace some of the theory’s assumptions with another which is less likely, hence reducing the maximum value of  $V_s$  that the theory can get. This explanation is coherent with some other recent arguments about the superiority of prediction over accommodation, in particular those linking the former to the expectation of future empirical success (see Douglas and Magnus, 2013, for a survey).

- Frost---Arnold, G., 2010, "The No-Miracles Argument for Realism: Inference to an Unacceptable Explanation", *Philosophy of Science*, 77: 35–58.
- Horwich, P., 1990. *Truth*, Oxford: Blackwell.
- Ladyman, J., 2007, "On the Identity and Diversity of Individuals", *The Proceedings of the Aristotelian Society*, Supplementary Volume 81:23-43.
- Laudan, L., 1981. "A Confutation of Convergent Realism". *Philosophy of Science* 48:19-49.
- Norton, J.D., 2014, "A material dissolution of the problem of induction", *Synthese*, 191: 671-690.
- Olsson, E.J., 2002. "What is the problem of coherence and truth?", *The Journal of Philosophy*, 99:246-278.
- Psillos, S., 1999. *Scientific Realism: How Science Tracks Truth*, London: Routledge.
- Putnam, H., 1981, *Reason, Truth and History*, Cambridge: Cambridge University Press.
- Stanford, P.K., 2006, *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives*. Oxford, Oxford University Press.
- Worrall, J., 1989, "Structural Realism: The Best of Both Worlds?", *Dialectica*, 43: 99–124.
- Worrall, J., 2007, "Miracles and Models: Why Reports of the Death of Structural Realism May Be Exaggerated" *Royal Institute of Philosophy Supplements*, 82(61): 125–154.
- Zamora Bonilla, J. P., 1996, "Verisimilitude, Structuralism and Scientific Progress", *Erkenntnis*, 44:25-47.
- Zamora Bonilla, J.P., 2002a, "Scientific Inference and the Pursuit of Fame: A Contractarian Approach", *Philosophy of Science*, 69:300-23.
- Zamora Bonilla, J.P., 2002b, "Verisimilitude and the Dynamics of Scientific Research Programmes", *Journal for General Philosophy of Science*, 33:349-68.
- Zamora Bonilla, J.P., 2013, "Why are good theories good? Reflections on epistemic values, confirmation, and formal epistemology", *Synthese*, 190:1533-1553.



---

*Mathematical Evidence: Pure vs Applied*

JAMES ROBERT BROWN

UT

jrbrown@chass.utoronto.ca

---

Philosophers distinguish pure from applied mathematics by saying that pure involves only mathematical concepts while applied uses a mixture of mathematical and non-mathematical notions. A trivial example: “ $2+3=5$ ” is pure; “2 apples + 3 apples = five apples” is applied. Mathematicians, by contrast, often cite examples from physics (or biology or finance) and nevertheless call them instances of pure mathematics. Why this difference?

Philosophers are motivated by *epistemology*; they want to know if and how it is possible to justify claims that have no possibility of empirical content. Mathematicians, on the other hand, draw their distinction based on whether the mathematics is *interesting*. Thus, for example, General Relativity and Quantum Field Theory attract their attention for mathematical reasons. G.H. Hardy famously called applied mathematics “ugly,” but claimed that GR and QM are “real mathematics,” worthy of his attention.

With these rival distinctions in mind, we can ask whether philosophers are approaching the epistemic issues the right way? Can we learn mathematical facts by thinking about situations that involve non-mathematical entities?

Let me make a second distinction between pure and applied ethics. (I’m running several existing distinctions together, eg, metaethics, practical ethics, theoretical ethics, normative ethics, and so on.) Typically, discussions about abortion are instances of applied ethics; not surprisingly, they are about abortion. Discussions of the trolley problem, however, are not about runaway trollies; rather, they are about utilitarian principles, and as such they are instances of pure ethics. The aim of an abortion debate is a policy on abortion, but a debate on the trolley problem will not culminate in a public policy on runaway trollies that requires people to throw (or not to throw) a switch killing just one innocent person to save five.

In spite of the important difference between the pure trolley case and the applied abortion case, the methods of investigation are strikingly similar. Both use thought experiments. The trolley example is obviously a thought experiment, and in the abortion case, the most famous argument is based on Thomson’s thought experiment involving a sick violinist.

Ethical reasoning (at least some of it), whether pure or applied uses the same technique. The pure-applied ethics distinction is similar to the mathematicians’ pure-applied distinction. Not only do they make similar distinctions between pure and applied (unlike the distinction made by philosophers of mathematics), but they both appeal (if only implicitly) to intuitions involving imagined concrete situations.

This should come as no surprise, because in mathematics and ethics intuitions are of prime importance, since direct empirical evidence is more or less out of the question. (In the talk I will digress briefly on this important point.)

The main claim of the talk is this: The right model for philosophers concerned with the epistemology of mathematics should be ethics. Just as thought experiments can work in ethical reasoning, they can also work in mathematical practice. I will illustrate this claim with examples. These will be mostly simple examples that are accessible to a general audience, such as the pigeonhole principle.

The *pigeonhole principle* says: If there are  $n+1$  pigeons distributed in  $n$  pigeonholes, then at least one hole must have at least two pigeons. This principle which seems so obvious is a hugely important principle in combinatorial mathematics. In the philosophers' sense of applied mathematics, the principle implies that in a room of 367 people, there is at least one pair of people with the same birthday. And it implies that in Dusseldorf (population: 600,000) there are at least two people with the same numbers of hairs on their heads (typical full head of hair is 100,000 hairs). In the philosophers' sense of pure mathematics, the pigeonhole principle implies: If  $\{x_n\}$  is a sequence of real numbers lying in a bounded interval, then  $\{x_n\}$  contains a Cauchy subsequence. And it also implies: Every graph with two or more vertices has two vertices with the same degree. It is not difficult to prove the pigeonhole principle, but the important thing to notice is that a proof is not necessary for rational belief; we have sufficient evidence in the form of the obviousness of the principle itself in terms of pigeons and pigeonholes.

The kind of evidence that we possess for the pigeonhole principle is the same as we have in the ethics cases mentioned earlier. In each of these we reason from particular examples that are readily imagined to powerful general principles. It has the trappings of empirical experience, but this is misleading. It is in fact *a priori* evidence. The philosopher's distinction between pure and applied mathematics is perfectly legitimate and captures something objective. But it is not well-motivated. The more fruitful distinction between pure and applied, the distinction embraced by working mathematicians, does not correspond to a divide between empirical and non-empirical.

In conclusion, a case is made to liberalize how philosophy of mathematics sees pure mathematics and to broaden the realm of legitimate evidence in pure mathematics beyond standard proofs. It remains a challenge to account for how we are able to move from thinking about apparently empirical things such as pigeonholes to powerful general principles. I will close the talk with some brief speculations about this cognitive capacity we seem to possess.

---

*Absolute Measures of Effectiveness*

JACOB STEGENGA

University Of Utah

jacob.stegenga@utah.edu

---

Many ‘outcome measures’ are employed in clinical research. An outcome measure is an abstract formal statement describing a relation between the value of the measurand in the control group and the value of the measurand in the experimental group. When particular substantive values for such measurands are substituted into an outcome measure, the result is a quantitative estimation of the strength of an alleged causal relation—this quantity is usually called an ‘effect size’.<sup>1</sup> The results of clinical research are frequently reported with ‘relative’ outcome measures. Here I argue that relative measures promote the base-rate fallacy, and so results of clinical research should always be reported with ‘absolute’ outcome measures.

For both continuous and dichotomous parameters, the choice of outcome measure is important and can have significant influence on the estimation of effectiveness, and the basic issue I discuss below is salient for both kinds of parameters. However, the point can be made more simply by focusing on dichotomous parameters.

If the measured parameters are dichotomous (such as death), standard outcome measures include the odds ratio, relative risk, relative risk reduction, risk difference, and number needed to treat. To define these, construct a two-by-two table for a study that has an experimental group (E) and a control group (C), in which a binary outcome is measured as present (Y) or absent (N), where the number of subjects with each outcome in each group is represented by letters (a-d), as follows:

Group	Outcome	
	Y	N
E	a	b
C	c	d

Relative risk (RR) is defined as:

$$RR = [a/(a+b)] / [c/(c+d)]$$

Relative risk reduction (RRR) is defined as:

$$RRR = [[c/(c+d)] - [a/(a+b)]] / [c/(c+d)]$$

Risk difference (RD) is defined as:

---

<sup>1</sup> How an effect size relates to the strength of a causal relation is a tricky problem beyond the scope of the present paper. See (Broadbent, 2013) for a superb discussion of what he calls, in the epidemiological context, ‘puzzles of attributability’.

$$RD = a/(a+b) - c/(c+d)$$

Number needed to treat (NNT) is defined as:

$$NNT = 1 / [[c/(c+d)] - [a/(a+b)]]$$

It also can be useful to define these in terms of conditional probabilities. The probability of a subject having a Y outcome given that the subject is in group E,  $P(Y|E)$ , is  $a/(a+b)$ , and likewise, the probability of having a Y outcome given that a subject is in group C,  $P(Y|C)$ , is  $c/(c+d)$ . Thus, for example, we have:

$$RR = P(Y|E)/P(Y|C)$$

$$RD = P(Y|E) - P(Y|C)$$

A widespread and misguided practice is to report RR or RRR but not RD or NNT. The over-reliance on relative outcome measures in epidemiology is dubbed ‘risk relativism’ by (Broadbent, 2013). Broadbent canvasses several alleged justifications for the widespread use of relative measures like RR, and finds them all wanting. Here I add to this concern by noting a nefarious consequence of risk relativism.

Employment of relative measures, such as RR or RRR, promotes the base-rate fallacy, because relative measures do not take into account the baseline rates of the event in question (Worrall, 2010). Both physicians and patients overestimate the effectiveness of medical interventions when presented with only relative measures, and their estimates are more accurate when they are presented with both relative and absolute measures or with absolute measures alone.

To see that relative measures do not take into account the base rate of the outcome in question, consider RR. By Bayes’ Theorem, RR is equivalent to:

$$\begin{aligned} RR &= [P(E|Y)P(Y)/P(E)] / [P(C|Y)P(Y)/P(C)] \\ &= [P(E|Y)/P(E)] / [P(C|Y)/P(C)] \end{aligned}$$

The baseline probability of having outcome Y,  $P(Y)$ , has fallen out of the equation. RR is not sensitive to  $P(Y)$ . In contrast, absolute measures are sensitive to the base rate of the outcome in question. Consider RD. By Bayes’ Theorem, RD is equivalent to:

$$\begin{aligned} RD &= [P(E|Y)P(Y)/P(E)] - [P(C|Y)P(Y)/P(C)] \\ &= P(Y)[[P(E|Y)/P(E)] - [P(C|Y)/P(C)]] \end{aligned}$$

$P(Y)$  appears as a multiplier in RD. Thus RD is sensitive to  $P(Y)$ .

Thanks to the work of Kahneman and Tversky and others, we know that people reason poorly with prior probabilities. Therefore, since RD and other absolute measures take into account prior probabilities, whereas RR and other relative measures do not, RD should always be reported.

To illustrate the problem that arises when not taking  $P(Y)$  into account with relative measures of effectiveness, consider the drug alendronate sodium (Fosamax), marketed by Merck as causing an increase in bone density in women to avoid bone fractures. A large trial compared the drug to placebo over a four year period (Black et al., 1996). The evidence from the trial was touted as showing that the drug reduces the risk of hip fractures by 50%—this was a relative measure of risk reduction (RRR). However, as (Moynihan &

Cassels, 2005) note, only 2% of the women in the control group had hip fractures during the four years of the trial, while only 1% of the women in the experimental group had hip fractures. Thus the RD effect size was a mere 1%—the absolute difference in hip fracture rates between the experimental group and the control group was only 1%—after consuming the drug for four years.

(Worrall, 2010) rightly notes that the choice of using a medical intervention is a decision which ought to be modeled with an expected utility calculation. In this paper I give a further argument for the superiority of absolute measures along these decision-theoretic lines.

The reliance on relative outcome measures at the expense of absolute outcome measures is ubiquitous. This, together with the fact that people overestimate the effectiveness of medical interventions when provided with relative outcome measures, entails that on average people overestimate the effectiveness of medical interventions. Effectiveness always should be measured and reported in absolute terms (using measures such as RD). This would have the result that estimates of the effectiveness of medical interventions would be deemed lower than they now are.

## Symposia & Contributed Papers IV

**Émile Du Châtelet's Institutions de Physique**

**Symposium**

Organizer: John A. Hanson

Chair: Nina Retzlaff

Room 5F, Thursday 15:30 – 17:30

---

*Émilie du Châtelet on Newtonian Attraction*

JAMEE ELDER

University of Notre Dame

jamee.c.elder.9@nd.edu

---

*Du Châtelet's Philosophy of Space and Time*

ADRIANA M. SOLOMON

University of Notre Dame

asolomo1@nd.edu

---

*Du Châtelet on the Law of Continuity*

JOHN A. HANSON

University of Notre Dame

johnarndthanson@gmail.com

---

*PSR and the Problem of Force: The Metaphysical Grounding of Physics in Du Châtelet and Wolff*

JEREMY STEEGER

University of Notre Dame

jsteeger@nd.edu

---

*Substance and Change in the Institutions de Physique*

AARON WELLS

University of Notre Dame

awells5@nd.edu

---

**General Description**

Émilie du Châtelet's *Institutions de Physique* is a fascinating, though largely forgotten, text in the history of early modern science and philosophy. Presented as a series of physics lessons for her son, the *Institutions* attempts to critically wed the major themes of Newton's physics with the natural philosophy of Leibniz and Wolff. The resulting work is incredibly ambitious, presenting a picture of the physical world built from its metaphysical foundations up to a comprehensive view of a law-governed cosmos. En route, du Châtelet delves into many of the most challenging issues of her time, including the methodology of scientific enquiry, the nature of space and time, the metaphysical and physical composition of bodies, and the question of whether *vis viva* ought to count as a genuine measure of the force of bodies. In the process, she stakes out a unique position between the approaches of the empiricist disciples of Newton and the continental followers of Descartes, Leibniz, and Wolff.

Yet in spite of all this, *Institutions de Physique* remains a neglected topic of study. This is partly the effect of du Châtelet having also produced a still-respected translation of Newton's *Principia* into French, an accomplishment which has overshadowed much of her original work. It is also certain that the tawdry details of her relationship with Voltaire has done much to distract from her intellectual accomplishments, as did accusations of plagiarism on the part of her tutor, Samuel König. But it is also undeniable that, through many different routes and many different reasons, many of them explicitly sexist, du Châtelet was effectively written out of the standard history of natural philosophy of this time period, in spite of being widely read and recognized as a significant contributor to the debates of her time.

For all of these reasons, producing an accurate, thorough, and fair assessment of du Châtelet's work in the *Institutions* should be a concern for scholarship in early modern philosophy. Unfortunately, a major hindrance to such work in the Anglophone world is the lack of a complete translation of the *Institutions de Physique* into English. Over the course of 2014, our group has worked under the guidance of Prof. Katherine Brading to remedy this situation by producing a translation and begin expanding on the small critical literature surrounding her work. The exegetical tasks are considerable. Beyond the basic challenges of translation, du Châtelet's conceptual vocabulary differs considerably from our own, and stands in serious need of contextualization and rearticulation. Nonetheless, even at this early stage of this research, it is clear

that the study of Du Chatelet's works is a rich source for potentially important scholarship in early modern science and philosophy in Europe.

The proposed symposia will bring the early fruits of our work into public view, in the form of paper-based talks ranging over different aspects of du Châtelet's thought in the *Institutions*. The topics covered will include du Châtelet's view of the uses and abuses of hypotheses, her treatment of space and time, her criticisms of Newtonian views of gravitation, and her account of Leibniz's law of continuity. These talks should be of great interest to historians of early modern science and philosophy, and of use to philosophers of science interested in the 17th and 18th Centuries and the evolution of scientific practice during this period. Moreover, for those interested in the role of women in philosophy, Emilie du Châtelet and the *Institutions de Physique* provide an important case for consideration, both sociologically and content-wise. For all these reasons, we believe that this symposium would be an excellent addition to EPSA's 2015 lineup.

## Abstracts

### 1. Jamee Elder: *Émilie du Châtelet on Newtonian Attraction*

In this paper I will discuss arguments presented by Émilie du Châtelet in Chapter 16 of her *Institutions de Physique*, "On Newtonian Attraction". My focus will be on her two related arguments for the conclusion that "attraction, as the Newtonians propose it, that is to say, as far as we make it a property of matter and the cause of the majority of the phenomena, is inadmissible" (§394). In Chapter 16, du Châtelet examines the notion of attraction, conceived of as inherent to matter and as a *cause* in its own right. Both of her main arguments against this conception of attraction are based upon the *principle of sufficient reason* (PSR). In her first argument, she argues that two bodies at rest and separated by a void could have no sufficient reason to begin to move toward each other according to the laws of attraction. Her second argument is based on the variation in the direction and speed that result from attraction and relates to her views concerning the essence of matter. She argues that, by PSR, attraction cannot be a property of matter and that the phenomena of attraction require a mechanical explanation.

My objective in this paper is to carefully elucidate and analyze these arguments within the context of Du Châtelet's broader Wolffian-Leibnizian metaphysics, as outlined in the earlier chapters of the *Institutions*. Some key issues are her conception and application of PSR and her ideas concerning essences and continuity. I will also elucidate the consequences of this analysis for du Châtelet's physics and philosophy as a whole. The question of the status of attraction is at the juncture of du Châtelet's Leibnizian and Newtonian commitments, and is therefore a crucial moment at which to analyze the explicit interaction of these commitments.



## 2. Monica Solomon: *Du Châtelet's Philosophy of Space and Time*

Du Châtelet's *Institutions de Physique* was first printed in 1740 and, as the title suggests, it aims to provide a pedagogical introduction to the fundamentals of physics. My main aim here is to place du Châtelet's work in the history of arguments about the nature and properties of space and time as foundations for natural philosophy. I argue that, while du Châtelet's *Institutions* does not present a robust *theory* of space and time (a theory which would clearly present a coherent web of concepts in which those of space and time would be embedded), it contains a rich *philosophy* of space and time. First, I clarify some of the distinctions, definitions, and properties encountered in du Châtelet's chapters on space and time. Secondly, I delineate the sources for her arguments. For instance, it is clear that du Châtelet's view has a strong Leibnizian influence: space and time are not real (things). Space is nothing but the order or coexistent beings, while time is nothing other than the order of successive beings. But although Leibniz is a clear source, it is less clear that her main arguments for the unity underpinning the ideas of space and time (and which are presented by some of her most interesting arguments) are entirely Leibnizian. Thus, thirdly, in the main part of my paper, I present du Châtelet's particular type of abstractionism. For example, she argues that space and time arise from representations of how bodies coexist or of their successive states: space is to coexistent things just as numbers is to numbered things. In other words, only a multitude of things makes space necessary. The idea of time is conceived by analogy. The analysis will allow me to draw unexpected parallels between du Châtelet and well-known empiricist conceptualizations such as Hume's and Berkeley's.

## 3. John Hanson: *Du Châtelet on the Law of Continuity*

The Law of Continuity is one of the most fascinating, far-reaching, and difficult notions in Émilie du Châtelet's *Institutions de Physique*. Originally due to Leibniz, but developed in novel directions by Du Châtelet, this principle states that there are no jumps in nature. Alas, careful readers of the *Institutions* will find few clear statements of the law and a bewildering range of applications for it, ranging from arguments about the composition of physical bodies to the laws governing collisions, the nature of geometrical figures, and causation more generally.

Here I will offer a formalization of the principle, drawing from both explicit discussions and applications in the *Institutions*. I contend that Du Châtelet's notion of continuity is centrally concerned with issues of the unity of series, that is, what it means to say a series counts as a unity. A unitary series, on her picture, is one which is incapable of having new members added, thus being free of gaps or jumps. Series that are unitary in this way are continuous. The law of continuity says, in so many words, that there can be no unitary series which is discontinuous, and that nature itself constitutes such a unitary series. I will then assess how consistently this principle finds application throughout the text, and whether it meshes well with the examples she deals with. In some cases, I will argue, this principle is a sensible development of her ideas, but in some cases she appears to be seeing applications where there are none, as in her treatment of physical

bodies as composites of simple monads. These exceptions aside, however, I will conclude that Du Châtelet's treatment exhibits, at a minimum, a serious attempt to bring clarity to a perplexing and challenging issue.

#### **4. Jeremy Steeger: *The Influence of Newton on Du Châtelet's Methodology: Certainty of Phenomena and Counterfactual Dependency***

To contribute to the recent revival of interest in du Châtelet's masterwork, *Institutions de Physique* (IP), I take an unexplored tack: the continuity of her methodology with that of Newton. In Chapter 4 of IP ("Of Hypotheses"), du Châtelet details her conception of scientific method: in short, science requires the use of hypotheses to discover the true causes of phenomena. Du Châtelet's conception of phenomena originates from Wolff: she takes "phenomena" to denote images or appearances that may diverge from reality. But I argue that her approach towards the discovery of true causes given highly certain phenomena is thoroughly Newtonian: she mirrors Newton's *quam proxime* approach to certainty and his counterfactual analysis of secondary phenomena. In the body of English-language du Châtelet scholarship, only Karen Detlefsen has studied the philosopher's methodology; Detlefsen focuses almost exclusively on a comparison of du Châtelet and Descartes's respective approaches to hypotheses. While I share Detlefsen's conviction that du Châtelet intends the "certainty" with which we may know some hypotheses to be psychological, that is not the full story. As Detlefsen notes, Du Châtelet stresses both the "moral certainty" of some hypotheses and the requirement that we have "certain knowledge" of phenomena in Chapter 4. Taking into account passages from Chapter 8 ("Of the Nature of Bodies") and Chapters 13–14 (discussing gravity), I argue (1a) that the "psychological" conception of the certainty of hypotheses applies the certainty of phenomena as well and (1b) that this certainty mirrors Newton's *quam proxime* approach. Finally, following George E. Smith, I conceive of Newton's "derivation" of his gravitational law from phenomena to be an exercise in counterfactual analysis; thus, I argue (2) that du Châtelet's discussion of experiments testing the "phenomena of gravity" in Chapter 14 mirrors Newton's concern with counterfactual dependency.

#### **5. Aaron Wells: *Substance and Change in the Institutions de Physique***

I discuss Emilie du Châtelet's response to an under-appreciated 18th-century puzzle about substance and change. The puzzle is as follows. Created substances have their essences necessarily. But essences seem immutable. It's assumed that essences ground any other features those substances might have, and that something immutable can't be the ground of change. If so, however, nothing in the created world can be the ground of change. In that case, barring divine intervention, change in created substances is impossible.

Du Châtelet, Christian Wolff, and the early Kant all attempt to solve this puzzle without committing to a Leibnizian picture – on which reality has two or more levels and there's no change at the fundamental level. I argue that Du Châtelet's solution compares favorably to the inconclusive or unstable efforts of Wolff and Kant. She argues that modes have a causal order independent of the essential properties of the substances

they inhere in. In particular, modes are temporal. A given mode depends on prior modes – either of its substance or of others. Du Châtelet apparently takes inherence and grounding to come apart. A mode can inhere in a substance, such that the substance grounds the mode's existence, but not the mode's qualitative features. Such a view is unusual but not incoherent, and allows du Châtelet to address the puzzle without following Leibniz in committing to radically distinct levels of reality.

**Social Norms Across Disciplines****Symposium**

Organizer: Chiara Lisциandra

Chair: Alexander Christian

Room 5E, Thursday 15:30 – 17:30

---

*Norm Compliance and Humeanism. A Neurocomputational Account*

MATTEO COLOMBO

Tilburg University

m.colombo@uvt.nl

---

*Insult versus Accident: Caste Culture and the Efficiency of Coordination*

KARLA HOFF

World Bank

khoff@worldbank.org

---

*Modelling Norms*

CHIARA LISCIANDRA

University of Helsinki

chiara.lisciandra@helsinki.fi

---

**General Description**

The study of social norms is an interdisciplinary field of research whose contributions come from several areas of inquiry: namely philosophy, economics, sociology, political science, psychology, neuroscience and anthropology. It is in the nature of the topic to require a range of perspectives for an adequate explanation of its complexity. Consequently, however, a rather scattered picture of norm-driven behavior has emerged in the literature. A plethora of theories of social norms have been provided, each of which attempts to address some of the shortcomings of previous accounts or to fill their explanatory gaps. At first glance, it does not seem that the different contributions combine in a unified picture.

In this symposium, we will focus on the foundational aspects of social norms. We will consider some of the main approaches that have been put forward in the literature and we will ask: Which are the targets that different explanations aim at? Is there a hierarchy of explanatory levels? If so, does the hierarchy reflect an ontological, an epistemological or a pragmatic commitment? We will see that different positions are being defended by their supporters as more legitimate, appropriate and comprehensive; but a clear reconstruction

of the role that each one has is still missing in the literature. In this symposium we shall try to clarify the contribution that each of the different accounts has made to our theoretical understanding of social norms and to explore possible directions for future research.

Ideally, if the division of labour across disciplines was clearly delineated, the analysis of norms would look as follows. Economists would focus on norm-driven behavior to explain macro-phenomena and how they result in institutions. Psychologists would explore the motivational factors of norm compliance and neuroscientists the underlying neural mechanisms. Legislators would deal with the problem of how to replace negative social norms by introducing laws for the benefit of citizens. Anthropologists would study how different societies elaborate different social norms over time. According to the specific domain of inquiry, researchers would conduct theoretical analysis or experimental work, both in the laboratory and in the field. Each discipline would cover one piece of the mosaic and interactions among disciplines would take place at the boundaries. Collaborations between legislators and psychologists, for instance, would be particularly helpful when considering how to enforce new norms. Economists' models would be particularly helpful to psychologists as a benchmark against which to test individual behavior in decision-making contexts.

According to this analysis, the domains of inquiry of different disciplines have their spheres of competence within which to assess their epistemic values. However, this representation of the division of labour does not accurately reflect academic reality. Scientists disagree about how social norms should be studied. Psychologists and economists criticize each other's way of analyzing norm-driven behavior. Even the economists themselves disagree as to how economists should make sense of norm-driven behavior. Moreover, current explanatory frameworks for social norms pay little attention to why and how the brain might carry out the computational functions that generate norm compliance behavior.

These criticisms give rise to two main questions, which will be addressed in the symposium. First, how to build psychologically more realistic models that are still relevant to economic theory? Secondly, how should a measurement tool for socio-normative behavior be built to test the different and competing hypotheses? As an example of one of the main points of disagreement between psychologists and economists, consider the role of social preferences. On the one hand, psychologists insist on the necessity of enriching classical economic models with psychological parameters to represent social preferences. Social preferences should explain why we tend to cooperate in situations where otherwise we would behave according to our personal interest. On the other hand, a widespread view among economists is that social preferences are superfluous explanatory factors. Individual compliance with social norms can be explained to the extent that social norms provide rational solutions to coordination problems in repeated interactions.

To improve upon this theoretical standpoint, the role of social preferences deserves further clarification. Cristina Bicchieri is one of the most influential proponents of the view that social norms transform a choice problem characterized by conflict between individual and group outcome into a choice problem where that conflict has been solved. This happens because social norms introduce negative sanctions, either

psychological or monetary, that dis-incentivize deviation from cooperation. However, at times, social norms are presented as rules related to principles of justice, equity and equality. At other times, social norms prescribe negative behaviors such as child marriage or discriminatory policies. Depending on how we define the set of social norms, the role of social preferences changes accordingly. From something akin to a *moral emotion* which describes that— at least in certain conditions—we care about other people’s wellbeing, compliance with social norms becomes the effect of a utilitarian cost-benefit analysis for the purpose of avoiding sanctions. Furthermore, once social norms include even negative behaviors, then it ceases to be obvious that social norms solve individual vs group problems rather than introducing a conflict in the first place.

According to whether social norms are considered as the result of other-regarding preferences or as rules which individuals follow to avoid sanctions, people’s compliance in cases where deviance would not incur costs is explained differently. In the first case, compliance is taken as evidence that individuals value the content of norms; in the second case, by contrast, it is taken as evidence that the automatic tendency to follow social norms leads us to follow norms even in cases where it would be more convenient to do otherwise.

Within economics itself, one of the main points of debate concerns how more realistically to model interactive decision-making problems. To be sure, economists agree that full rationality is an unrealistic assumption. Yet how should we represent boundedly rational individuals? There are several ways in which this can be done. Memory loads and cognitive constraints change from subject to subject; even systematic deviations from full rationality are unrealistic to a certain extent. Thus, it might be asked, can we ever derive more interesting predictions under the assumption of bounded rationality than under that of full rationality? Three scholars who have worked within different frameworks will face one another to discuss and defend their approaches to the study of social norms. It is the purpose of this symposium to present three papers drawing respectively on behavioral economics, neuroscience and economics. Karla Hoel is an economist and director of the World Bank Report 2015 on social norms, mental models and behavioral economics. Chiara Lisciandra is a researcher at the University of Helsinki, who has extensively worked on models of the emergence and diffusion of norms in society. Matteo Colombo is a researcher at Tilburg University, who explores the foundations of computational cognitive neuroscience and moral psychology. The common thread that will run through the participants’ talks is a reflection on how to make the pieces of mosaic of the explanation of social norms better fit together.

## Abstracts

### 1. Matteo Colombo: *Norm Compliance and Humeanism. A Neurocomputational Account*

According to the Humean theory of motivation, belief and inference are insufficient for motivation; in addition to belief and inference, motivation requires a desire or some other conative state. However, the

mechanism of motivation and, in particular, the nature of desire remain opaque. This article explores the question of whether the Humean theory of motivation is vindicated by current work in social cognition by reviewing current research in computational neuroscience and in philosophy about motivation and social norms. I endorse Bayesian decision theory and reinforcement learning as two fruitful frameworks to uncover the mechanism of social norm compliance, and as productive guides to assess the empirical adequacy of the Humean theory of motivation. On the basis of a range of theoretical and experimental work on social norm compliance, I identify distinct ways in which the Humean theory of motivation is vindicated.

## **2. Karla Hoff: *Insult versus Accident: Caste Culture and the Efficiency of Coordination***

Much of what we value in society depends on coordination: for example, language, fiat money, standardization, and the rule of law. The particular conventions and norms that emerge vary widely across societies, with some being more efficient than others. Development economists have provided ample evidence that inefficient conventions and norms can impede economic growth, so one route to development is to improve the efficiency of coordination. But in order to do that, we need to understand why inefficient outcomes obtain.

We study the relationship between culture and the efficiency of coordination on a convention or norm. In a field experiment in India, men from high and low castes repeatedly played a simple coordination game with an efficient and an inefficient equilibrium. Compared to their low-caste counterparts, the high-caste men coordinated far less efficiently. 73 percent of low-caste pairs played the efficient Nash equilibrium in the final round of the partnership, compared to 50 percent of mixed low-high caste pairs, and only 32 percent of high-caste pairs. We trace the divergence in outcomes to how individuals respond to the low payoff that results when a player attempts efficient coordination but his partner does not. After this event, high-caste men are much less likely than low-caste men to continue trying for efficiency. This difference can be explained by the culture of honor among the high castes, which may lead them to see this low payoff as an insult rather than an accident and to respond in a manner that impedes efficient coordination.

## **3. Chiara Lisciandra: *Modelling Norms***

A novel approach to the study of social norms relies on probabilistic agent-based models. Agent-based models are a class of computational models that study the dynamics of interactive systems. By relying extensively on computer simulations, these models considerably increase the predictive power of traditional models: they make it possible to analyze phenomena involving a large number of factors and their aggregated effect, thereby overcoming the problem of tractability of non-simulated models. Through this method, it is possible to formulate hypotheses based on fewer idealizations and whose degree of proximity to the target system is higher than it would be without such simulating devices. To model, for instance, the emergence of a new norm in society, the starting point is the translation of a decisional rule into a mathematical model,

whose predictions can be observed by means of computer simulations, and tested afterwards by means of laboratory experiments, which in turn can provide feedback about the initial model. Given the features listed above, it would be natural to expect that computer simulations were welcomed in economics to study the diffusion of norms. However, when compared to analytical models and experimental work, computer simulations are often considered as a secondary option at best. A widespread tendency in economics is to legitimize computer simulations only when models become too complex to be analytically solvable; or when the volume of data collected is such that only high-powered computers can process them. In this paper, the reasons for the economists widespread skepticism will be investigated. It will be argued that there are no clear criteria to distinguish when a problem is intractable or not: hence, the claim that simulations should be limited to intractable problems looks unjustified after all.



---

*Disease-Mongering through Clinical Trials*

DAVID TEIRA

CHRISTIAN SABORIDO

MARIA GONZALEZ-MORENO

UNED

UNED

U San Pablo-CEU

dteira@fsof.uned.es

cristian.saborido@fsof.uned.es

mglezm2012@gmail.com

---

*Disease-mongering* (DM) generally refers to a purported commercial strategy of the pharmaceutical industry, consisting in tinkering with the definition of a given disease (sometimes to the point of creating a new one) in order to promote the sales of one of their drugs. Disease-mongering has been featured prominently in special issues of the *British Medical Journal* (2002) or *Plos Medicine* (2006), although its existence is for some controversial –and it has probably been so for more than four decades, since the earliest discussions about *medicalization* or the more current debates about *pharmaceuticalization*.

In this paper we want to articulate a more precise concept of DM. We want to show how pharmaceutical marketing can commercially exploit certain diseases when their best definition is given through the success of a treatment in a clinical trial. We will distinguish two types of disease-mongering according to the way it exploits the definition of the trial population for marketing purposes. We are going to argue that behind these two forms of disease-mongering there are two well-known problems in the statistical methodology of clinical trials and overcoming them is far from simple.

From a purely commercial standpoint, the industry wants any treatment to: (a) work on a given class of patients, in order to earn regulatory approval and get market access; and (b) ensure that this class is as large as possible, in order to increase sales. Pharmaceutical marketing has exploited a methodological misconception about trials that prevails among both physicians and patients. Namely, that they provide a general assessment of treatments independently of the reference class they are tested on. Hence, physicians may prescribe them off-label, assuming that a patient will benefit from them as much as the participants in the trial, even if this patient would not have been eligible.

When the definition of the trial population is so loose that physicians can be persuaded that it would suit most patients they see, we will speak of *mild disease-mongering*, since it does not target the trial as such, but medical prescription based on its outcome. However, there is also *strong disease-mongering*, where the very definition of the patient population is targeted for marketing purposes. The goal here is to find a growing group of patients where we can reach a statistically significant difference between treatments. Inasmuch as

the latter is obtained, there will be grounds to get regulatory approval for the drug and sell it to this larger audience.

Behind the two types of disease-mongering we have distinguished we will find two classical problems in statistical methodology: the reference class problem and the dichotomy between substantive and statistical significance. *Mild disease mongering* occurs when the set of eligibility criteria is vague enough as to allow the inclusion of patients that, under a more strict definition of the disease, would not qualify as members of the trial population. *Strong disease-mongering* exploits the confusion between clinical and statistical significance: if we judge how good a treatment is only on the basis of the statistical difference between the outcomes observed in a trial, we might declare a treatment effective without a real assessment of how important that difference is from a clinical standpoint. For none of these problems there is a purely statistical solution.

In order to substantiate our claims, we present our two types of disease-mongering through two case studies. The best-selling tranquilizer Valium will illustrate how sales can be boosted by over-prescription beyond the populations originally targeted in a trial. The development of the first generation of statins will show how the trial population was expanded in search of statistically significant results that allowed sales to increasing audiences. By way of conclusion, we will discuss some possible solutions to both mild and strong disease-mongering.

---

*Convergent Perspectivism*

NINA ATANASOVA

The University of Toledo

nina.atanasova@utoledo.edu

---

This paper articulates convergent perspectivism as an account of experimental neurobiology that makes sense of the multiplicity of different experimental protocols for the study of presumably identical phenomena employed in the field. The thesis is that such multiplicity is necessary under considerations for the validity of animal models which are the main experimental tool in the field. Animal models in neurobiology are used as representational models in which the experimental animals serve as proxies for humans. Their representational validity is established through a validation strategy which requires the results of multiple integrated experiments aiming at explaining identical phenomena to converge.

I develop this account in response to the challenges Sullivan (2009) raises for the validity of the knowledge produced in experimental neuroscience as well as for the philosophical accounts of the unity of the discipline. Sullivan argues that neither ruthless reduction nor mosaic unity – the main competitors for an account of the unity of neuroscience proposed by Bickle (2006) and Craver (2007) respectively – adequately handle the multiplicity of experimental protocols in neurobiology. On her account, the profound lack of uniformity of experimental designs and protocols in neuroscience precludes the reliable translation of knowledge claims produced in the idiosyncratic context of one laboratory to effects produced and studied in another laboratory.

This, according to Sullivan, stands in the way of integration of neurobiological knowledge. She thus concludes that a unified account of neuroscience is not likely to be successful.

While a disunity of neuroscience would not be considered necessarily bad for the discipline, Sullivan insists that it also puts in jeopardy the validity of the extrapolation of knowledge claims articulated on the basis of laboratory animal experiments to the targeted naturally occurring phenomena. If this is indeed true, one will have to agree with Sullivan that contemporary neurobiology – and thus neuroscience – is in a state of crisis. This is a serious challenge for the legitimacy of neuroscientific knowledge itself which has to be addressed by experimental neuroscience as well as philosophy of neuroscience.

While it is easy to agree with Sullivan that the main competitors for the account of the unity of neuroscience fail to provide it, the inference that this is so because neuroscience is in a state of crisis needs a cautious examination. Neurobiology is indeed characterized by a multiplicity of experimental designs and protocols. However, the careful study of the experimental practices of neurobiology shows that it is not due to negligence. The multiplicity of experimental designs and protocols is purposely pursued by practicing

experimental neurobiologists and they have good reasons to maintain it. Thus, my purpose in this paper is to articulate the epistemic justification for maintaining a multiplicity of experimental protocols in neurobiology. In doing so, I will show that neuroscience is far from being in a state of crisis.

Furthermore, I will show that Craver's and Bickle and colleagues' accounts of the unity of neuroscience fail not because the discipline is in a state of crisis. They fail because each of them focuses on a narrow set of experimental practices and the conclusions they draw on this basis simply do not generalize to the entirety of the discipline. While Craver studies experiments aiming at articulating entities and mechanisms underlying neuro-cognitive phenomena, Bickle and colleagues focus on experiments that aim at identifying causal relationships between the neuro-physiological entities that produce those neuro-cognitive phenomena. However, both approaches neglect the need to explore the experimental practices employed by neurobiologists in the validation of the animal models they use to articulate knowledge about the human nervous system.

This validation is a prerequisite for the success of the knowledge claims produced via the experiments studied by Craver and Bickle. My analysis shows that in the process of validation of animal models as representations of human conditions neurobiologists intentionally construct multiple experimental arrangements to study the phenomena they explore from multiple partially overlapping perspectives. The results produced this way, if they converge sufficiently, can be integrated. The technical term used by neurobiologists to designate these integrated experimental perspectives is "test batteries". Test batteries are semi-standardized sets of tests, which often overlap with respect to the functions they measure, used in neurobiology to study the targeted neuro-physiological phenomena. Vorhees (1996) argues that the employment of different tests for measuring the effects studied in neurobiological animal models provide converging data on a given functional domain targeted for study. This strategy ensures that the measured effects are not artifacts of the experimental setup.

The analysis of this practice of validation of animal models motivates my articulation of the principle of convergent perspectivism as a principle for maximizing experimental knowledge in neurobiology that justifies the multiplicity of experimental protocols employed in the discipline. According to convergent perspectivism, neurobiologists should employ multiple experimental designs and tests for the study of the targeted neuro-cognitive phenomena in order to validate the knowledge claims they articulate on the basis of experiments. When the results of the multiple experimental perspectives converge, the knowledge claims thus produced are considered validated. For this reason, maintaining the multiplicity of experimental designs and protocols in neurobiology is not a sign of crisis. It is rather a sign of thriving of the discipline.

Convergent perspectivism addresses Sullivan's challenge for the validity of neurobiological knowledge. It also puts a common denominator under the two traditionally competing account of the practice of neuroscience, Craver's and Bickle's. They both identify the need for employing multiple experimental approaches to the explanatory targets in neuroscience. For example, Craver (2007) argues for the desirability

of the employment of multiple techniques for detecting the components of mechanisms the description of which is the ultimate goal of experimentation in neuroscience on his account. Craver's purported rival – Silva and Bickle (2009) and more recently Silva, Landreth and Bickle (2014) – is explicit about the importance of the convergence of results of multiple experimental approaches for the articulation of causal explanations in neuroscience. Convergent perspectivism, thus, provides a promising platform for articulating a philosophical account of the integration of knowledge in – if not unity of – neuroscience.

### References

- Bickle, J. (2006). "Reducing mind to molecular pathways: Explicating the reductionism implicit in current cellular and molecular neuroscience". *Synthese*, 151(3): 411-434.
- Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press; NY: Oxford University Press.
- Silva, A. J. and J. Bickle (2009). "The science of research and the search for molecular mechanisms of cognitive functions". In J. Bickle (Ed.), *The Oxford Handbook of Philosophy and Neuroscience* (pp. 91-126). Oxford: Oxford University Press.
- Silva, A., A. Landreth and J. Bickle (2014). *Engineering the Next Revolution in Neuroscience*, NY: Oxford University Press.
- Sullivan, J. A. (2009). "The Multiplicity of Experimental Protocols: A Challenge to Reductionist and Non-reductionist Models of the Unity of Neuroscience." *Synthese* 167: 511-539.
- Vorhees, C.V. (1996). "Design Considerations in the Use of Behavioral Test Batteries for the Detection of CNS Dysfunction in Laboratory Animals". *Mental Retardation and Developmental Disabilities Research Reviews*, 2: 227-233.

---

*Extended Inheritance as Persisting Extended Organization*

GAËLLE PONTAROTTI

IHPST, Université Paris 1 Panthéon Sorbonne

gaelle.pontarotti@gmail.com

---

During the major part of the 20th century, investigations about biological inheritance were dominated by a gene-centered vision, according to which duplication and transmission of parental genes is the sole mechanism responsible for transgenerational similarities (Mameli, 2004). However, this vision has been seriously challenged for the last two decades. Whereas many studies have been dedicated to epigenetic, ecological or symbiotic channels ensuring the transmission of traits across generations, inclusive conceptual frameworks have been proposed to take into account non-genetic systems of inheritance and the related diversification of biological legacies.

Some leading thinkers have adopted an informational perspective according to which these legacies include genetic and non-genetic information contributing to phenotypic resemblance across generations (Jablonka, 2002; Jablonka & Lamb, 2005; Shea, 2007; Danchin *et al.*, 2011; Mesoudi *et al.*, 2013). Other have defined inheritance in terms of stability or availability of genetic and non-genetic developmental resources (Griffiths & Gray, 1994; Oyama, 2000; Stotz, 2010). As required by studies on ecological inheritance, most of these important contributions include elements traditionally belonging to environment in the category of inherited determinants. However, such an inclusion overrules a key theoretical requirement in the studies on inheritance, which consists in keeping a clear line of demarcation between biological systems – and consequently biological lineages – and their environment, even if this environment is involved in the reoccurrence of traits and can be stabilized by previous generations. While current frameworks resort to an evolutionary argument to capture the limits of the extended phenomenon under investigation (see Griffiths & Gray, 1997; Griffiths, 2001; Stotz, 2010 for DST accounts; see Jablonka, 2002; Shea, 2007 for informational accounts), I propose an alternative theoretical argument.

The main purpose of this paper is to outline the contours of an organizational perspective on extended inheritance. Based on theoretical studies on biological organization (Locke, 2001; Mossio *et al.*, 2009; Mossio & Moreno, 2010) and extended physiology (Turner, 2002, 2004), this perspective allows thinking about diversified and spatially extended biological legacies – including genetic and non-genetic determinants – while maintaining a theoretically indispensable line of demarcation between biological systems and their conditions of existence. In this context, this line of demarcation is outlined by an organizational criterion, and by the related theoretical distinction between *organizational constraints* and *environmental resources*. Biological legacies are restricted to persisting *organizational constraints* – like genes and gene's products,

niche's artefacts, symbionts, etc. – which collectively determine the maintenance of *an extended organization* within and across generations of *composite biological systems*. Persisting constraints, whose causal role is to harness flows of matter and energy so as to allow what contemporaneous biologists would call a “thermodynamically open system” to keep producing its components, belong to successive biological systems. They must be distinguished from the stable *resources* they contribute to exploit, even if these resources are stabilized or modified by biological lineages. As a result, extended inheritance is neither defined as transmission of information across generations nor as stability of a set of developmental resources; it rather appears as *persisting extended organization*.

The well-studied case of symbiotic transmission is presented as a paradigmatic system for an organizational perspective on extended inheritance. Whereas persisting symbionts are known to complete metabolic pathways and to ensure other physiological functions in the composite system they form with their host (Gilbert, Sapp & Tauber, 2012), the specific example of persisting shared metabolism between hosts and symbionts provides an interesting example of persisting extended organization. It also relevantly illustrates the theoretical difference between persisting *constraints* and stable *resources*. In this case, persisting microbial partners ensure biosynthesis and nutriment degradation for their host (Turner, 2004; Margulis & Sagan, 2001; Douglas, 2009: 43; Hansen & Moran, 2011). In doing so, they co-determine nutrients channeling across generations of composite biological systems and appear as *organizational constraints* which are not to be mistaken with the stable *resources* they contribute to harness. From an organizational point of view, these persisting microbial partners are genuine parts of composite biological systems and of diversified biological legacies, be they transmitted vertically or recruited in the environment. In the same way, other ‘environmental’ elements having the theoretical status of constraint will be included into composite biological systems and into diversified biological legacies.

To sum up, the aim of this paper is to draw the contours of an organizational account of extended inheritance, which appears as an alternative to other conceptual frameworks. In this perspective, extended inheritance is defined as persisting extended organization. Biological legacies are restricted to persisting *organizational constraints* whose specific causal role is to harness flows of matter and energy across generations of composite biological systems. The new line of demarcation between successive biological systems and their environment, which is a theoretical requirement in studies about inheritance, is outlined by an organizational criterion and by the related theoretical distinction between *constraints* and *resources*. In future works, I will study the characteristics of the diverse constraints involved in extended inheritance. I will also address the related and difficult issue of transgenerational – i.e. temporal – boundaries between parents and offspring, in a context where biological systems are made of heterogeneous parts whose cycles of reconstruction are not necessarily synchronized (see Sterelny *et al.*, 1996 for a similar idea). These elements will be of the highest importance to further assess the impact of extended inheritance on evolutionary thinking. Here again, the case of symbiotic transmission will be taken as a paradigmatic system.

---

*Natural Selection: Deriving Causality from Equilibrium*

HUGH DESMOND

KULeuven

hugh.desmond@hiw.kuleuven.be

---

Natural selection is often represented as one of the main causal mechanisms driving evolution, and is often modeled in biology textbooks as some kind of Newtonian force, with magnitude and direction. However, this picture is complicated when one takes into consideration how evolution by natural selection is constituted by individual births and deaths. Following a number of articles by Walsh, Ariew and Matthen, there is now a significant challenge that natural selection may not even be a cause, let alone a Newtonian force. Individual births and deaths can be described independently of natural selection. Natural selection is no causal propensity over and above individual-level processes; rather, it is a statistical effect, a mere book-keeping of the genuinely causal interactions that take place between individual organisms (Matthen and Ariew 2009; Walsh, Lewens and Ariew 2002; Walsh 2007).

In the extensive literature that has ensued, the statisticalist approach has mainly been used to argue for a *deflationary* position “fitness and natural selection have no reality except as accumulations of more fundamental events” (Matthen and Ariew 2002, 82). In this paper I will investigate the underexplored possibility of a non-deflationary statisticalist analysis of selection. This adopts the statisticalist, bottom-up analysis of population change, but tries to reconcile it with certain causalist intuitions. The inspiration for this is that, while statisticalist considerations may preclude certain naïve ways of understanding the causal nature of selection, causalist intuitions cannot be entirely wrong either. At the very least, it cannot be denied that most of biological practise is not threatened by these considerations. While it may be metaphysically inaccurate, it is often empirically accurate to model selection as a causal force (for example in cases of stabilizing selection, where component pressures cancel out). This suggests that causalist intuitions must be legitimate in some way.

My approach in this paper will be to use the notion of *equilibrium* as a way of understanding how the causal nature of selection can be real, thus grounding causalist intuitions. Equilibrium is a central concept in modeling the behavior of complex systems. In particular, stable equilibria are empirically important because they act as attractors and allow for a long-term prediction of the behavior of the system, even though the behavior in the middle-term may be chaotic and too complex to calculate. However, they are also philosophically important as they can allow a well- defined *direction* to be assigned to a complex process. Thus a concept of directionality can be formulated that is grounded in a statistics of individual-level dynamics and that allows us to understand why natural selection can be legitimately called causal.



To establish such a framework, I will need to do three things. The first task will be to lay the ground by disentangling some different notions of causality at play, in particular process and difference-making causality. Each highlights a different aspect of natural selection and confusion results if these are not kept separate. In this paper I will focus on difference-making causality alone, mainly because this notion has been more controversial. Difference-making is, broadly, counterfactual dependence. The statisticalist arguments have endeavored to show that, even if natural selection were not present, evolutionary change would occur.

One argument has been that natural selection is established only retroactively, by a statistical regression on actually occurred births and deaths (where selection is the correlation between traits and births). There is no description-independent way of establishing fitness or natural selection (and this is related to the reference class problem). Another argument has concerned the inseparability of natural selection from the causal processes affecting the behavior of organisms. The probabilities that characterize the possible outcomes by natural selection are only a measure of our ignorance of the individual-level processes determining the births and deaths. They do not correspond to any putative ‘causal propensity’ that could be used to ground natural selection.

The second task will be to formulate the condition of equilibrium, and to show how, if it is accepted, it can resolve certain key issues regarding difference-making causality. For this I will use an extension of the Price equation to the multigenerational case. The Price equation gives an exact relationship between the phenotype distribution of different generations:

$$\bar{z}^{(k+1)} - \bar{z}^{(k)} = \text{Cov}(\omega(k), z^{(k+1)}) + E[z^{(k+1)} - z^{(k)}],$$

where  $z^{(k)}$  is the phenotype variable of the  $k^{\text{th}}$  generation,  $\text{Cov}(\omega(k), z^{(k+1)})$  a measure for how relative fitness  $\omega$  covaries with phenotype  $z$ , and  $E[z^{(k+1)} - z^{(k)}]$  the expected transmission bias.

I will show how this equation can be simplified considerably under assumption that an equilibrium is reached after a certain number of generations. This assumption then allows one to uniquely define a direction of an evolutionary process: the tendency towards equilibrium.

This is important because it allows one to argue that the probabilities defining fitness are not purely description-dependent. Neither is natural selection merely a measure of subjective uncertainty; rather, it reveals an objective feature of certain evolutionary processes, namely the presence of stable equilibrium. Natural selection is causal in the difference-making sense: if it were not present, an evolution towards stable equilibrium would not be observed.

Finally I will need to argue why the equilibrium condition is a plausible assumption. To this end, I will show that given evolutionary change, either a stable equilibrium is reached, or if it is not, then the concept of fitness is not meaningful. I discuss certain results from Markov process literature, where the conditions for equilibrium are established (Doebelin’s theorem). From this it can be seen that the notion of equilibrium is intertwined with natural selection, and that this is a natural way to reconcile both statisticalist and causalist approaches.

---

*In Defense of Historical Theories of Confirmation*

CORNELIS MENKE

University of Bielefeld

cmenke@uni-bielefeld.de

---

In a seminal paper of 1974, Alan Musgrave contrasted *Logical* with *Historical Theories of Confirmation*. According to Musgrave, a long tradition in methodological thought, from Descartes and Leibniz to Whewell and Duhem, favored historical approaches to confirmation (or HTC for short) and held that a good hypothesis should not only account for the known phenomena (or the phenomena it was devised to explain), but should predict novel phenomena besides. According to modern ‘orthodoxy’, on the other hand, all that matters in assessing the support of a hypothesis is the hypothesis itself, the empirical evidence, and the logical relations between both. This latter approach Musgrave called a (purely) logical theory of confirmation (or LTC for short). In particular the question of the evidential weight of successful predictions as opposed to accommodations is still lively debated (Douglas & Magnus 2013) – and still open.

I shall argue that the present debate on predictivism vs. accommodationism is not identical (although obviously connected) to the pivotal question Musgrave raised, viz., whether to prefer *logical* or *historical* theories of confirmation, and on what grounds. In fact, despite the growing number of suggested explanations of the (supposed) evidential weight of predictions, the debate on predictivism has ceased to address this more fundamental issue – and not, or so I shall argue, for fully convincing reasons.

First, I shall try exemplarily to show that basically all major participants in the present debate – advocates and critics of predictivism alike – accept and rely on *logical* approaches to confirmation (1). Secondly, I shall consider the main reasons for this preference for LTC, and shall propose that partly the arguments beg the question against adherents to HTC, partly the cogency of the arguments is disputable or unsettled, and I shall indicate possible responses to these objections (2). Finally, I shall discuss the prospects of HTC and conclude that, from a pragmatist point of view, HTC are not nearly as counter-intuitive as they are widely believed to be (3).

**1.**

‘Predictivism’ holds that successful predictions of novel phenomena carry special epistemic weight. At first sight, predictivism could be conceived of as a special case of HTC; but it is plain that there is more to HTC than predictivism: for instance, the methodological rule of predesignation of statistical tests (Peirce 1883) as

well as generativist methodologies (e.g., Nickles 1989) are different versions of HTC; thus, HTC is the wider concept. But on the other hand, not all champions of predictivism would necessarily adhere to HTC, at least not in Musgrave's sense – I shall argue that, on the contrary, most participants in the predictivism debate rely on LTC. That this holds for those who deny a special value of predictions is plain. But also amongst predictivists there is a widespread *consensus* that (a) predictivism is in need of justification and (b) that the means of this justification lies in reducing the alleged value of predictions to logico-structural relations between theory and evidence. In other words, predictions count because successful predictions are connected (a sign or indicator) for other, genuine epistemic values, and insofar only of 'instrumental' (Douglas/Magnus) or 'symptomatic' (Schindler) value. There is, to be sure, no consensus as to what epistemic values in particular predictions are an indicator of: severe tests (Mayo), explanatory power (Harker), avoidance of 'over-fitting' (Hitchcock/Sober), or all of them, depending on the context (Douglas/Magnus). – John Worrall's version of predictivism is a special case. Worrall holds that only use-novel evidence can genuinely support a theory (in this respect, his claim is much stronger than modest predictivism); but at the same time, he considers the actual historical development unimportant. What matters is not, whether some evidence has been used to develop a theory (or fix a parameter), but whether it *could have been* used: thus Worrall, too, favors LTC.

## 2.

Since the debate on predictivism is a debate nearly exclusively amongst adherents of LTC, it comes as no surprise that the arguments against HTC are rarely fully articulated. One might distinguish *logical* and *empirical* considerations. Logical objections to HTC mostly refer to (a) the lack of a rationale for HTC, (b) the claim that an evidential weight of 'contingent' historical facts is counter-intuitive, and (c) the claim that "from a logical point of view, the strength of the support that a hypothesis receives from a given body of data should depend only on what hypothesis asserts and what the data are." (Hempel 1965) I shall argue that these logical objections are not cogent: (a) leads to a circular justification (either it reduces the value of predictions to other historical facts, in which case it will beg the question against proponents of LTC; or it reduces it to logico-structural values (as above), thereby explaining the value of historical facts *away*. (b) simply begs the question against proponents of HTC (besides, the history of methodology shows that the intuition that historical facts are clearly irrelevant to confirmation is of quite recent origin). (c), finally, is plainly false (Duhem).

I shall argue that the 'empirical' objections to HTC are more cogent, and discuss the most prominent ones and sketch possible rejoinders: (d) scientists don't usually refer to historical facts when assessing the merits of hypotheses; (e) scientists often feel comfortable to assess theories even without knowledge of the historical circumstances of their development; (f) HTC are not compatible with the *objectivity* of science; (g) HTC are not the only game in town and inferior to LTC.

**3.**

Finally, I shall briefly consider the prospects of HTC in general. In particular, I shall propose that HTC are wrongly considered counter-intuitive: 'contingent' historical facts about the development of a theory are correctly regarded as irrelevant (if only due to the meaning of 'contingent'); but historical facts are only contingent if one assumes that there are no rules or patterns of development in science whatsoever (a claim which contradicts much of the practice of philosophy of science, and most theories of scientific development except for, perhaps, Darwinian approaches). That there is no algorithm for discovery and theory-formation is granted; but from a pragmatist point of view one might very well argue that we (and researchers in particular) know more about the process of research and its patterns than about the outcomes. This would explain the intuitive appeal that HTC – rightly – had and still has.

---

*From Zymes to Germs: Discarding the Realist/Antirealist Framework*

DANA TULODZIECKI

Purdue University

tulodziecki@purdue.edu

---

According to one of the main anti-realist arguments, the pessimistic meta-induction, we have reason to believe that our current theories are just as false as their predecessors. Proponents of this argument draw attention to a list of theories that were once regarded as highly successful, yet ended up being discarded and replaced by radically different ones. Scientific realists, in response, have argued, first that the anti-realists' list is too permissive, and ought to be restricted only to theories that enjoyed 'genuine' success, which, according to realists, consists in a theory's ability to make (use-) novel predictions, i.e. predictions that played no role in the generation of the original theory. Second, in dealing with the remainder of the so diminished list, realists have proposed and endorsed a variety of selective realisms (notably those of Kitcher, Worrall, and Psillos) which emphasise the carrying over of stable and continuous elements from earlier to later theories and which are then used to argue for the approximate truth of those earlier theories.

In this paper, I argue that neither realist nor anti-realist accounts of theory-change can account for the transition from zymotic views of disease to germ views. I begin by explaining the zymotic theory of disease, one of the most sophisticated and popular versions of the mid-nineteenth miasma theory. The zymotic theory drew on some of the most successful science at the time, such as Liebig's chemical theories, thereby allowing it to propose highly detailed mechanisms about the exact manner of disease causation. According to the zymotic theory, diseases occur as a result of introducing into the body various zymotic materials, either through direct inoculation or through inhalation after being dispensed in the air. Essentially, these zymotic materials were thought to be putrefying organic matter that would communicate its process of decomposition to pre-existing materials in the victim's blood where it would act in a manner similar to ferment, thus causing diseases.

After explaining the basics of the zymotic theory, I then show (i) that the zymotic theory and its successor, the germ theory, are strikingly different in almost every respect and (ii) that, despite the fact that the zymotic theory was so different from its successor, it was highly successful. Moreover, I show (iii) that this is so even according to the realists' own, more stringent, criterion of success as consisting of use-novel predictions. Some examples of such use-novel predictions were the zymotic theory's predictions about what geographical regions ought to be affected by diseases to what degrees, and, strikingly, a number of numerically very precise predictions resulting from Farr's so-called elevation law of 1852, relating cholera mortality and the elevation of the soil. Other novel predictions concerned the course and duration of epidemics, the relation

between population density and disease morbidity and mortality, the relation between mortality rates and different occupations, and relations between mortality from various diseases and age. I argue, however, that despite the zymotic theory's successes, realists cannot account for the zymotic case. According to selective realists, precisely those parts that were indispensable to a theory's genuine success are the ones that ought to be retained; yet, as I show, there is no discernible continuity between the zymotic theory and the germ theory: the zymotic theory had an entirely different ontology and structure from that of the germ theory, and it was also radically conceptually different in other ways, such as in its focus on processes of decay as opposed to pathogenic entities. Thus, there were no stable or invariant elements that were carried over from the zymotic to the germ theory: neither its entities, nor its mechanisms or laws, nor its processes, or even the structure of diseases themselves was retained.

It thus appears that the zymotic theory is exactly the kind of case that anti-realists are looking for as support for the pessimistic meta-induction: it was highly successful, discarded, and had very little in common with its successor. However, I argue that, in fact, anti-realists fare no better than realists, since there was also no radical conceptual change or discontinuity between zyme and germ views: despite the fact that the zymotic theory and the germ theory – viewed as different products – are radically different, the transition from the former to the latter was neither radical nor sudden.

To make this point, I show that there were no clearly defined and opposing germ and anti-germ research programmes, as is often claimed; in particular, there was no switch from one of these views to the other, but, instead, a gradual transition during which different aspects of a number of germ views were slowly assimilated into the zymotic theory. Elements of zymotic and germ views co-existed for some time, until, eventually, various parts of the zymotic theory were discarded, little by little, as increasingly well-defined versions of the germ theory emerged and started taking hold. The specific examples I use to argue for this position are (i) the changing views about the media of disease transmission, (ii) the changing views about the nature of zymes, and (iii) the change from chemical views of disease to biological ones. I conclude that neither realist nor anti-realist views can adequately account for the transition from zymes to germs. However, I argue that the problem lies not with specific realist or anti-realist proposals, but, rather, with an unwarranted assumption they both share, namely the assumption that there are well-delineated theories that can be compared and assessed on terms set by the realism-debate in the first place, an assumption that does not hold in this case.

---

*The Gap Between Psychological Explanation and Mechanistic Explanation*

SHELDON CHOW

Leibniz Universität Hannover

sheldonjchow@gmail.com

---

So-called New Mechanists (e.g., Bechtel, 2008b; Craver, 2007) argue that the norms and methodology of philosophical accounts of scientific explanation should be modeled on the mechanistic explanatory practices of the life sciences (biology being the paradigmatic example). Bechtel and Abrahamsen (2010) have offered perhaps the most constructive characterization of mechanism:

A mechanism is a structure performing a function in virtue of its component parts, component operations, and their organization. The orchestrated functioning of the mechanism, manifested in patterns of change over time in properties of its parts and operations, is responsible for one or more phenomena. (p. 323)

According to the New Mechanists, psychological explanation should likewise proceed by describing the parts, operations, and the organized functioning of the relevant neural and biochemical structures that are responsible for psychological phenomena (e.g., Bechtel, 2008b).

However, a simple adoption of methods and norms found in biology isn't a straightforward matter for explaining psychological phenomena. If we consider biological phenomena, we see that explanations involve *material transformations* (e.g., moving ions, oxidizing a substrate, activities of proteins). However, if we look at the sorts of phenomena that are of interest to psychologists, we see that they don't fit this model of explanation quite so well. Indeed, many phenomena that interest cognitive psychologists concern transformations of *information*, where structures and processes of underlying substrate is less important than what the information *is about* (Bechtel, 2008a, 2009). As such, the methodology of psychology typically differs from that of biology. Instead of attempting to identify physical loci and material transformations, psychologists typically attempt to identify the functional systems that are responsible for taking information as input and producing outputs, and then hypothesize about how the information is transformed by these systems (Cummins, 1983, 2000).

Even though they don't deliver the right sort of explanation for mechanistic philosophers, functional analysis and information-processing models of psychology are understood to be useful insofar as they facilitate transitioning to mechanistically explaining psychological phenomena as science advances (Bechtel 2008a; Piccinini & Craver, 2013). However, I argue that the disconnect between mechanistic explanation and psychological explanation is more severe than what these mechanist philosophers make it out to be, and that this gives us pause to consider the limits of mechanistic explanation. I pose two problems in particular that show how psychological phenomena preclude mechanistic explanation.

I first turn to a problem concerning the relevant entities in psychological explanation. A central feature of mechanistic explanation is that the mechanism, whatever it happens to be, must be identified in terms of a system's parts, operations, and the organized interaction of these. Yet, psychological entities don't seem to be of an ontological type to fit this description. Therefore, it is questionable whether psychological phenomena are amenable to mechanistic characterization. To illustrate, I discuss a specific sort of psychological phenomenon that is particularly germane to psychological explanation: inference. Inference is commonly conceived as a mental act of adding new beliefs and/or giving up old beliefs in light of other held beliefs (Harman, 1986; Boghossian, 2014; Rips, 1998; Evans, 2013). This characterization of inference puts *contents* as the driving force—what's important is what the beliefs are about, and the role this plays in belief change, and not how the relevant beliefs and inference are realized.

Secondly, inference requires that one's representations must be appropriately transformed—one must adopt/reject a belief because of a semantic relation that holds between it and other held beliefs, and a simple causal-mechanical relation won't do. But this admits a normative vocabulary that can't be captured in mechanistic terms. When we ask why *S* made the inference she did, we are asking for reasons why *S* came to adopt the belief(s) she did, not for the mechanisms that made it possible or gave rise to it. Put another way, mechanistically explaining inference is possible *only if we reconstrue what it is we are trying to explain*—are we attempting to explain why *S* inferred what she did, or how *S*'s neural systems give rise to beliefs and how these are altered in a manner that we recognize as inference? Going from the former to the latter would entail a shift in the explanatory question from *why* to *how*, which is consistent with the mechanists' programme (Bechtel, 2008b). But this shift means that mechanistic explanation isn't addressing the same questions that philosophers and psychologists typically address.

These arguments are reminiscent of the classical debates over psychological reduction, but the purpose of the present paper is not to explicitly revisit these debates. Rather, it is to show that mechanistic explanation is incomplete for psychological phenomena. In exposing the limits of mechanistic explanation, I go on to develop a view according to which scientifically explaining psychological phenomena requires different sorts of explanations, mechanistic explanation being just one of several. To explain inference, for example, we may need in addition to mechanistic explanation, rational explanation (explaining the reasons for belief change), semantic explanation (explaining what kind of belief change is inference), functional explanation (explaining how belief states function to enable inference), and representational explanation (explaining how manipulations of representations can effect changes in belief content). This position is developed in reference to Wimsatt's (e.g., 2007) notion of robustness, where a wide range of means of detecting a phenomenon provides corroboration of its existence, and increasing the means of detection increases our understanding of the phenomenon. In short, to achieve robust explanations of psychological phenomena, we will need to go beyond mechanistic explanation to reveal how different varieties of



explanations explain different aspects of the relevant psychological phenomena, and how these explanations relate to each other in order to increase our understanding of the phenomena.

I conclude by drawing lessons for scientific explanation generally. My analysis shows that scientific explanation itself is a complex phenomenon. We might apply the New Mechanists' motivation to pursue mechanistic explanation to explanation itself, whereby scientific explanation is (ought to be?) realized by the orchestrated activity of many different parts, i.e., different explanations that target different facets of the explanandum phenomenon.

## References

- Bechtel, W. (2008a). Mechanisms in cognitive psychology: What are the operations? *Philosophy of Science*, 75, 983-994.
- Bechtel, W. (2008b). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. London: Routledge.
- Bechtel, W. (2009). Constructing a philosophy of science of cognitive science. *Topics in Cognitive Sciences*, 1, 548-569.
- Bechtel, W., & Abrahamsen, A. (2010). Dynamic mechanistic explanation: Computational modeling of circadian rhythms as an exemplar for cognitive science. *Studies in History and Philosophy of Science*, 41, 321-333.
- Boghossian, P. (2014). What is inference? *Philosophical Studies*, 169, 1-18.
- Craver, C. F. (2007). *Explaining the brain: The mosaic unity of the cognitive neurosciences*. Oxford: Clarendon.
- Craver, C. F. (2014). The ontic account of scientific explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the special sciences: The case of biology and history* (p. 27-52). Berlin: Springer.
- Cummins, R. (1983). *The nature of psychological explanation*. Cambridge, MA: The MIT Press.
- Cummins, R. (2000). 'How does it work?' versus 'what are the laws?': Two conceptions of psychological explanation. In F. Keil & R. Wilson (Eds.), *Explanation and cognition* (p. 117-144). Cambridge, MA: The MIT Press.
- Evans, J. S. B. T. (2013). Reasoning. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology*. Oxford: Oxford University Press.
- Harman, G. (1986). *Change in view: Principles of reasoning*. Cambridge, MA: The MIT Press.
- Piccinini, G., & Craver, C. (2013). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, 183, 283-311.
- Rips, L. J. (1998). Reasoning. In W. Bechtel & G. Graham (Eds.), *A companion to cognitive science* (p. 299-305). Malden, MA: Blackwell Publishing.
- Wimsatt, W. C. (2007). *Re-engineering philosophy for limited beings: Piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

---

*Measurements, Coordination, and the Problem of Representation of Physical Quantities*

FLAVIA PADOVANI

Drexel University

fp72@drexel.edu

---

The problem of measurement is a central issue in the epistemology and methodology of the physical sciences. While our epistemic access to the phenomena is supplied by theories and models, this access is primarily achieved through experimentation, and hence measurements play a fundamental role in the construction of our scientific knowledge.

Physical theories typically consist of principles, laws, and equations that are expressed in mathematical form and that include parameters representing relevant physical quantities such as velocity, pressure, and temperature, as theoretical terms. Although such quantities are already sufficiently interpreted within each theory, they become meaningful only when they are related to some measurement procedure. But the parameters that appear in scientific theories and equations are not pre-existing quantities. As the history of science illustrates, simple quantity terms of our scientific knowledge that we take for granted, and the instruments that we use to measure them, actually arise as outstanding achievements of our scientific conceptualization and technical progress. Their individuation *as* parameters for the relevant laws and equations often goes together with the creation of the corresponding measurement procedures. Furthermore, measurement procedures are grounded on, and heavily depend on, a pre-constituted conceptual framework, which in turn they themselves contribute to forge: in other words, measurements are theory-dependent. At the same time, though, scientific theories are empirically tested by means of measurements. That raises a threat of circularity, in that the reliability of the measurement procedures used in scientific theories is often established by appeal to the same theories that are supposed to be confirmed by such measurements. In this context, it thus becomes crucial to determine the conditions for the objectivity we ascribe to our scientific theories.

This task is closely related to the problem of scientific representation. Thus, a condition for the objectivity of scientific knowledge rests on the ability to coherently represent the behaviour of measured objects as a good approximation of a theoretical ideal, which appears as some form of “natural prior” with respect to actual measurements. Measurement outcomes can be inferred from instrument indications only against the background of an idealised model, which strictly depends on the scientific theory in use. What one obtains is thus a construct, rather than a “brute fact”. In order to enhance the understanding of scientific knowledge, one then ought to investigate the adequacy of the representation of such a construct in the process of measurement and within the conceptual model that accommodates it. In his *Scientific Representation* (2008),

van Fraassen has indeed emphasised how measuring should be considered as a form of representation. In fact, every measurement pinpoints its target in accordance with specific operational rules within an already-constructed theoretical space, in which conceptual interconnections can be represented. So, this “logical space” provides the range of possible features pertaining to the measured items described in the domain and in the language of the relevant theory. Without this space of pre-ordered possibilities there can be no objects of representation *for us*. In this sense, the act of measuring is “constitutive” of the measured quantities as it allows for the coordination of mathematical quantities to “pieces of reality,” thereby providing meaning to the abstract representations through which we seek to capture physical phenomena.

In recent years, there has been a revived interest in the notion of “coordination” especially in relation to the issue of scientific representation as van Fraassen has described it in his (2008, ch. 5). In this connection, Hans Reichenbach’s 1920 account of coordination has revealed to be particularly interesting. In his early work, however, the idea of “coordination” was employed not only to indicate a class of general, theory-specific fundamental principles such as the ones suggested by Michael Friedman’s account of a relativized a priori, but also to refer to a number of other “more basic” principles. In Reichenbach’s early work, these “basic” principles are related not much to the structural features of a theory, but rather to the conceptual presuppositions required in order to approach the world through measurement in the first instance. Those basic principles are primarily necessary to translate the unshaped material from perception into some quantities that can be used within the mathematical language of physics. Quite interestingly, in his early writings many of these coordinating principles are conceived as preconditions both of the individuation of physical magnitudes *and* of their measurement. In other words, they are not limited to the definition of quantity terms but they also involve the individuation of what these quantity terms are supposed to be coordinated to.

The aim of this paper is to reassess Reichenbach’s approach to coordination and to the representation of physical quantities in light of recent literature on measurement and scientific representation, in contrast to the work of van Fraassen and of Friedman.

---

*Interventions at the Core of Scientific Reasoning – On De-Idealizing and  
Re-Idealizing Formal Logic*

MARTIN MOSE BENTZEN

Technical University of Denmark

mmbe@dtu.dk

---

Classical logic can only deal correctly with reasoning about closed systems, whereas the world described by science is an open and dynamic system. Therefore most patterns of scientific reasoning, such as inferences about causation, abduction, model fit, etc. are *ceteris paribus* and in principle defeasible. However, this does not imply that logic and theory of science must be separated forever. There is a promising and refreshing trend in current logic which takes the actual reasoning of human beings seriously. This movement has been dubbed *a cognitive turn in logic*, see (Bentzen, 2008), or a *naturalization of logic*, see (Woods, 2013). Common to this trend is a renewed interest in investigating the tension between the descriptive and normative aspects of logic, and an interest in strategies of coping with anomalies of traditional (classical and non-classical) logics which go beyond mere gesturing towards pragmatics. In this paper, I approach this call for a *de-idealization* of logic from a philosophy of science angle. A de-idealization of logic will require that logicians must come down from the ivory tower of their formal modelling and take a new look at actual reasoning in its natural (or perhaps social) context. However, it has been suggested at least since Kuhn that science does not deal directly with objects in themselves but with versions of objects which are at least partially shaped by the very scientific approach to them. As a consequence, naturalization is not an unproblematic term. Since we do not have direct access to our own naked reasoning, it is not just a matter of adopting empirical methods in order to establish adequacy between formal logic and natural reasoning. Rather, the process of establishing any object as an object of scientific inquiry, whether in natural science or in logic, is itself one that involves a preparation process which consists in a delimitation, abstraction and idealization of said object. As a seemingly simple example, take the preparation process leading from everyday language to logical inferences in semantical accounts of classical propositional logic. In the first step, we delimit the set of syntactically possible sentences to declarative sentences, leaving out e.g. imperatives and interrogatives. Then we abstract from most features of declarative sentences to arrive at the idea of a propositional content. Then we abstract further to the idea of a truth value. Then we idealize by concentrating only on sentences which can be regarded as instances of generic objects capable of having a truth value. In a further step of the process we delimit the theoretical context such generic objects can

interact within via the classical connectives and rules of inference. At all steps of this process, choices are made which themselves require theory to be justifiable. For example, a syntactical categorization dividing sentences into declaratives, interrogatives, imperatives, etc., must be theoretically grounded in linguistics. In general, what do we do to prepare and delimit the objects of reasoning? What aspects do we isolate and how do we remove irrelevant and disturbing factors? How is reasoning shielded from disturbing contextual factors? Such questions are the focus of the first part of this paper.

A further problem with requiring a naturalization of logic is that if we do not want to give up the normative aspect of logic we can never define reasoning in completely naturalistic terms. Giving up this aspect and reducing logic to empirical psychology or cognitive science, seems either to imply giving up any concept of reasoning being *correct* which could have devastating consequences for mathematics and philosophy, or to shift the role of a normative justification of reasoning to these sciences which is equally problematic. Even if we grant that natural language or cognitive processes can provide empirical triggers for a de-idealization of logic, such a process must be seen as a possibly revolutionary intervention with core elements of our current understanding of reasoning. In order not to revert to a state of unclear informal reasoning, following de-idealization there is a need for a subsequent step of re-idealization. However, this step is, perhaps not surprisingly, proving itself rather difficult. In philosophy of science, we do have important studies of *ceteris paribus* reasoning, see e.g. (van Benthem, Girard, and Roy, 2009), abduction, see e.g. (Aliseda, 2006), causation, see e.g. (Pearl, 2009) and default reasoning, see e.g. (Horty, 2012), also (Woods, 2013). However, although these studies seem to indicate that what is needed is a proper account of defeasible non-monotonic reasoning and that the boundary between pragmatics and semantics must be reinterpreted, there is as of yet no single satisfactory account of such reasoning in any of these areas, despite obvious merits of the various accounts mentioned above. Why is this? I first look at several kinds of scientific interventions of increasing depth (normal scientific/hypothetico-deductive interventions, definitional interventions, and interventions regarding interpretation of simple terms). In the second part of the paper, I argue that a de-idealization and subsequent re-idealization of logic will require an intervention of a deep kind. More specifically, I argue for a de-idealization of logic by reinterpreting its simple terms via awareness of the fact that they are objects constituted by the preparation process outlined in the first part of the paper and a subsequent re-idealization of logic which is aimed at establishing patterns of correct reasoning within the general context suggested by recent studies of non-monotonic reasoning.

## References

- A. Aliseda, 2006, *Abductive Reasoning. Logical Investigations into Discovery and Explanation*, Springer.  
J.F. Horty, 2012, *Reasons as defaults*, Oxford University Press.  
J. Pearl, 2009, *Causality*, 2<sup>nd</sup> edition, Cambridge University Press.  
J. van Benthem, 2008, 'Logic and Reasoning: Do the Facts Matter?', *Studia Logica*, 88: 67–84.

- J. van Benthem, Girard, P., and Roy, O. 2009. Everything else being equal: A modal logic for ceteris paribus preferences. *Journal of Philosophical Logic* 38(1):83–125.
- J. Woods, 2013, *Errors of Reasoning - Naturalizing the Logic of Inference*, College Publications.

---

*A Logic for the Discovery of Causal Regularities*

MATHIEU BEIRLAEN

Ruhr University Bochum  
mathieubeirlaen@gmail.com

BERT LEURIDAN

University of Antwerp  
bert.leuridan@uantwerpen.be

---

In the past decades a large number of search algorithms for causal discovery has been developed. Perhaps the most influential theoretical framework for developing such search algorithms is the *causal Bayes' nets* framework which has given rise to, among many others, the **IC**- and **IC\***-algorithms by Pearl (2000, 2009) and the **SGS**-, the **PC**- and the **IG**-algorithms by Spirtes et al. (2000). As is well-known, these algorithms are based on a number of axioms, to wit the *Causal Markov Condition* and the *Faithfulness Condition* (which is also known as *Stability*).<sup>1</sup>

These conditions or axioms are not absolute. In the words of Spirtes et al. (2000, 9): “The Markov Condition is not given by God; it can fail for various reasons [...]. The reliability of inferences based upon the Condition is only guaranteed if substantive assumptions obtain. But the Condition is weak enough that there is often reason to think it applies.”<sup>2</sup> One domain in which the Causal Markov Condition fails, comprises deterministic causal structures (Baumgartner, 2009, 72). Such deterministic structures may also violate Faithfulness (Spirtes et al., 2000, 81ff).

In our paper we present a logic, called **ELIM'**, for the discovery of deterministic causal regularities starting from empirical data. It is an adaptive logic (we shall shortly explain what this means) that is inspired by Mackie's theory of causes as *inus*-conditions.

Mackie's theory of causes as *inus*-conditions is not probabilistic. It focusses on deterministic causal relations at the generic or type level. From Mill, Mackie borrows the idea that causation is seldom, if ever, an invariable sequence or regularity between a single antecedent (e.g. a short circuit) and a single consequent (e.g. a fire). Instead, it is often the case that the effect *P* occurs when some conjunction of factors (e.g. *ABC*; a short circuit, the presence of oxygen, the presence of inflammable materials) occurs, but not when any of these conjuncts fails to occur. Moreover, alternative conjunctions of factors (e.g. the conjunctions *DGH* and *JKL*) may also be followed invariably by *P*. *A*, in this example, is an *insufficient* but *non-redundant* part of an *unnecessary* but *sufficient* condition for *P*. In short, using the first letters of the italicized words, it is an *inus*-condition for *P*.

---

<sup>1</sup>For the Causal Markov Condition, see Spirtes et al. (2000, 54) and Pearl (2000, 30); for the Faithfulness Condition or Stability, see Spirtes et al. (2000, 56) and Pearl (2000, 48) respectively.

<sup>2</sup>In this quote they discuss the Markov Condition, but their claim applies to the Causal Markov Condition as well.

Mackie stresses the fact that our knowledge of complex causal regularities is seldom, if ever, complete. “What we know are certain *elliptical* or *gappy* uni- versal propositions.” (Mackie, 1974, 66) Moreover, he writes that “the elliptical character of causal regularities as known is closely connected with our characteristic methods of discovering and establishing them: it is precisely for such gappy statements that we can obtain fairly direct evidence from quite modest ranges of observation.” (Mackie, 1974, 68)

The logic to be presented in this paper will serve as an explication for Mackie’s views on these ‘characteristic methods’. As we will show, the gappy or elliptical character of our causal universal propositions gives the discovery of such propositions an interesting dynamics for which adaptive logics are well-suited.

Adaptive logics are tools for formalizing defeasible reasoning. They have been used to model a wide variety of reasoning patterns including explanatory reasoning, inductive generalization, and reasoning in the presence of inconsis- tencies. Moreover, Pearl’s **IC**-algorithm has served as the basis for **ALIC**, an adaptive logic for causal discovery (Leuridan, 2009). These patterns are non- monotonic: conclusions drawn from a set of premises may be withdrawn in the light of additional information (new premises). Adaptive logics are particu- larly suitable for capturing the non-monotonicity of defeasible reasoning. For a general introduction to adaptive logics, see Batens (2001, 2007).

The logic **ELIM<sup>r</sup>** consists of two preliminary logics: **ELI<sup>r</sup>** and **M<sup>r</sup>**. **ELI<sup>r</sup>** allows to derive logical equivalences — of a particular, Mackie-style type — from empirical data. **M<sup>r</sup>** then serves to minimize these equivalences; intuitively, **M<sup>r</sup>** serves to throw out redundant factors.

During our presentation, we will not spend too much time on the technical- ities of our approach. Instead, we will focus on:

1. giving a brief overview of Mackie’s theory of causes as inus-conditions with a special emphasis on the gappy or elliptical nature of our causal knowledge (and corresponding discovery methods);
2. giving a general overview of the preliminary logics **ELI<sup>r</sup>** and **M<sup>r</sup>** and the resulting logic for the discovery of causal regularities, **ELIM<sup>r</sup>**;
3. discussing the relation between our logic **ELIM<sup>r</sup>** and one recent and very interesting such discovery procedure for deterministic causation that also starts from Mackie’s theory, viz. the Boolean algorithm for coincidence analysis (*CNA*) proposed by Baumgartner (2009); and
4. discussing the relations between our logic **ELIM<sup>r</sup>** and some recent work on qualitative explications of inductive generalization and abductive or ex- planatory inference, viz. Batens (2011); Beirlaen & Aliseda (2014); Meheus & Batens (2006).

## References

Batens, D. (2001). A general characterization of adaptive logics. *Logique & Analyse*, 173–175:45–68.



- Batens, D. (2007). A universal logic approach to adaptive logics. *Logica Universalis*, 1:221–242.
- Batens, D. (2011). Logics for qualitative inductive generalization. *Studia Logica*, 97:61–80.
- Baumgartner, M. (2009). Uncovering deterministic causal structures: a Boolean approach. *Synthese*, 170(1):71–96.
- Beirlaen, M. and Aliseda, A. (2014). A conditional logic for abduction. *Synthese*, 191(15):3733–3758.
- Leuridan, B. (2009). Causal discovery and the problem of ignorance. An adaptive logic approach. *Journal of Applied Logic*, 7(2):188–205.
- Mackie, J. L. (1974). *The Cement of the Universe: A Study of Causation*. Clarendon Press, Oxford.
- Meheus, J. and Batens, D. (2006). A formal logic for abductive reasoning. *Logic Journal of the IGPL*, 14:221–236.
- Pearl, J. (2000). *Causality. Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.
- Pearl, J. (2009). *Causality. Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, 2nd edition.
- Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Cambridge, Massachusetts.

---

*A Resiliency-Based Approach to Chance*

PATRYK DZIURSZ-SERAFINOWICZ

University of Groningen

dziuroszerafinowicz@wp.pl

---

A number of recent philosophers have claimed that any notion of chance (objective probability) should satisfy certain conditions that appear to be constitutive of chance or can be regarded as platitudes about chance (e.g., Loewer 2001; Schaffer 2003, 2007). According to this line of thought, a notion of chance that violates these conditions either refers to something which only approximates genuine chance or does not refer to chance at all. Despite considerable discussion, there is no consensus among philosophers as to how many of such conditions a notion of chance should satisfy. Famously, David Lewis (1986) claimed that his Principal Principle captures all we know about chance. But, as shown by Frank Arntzenius and Ned Hall (2003), Lewis's claim cannot be rationally sustained. Similarly, Jonathan Schaffer (2007) has argued that besides the Principal Principle, there is a number of equally plausible conditions that inform our understanding of chance.

Two interesting, allegedly constitutive of chance, conditions have been proposed in Bigelow et al. (1993). To introduce these conditions, let  $Ch$  denote a chance function over a finite algebra of propositions  $A$  generated by a set of possible worlds  $W$ . Suppose that  $C_W = \{ch_w : w \in W\}$  is a finite set of other chance functions over  $A$ , indexed by the worlds in  $W$ . Further, let  $C_{ch_w}$  be the proposition that the chance distribution over  $A$  is given by  $ch_w$ . Following Bigelow et al. (1993, p. 458), call  $Ch$  the *present* chance function and the  $ch_w$ s the possible *later* chance functions. Assume that  $Ch$  and the  $ch_w$ s are probability functions over  $A$ . Then, the two conditions on chance may be presented as follows:

(C1) *Chance conditional on chance formulation*, i.e., for all  $A \in A$  and all possible later chance functions  $ch_w$ ,

$$Ch(A|C_{ch_w}) = ch_w(A).$$

(C2) *Chances are equal to the expected values of chances*, i.e., for all  $A \in A$ ,

$$Ch(A) = \sum_{w \in W} Ch(C_{ch_w} \in W) ch_w(A).$$

(C1) tells us that the present chance of some proposition  $A$  conditional on the proposition about some later chance of  $A$  should be set equal to that later chance of  $A$ . (C2) requires the present chance of some proposition  $A$  to be equal to the weighted average of possible later chances of  $A$ , where the weights are chances assigned by the present chance function to propositions about  $A$ 's possible later chances. Thus both (C1) and (C2) relate any present chance distribution to the possible later chance distributions in a certain specific way. It is convenient to frame the present chance function,  $Ch$ , the possible later chance functions, the  $ch_w$ s, and the relations between them covered by (C1) and (C2) within the theory of expert functions

developed by Haim Gaifman (1988) and Bas van Fraassen (1989, ch.8). This theory is framed in terms of higher-order probabilities. Each possible later chance function might be interpreted as a first-order chance function over  $A$ , and assuming that every proposition of the form  $C_{ch_w}$  belongs to  $A$ , the present chance function,  $Ch$ , might be regarded as a second-order chance function over  $A$  so enriched. Following Gaifman's terminology, we may interpret the possible first-order chance functions that figure in conditions (C1) and (C2) as 'expert' chance distributions for the present chance distribution.

The question arises: why should we believe that the two conditions are constitutive of chance? It seems that these conditions are neither trivially true nor do they follow from other platitudes about chance. My aim in this paper is to give a justification for conditions (C1) and (C2) by appealing to the resiliency of chance. I show that chances that violate these two conditions do not maximize resiliency suitably understood. Resiliency is taken to be a kind of stability property of chance: it reflects the approximate invariance of a chance distribution under variation of experimental factors that bear on a given chance set-up. I assume that probabilities that figure in statistical laws count as chances if they maximize resiliency over a given set of experimental factors. This idea draws on Brian Skyrms's (1977; 1978; 1980) resiliency-based account of chance. A modelling framework is introduced in which: (i) the experimental factors over which the resiliency of chance is evaluated form a partition, (ii) each cell of the partition singles out a possible world at which the cell holds true, (iii) we associate with every such world a possible later chance function defined as the present chance function updated by accommodating a given experimental factor (a given cell of the partition), (iv) for each possible world, we measure the resiliency of the present chance function by showing how close that chance function is to the possible later chance function at that world. Given this framework, it is then shown, for each of the conditions (C1) and (C2) separately, that (i) if a present chance function does not obey that condition, then there is another present chance function that is 'closer' to every possible later chance function, and (ii) if a present chance function obeys that condition, then there is no other present chance function that is closer to every possible later chance function. In other words, this result shows that any present chance function which violates that condition fails to maximize resiliency.

## References

- Arntzenius, F. and Hall, N. (2003). On What We Know About Chance. *British Journal for the Philosophy of Science*, 54(2):171–179.
- Bigelow, J., Collins, J., and Pargetter, R. (1993). The Big Bad Bug: What are the Humean's Chances? *British Journal for the Philosophy of Science*, 44(3):443–462.
- Gaifman, H. (1988). A Theory of Higher Order Probabilities. In Skyrms, B. and Harper, W., editors, *Causation, Chance, and Credence*, volume 1, pages 191–219. Morgan Kaufmann.
- Lewis, D. (1986). A Subjectivist's Guide to Objective Chance. In *Philosophical Papers*, volume 2, pages 83–132. Oxford University Press.
- Loewer, B. (2001). Determinism and Chance. *Studies in History and Philosophy of Science Part B*, 32(4):609–620.
- Schaffer, J. (2003). Principled Chances. *British Journal for the Philosophy of Science*, 54(1):27–41. Schaffer, J.

- (2007). Deterministic Chance? *British Journal for the Philosophy of Science*, 58(2):113–140.
- Skyrms, B. (1977). Resiliency, Propensities, and Causal Necessity. *The Journal of Philosophy*, 74(11):704–713.
- Skyrms, B. (1978). Statistical Laws and Personal Propensities. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1978:551–562.
- Skyrms, B. (1980). *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws*. New Haven: Yale University Press.
- van Fraassen, B. (1989). *Laws and Symmetry*. Oxford University Press.

---

*On the Preference for more Specific Reference Classes*

PAUL THORN

HHU

thorn@phil-fak.uni-duesseldorf.de

---

Typical instances of ‘direct inference’ satisfy the following defeasible inference schema (where “freq” denotes a function that takes a pair of sets, and returns the *relative frequency* of the first set among the second, and where PROB is a probability function that takes propositions as arguments, and is understood as designating the personal probabilities, or degrees of belief, that are rational for a respective agent):

From  $freq(T|R) = s$  and  $c \in R$  infer that  $PROB(c \in T) = s$ .

The doctrine that one should normally prefer frequency data for more specific reference classes in conducting direct inference is intuitively plausible. The prior intuitive plausibility of the doctrine probably explains why its advocates haven’t taken much care to argue for it, including Venn (1866), Reichenbach (1935), Kyburg (1974), Pollock (1990), Bacchus (1990), Kyburg and Teng (2001), and Thorn (2012). My aim in the present talk is to address this deficit.

In order to evaluate various ‘policies’ for forming personal probabilities, I introduce the notion of a population model  $M$ , which is a triple  $(U, T, \Pi)$ , consisting of a domain of objects,  $U$ , a subset  $T$  of  $U$  (where  $T$  stands for target class), and a partition  $\Pi = \{\pi_1, \dots, \pi_n\}$  of  $U$ , where  $\Pi$  corresponds to the set of maximally specific descriptions within which we are able to assign the elements of  $U$ .

For each object,  $x$ , in  $U$ , the task of a policy is to recommend a degree of belief in the proposition that  $x$  is in  $T$ . In other words, the task is to recommend a credence function,  $\chi$ , from  $U$  into  $[0,1]$ , which represents degrees of belief regarding the truth value of  $x \in T$ , for each  $x$  in  $U$ . Although the result presented below can be generalized, I begin by considering the case where it is known which objects are elements of which elements of  $\Pi$ , and our policies have access to the relative frequency of  $T$  among each element of  $\Pi$ , where  $freq(T|\pi) = |\{x : x \in \pi \wedge x \in T\}|/|\{x : x \in \pi\}|$ , for all  $\pi$  in  $\Pi$ . My intention here is to demonstrate the optimality of the following policy,  $\delta$ , which corresponds to using direct inference with the most specific applicable reference classes:

Relative to a respective population model  $M$ , let  $\delta(x \in T) = freq(T|\pi)$ , for all  $x$ , where  $\pi$  is the element of  $\Pi$  containing  $x$ .

While the policy corresponding to  $\delta$  is not optimal in comparison to all possible policies, with respect to all possible population models (for example, in comparison to the ‘oracular’ policy,  $\nu$ , that precisely tracks the truth value of all the relevant proposition, i.e.:  $\nu(x \in T) = 1$ , if  $x \in T$ , and  $\nu(x \in T) = 0$ , otherwise),

it is optimal (given a restriction on which accuracy measures we consider) in comparison to (the policies represented by) the following credence functions, whose value assignments are ‘principled’:

*Definition.* A credence function,  $\chi$ , is *principled* in  $M$  if and only if  $\forall \pi \in \Pi$ :

$$\forall x, y \in U: (x \in \pi \text{ and } y \in \pi) \Rightarrow \chi(x \in T) = \chi(y \in T).$$

The preceding definition tells us that a credence function is principled just in case, for each pair of objects, the same credence is assigned to both elements of the pair, regarding membership in  $T$ , if the two objects have exactly the same properties, among the set of properties that one is able to distinguish. Notice that  $\delta$  is principled. On the other hand, the restriction of our concern to principled credence functions excludes oracles, along with other policies that succeed by assigning different probabilities to objects that are indistinguishable, from the point of view of the policy.<sup>1</sup>

The optimality of  $\delta$  is dependent on how we measure accuracy. I here adopt the common parlance, and refer to accuracy measures as “scoring rules”. Formally, I here treat a scoring rule,  $S$ , as a function from pairs consisting of the credence assigned to a proposition,  $\chi(\alpha)$ , and the proposition’s truth value, as represented by a standard truth-valuation function, whose values are identical to those of oracular policy,  $\nu$ . So “ $S(\chi(\alpha), \nu(\alpha))$ ” would return the score for the credence function,  $\chi$ , regarding the proposition,  $\alpha$ , given  $\alpha$ ’s truth value,  $\nu(\alpha)$ . Since we only consider propositions that concern whether given elements of  $U$  are in  $T$  (according to a given population model  $M$ ), the application of scoring rules, in the present talk, takes the following form:  $S(\chi(x \in T), \nu(x \in T))$  (where everything is implicitly relativized to  $M$ ).

As it turns out, the optimality of the policy represented by the credence function  $\delta$  holds for a broad class of highly esteemed scoring rules, namely the set of all proper scoring rules (cf. Brier 1950, de Finetti 1974, Joyce 1998, Selten 1998, Greaves & Wallace 2006, Leitgeb & Pettigrew 2010, and Levinstein 2012). A scoring rule is proper just in case the expected score earned according to the measure is maximized by reporting one’s actual personal probabilities, i.e.:<sup>2</sup>

*Definition.*  $S$  is a *proper scoring rule* if and only if  $\forall M, r, s: r \neq s \Rightarrow S(s, 1) \cdot s + S(s, 0) \cdot (1 - s) \geq S(r, 1) \cdot s + S(r, 0) \cdot (1 - s)$ .

Strictly proper scoring rules satisfy the further condition:

*Definition.*  $S$  is a *strictly proper scoring rule* if and only if  $\forall M, r, s: r \neq s \Rightarrow S(s, 1) \cdot s + S(s, 0) \cdot (1 - s) > S(r, 1) \cdot s + S(r, 0) \cdot (1 - s)$ .

The first optimality result regarding  $\delta$  is as follows:

<sup>1</sup>Note that the oracular policy,  $\nu$ , will be principled in some population models, such as in population models where  $\Pi = \{\{x\}: x \in U\}$ . In all such cases,  $\delta(x \in T) = \nu(x \in T)$ , for all  $x$  in  $U$ .

<sup>2</sup>For the sake of uniformity, negatively oriented scoring rules (such as Briar scoring) are treated as *loss functions*, where the scores corresponding to such loss functions are determined by multiplying the loss earned according to such a rule by  $-1$ .

*Theorem.*  $\forall M, \chi$ : if  $\chi$  is principled in  $M$  and  $\chi \neq \delta$ , then  $\forall S$ : (1) if  $S$  is a *proper* scoring rule, then  $\forall \pi \in \Pi$ :  $\sum_{x \in \pi} S(\delta(x \in T), \nu(x \in T)) \geq \sum_{x \in \pi} S(\chi(x \in T), \nu(x \in T))$ , and (2) if  $S$  is *strictly proper*, then  $\sum_{x \in U} S(\delta(x \in T), \nu(x \in T)) > \sum_{x \in U} S(\chi(x \in T), \nu(x \in T))$ .

## Symposia & Contributed Papers V

**Newman's Objections to Structural Realism:**

**Symposium**

**New Approaches**

Organizer: Thomas Meier & Sebastian Lutz

Chair: Christian J. Feldbacher

Room 5D, Friday 09:30 – 11:30

---

*Overcoming Newman's Objection*

OTÁVIO BUENO

University of Miami

otaviobueno@me.com

---

*The Newman Problem and Ontic Structural Realism*

JAMES LADYMAN

University of Bristol

james.ladyman@bristol.ac.uk

---

*Newman's Objection is Dead, Long Live Newman's Objection!*

SEBASTIAN LUTZ

Ludwig-Maximilians-Universität München

sebastian.lutz@gmx.net

---

*A Carnapian Answer to Newman*

THOMAS MEIER

Ludwig-Maximilians-Universität München

thomas.meier@lrz.uni-muenchen.de



---

*Russell's Response to Newman: Space-Time Structuralism*

THOMAS PASHBY

University of Southern California

tom.pashby@gmail.com

---

**General Description**

Structural realism (sr) claims the ontological or epistemological primacy of relations over their relata. It has played an important historical role in philosophy and the philosophy of science and is if anything more influential in current debates about scientific knowledge. Unfortunately, Newman (1928) formulated an objection to Russell's sr (Russell 1927) that is often taken to show that all forms of sr are trivial. It is the aim of this symposium to provide precise accounts of sr that avoid Newman's objection.

A successful defense of sr would provide great rewards because of sr's central role in historical and current discussions about the nature of scientific knowledge. Historically, sr played an important role in the philosophies of, among others, Duhem, Poincaré, Carnap, and of course Russell (Gower 2000). A first, historical problem is the interpretation of Russell's position. Worrall (2007) and (taking their cue from Maxwell [1970]) Demopoulos and Friedman (1985) distinguish between observational and theoretical vocabulary and take Russell's sr to be explicated by the Ramsey sentence, which existentially quantifies on all theoretical relations. This is compatible with Russell's distinction between relations with which we are acquainted and those which we know only by description, but the modern notion of structure is not given by a quantification on relations. This points to the second problem in assessing early contributions to the sr debate, the differences in formal tools available then and now. Formal semantics in its current form was not developed until the 1930s, so when Russell and Newman speak of 'structure', they cannot have the exact modern notion of model theoretic structure in mind. Thus it is not obvious how their concept relates to modern notions from mathematical physics, mathematics, logic, or model theory.

In this symposium, Pashby contends that Russell's response to Newman indicates that his structuralist commitments are best explicated not by the Ramsey sentence but instead in terms sympathetic to Russell's attempted reconstruction of mathematical physics. Russell should be read as treating specific relations, spatio-temporal ones, as interpreted by both observable and unobservable objects. Lutz argues that Newman vacillates in his objection between two modern notions of structure, one involving quantifications over relations, and one involving isomorphisms between model theoretic structures (in today's sense). And while Pashby suggests specific relations as distinguished, Lutz argues that even if all relations are treated the same, the notion of structure based on isomorphism is non-trivial. Ladyman concludes that at least ontic sr

is immune even to generalizations of Newman's objection. He further outlines possible replies to a close kin of Newman's objection, van Fraassen's objection to "pure" structuralism.

In their contributions to the symposium, Bueno and Meier consider Carnap's version of sr in which "founded relations" are assigned a special status. Meier argues that, since Carnap considers founded relations to be experiential, they should be categorized as empirically accessible. This assumption invalidates an essential premise of Newman's objection and allows a formal explication of sr in structuralist meta-theory with its notion of intended applications. Bueno, on the other hand, points out that Carnap himself categorizes founded relations as logical relations and argues that under this assumption Carnap's version of sr fails. Instead of giving special status to specific relations, so Bueno argues, special status should be given to specific structures: Structures that have factual content. Since these are finite, he argues, Newman's objection fails. The replies to Newman's objection suggested in this symposium are broadly compatible, and thus span a logical space of possible viable accounts of sr. They also show that Newman's objection to sr is relevant for such diverse topics as the history of philosophy and logic, current metaphysics, philosophy of science, and physics. A reply to Newman's objection that satisfies the requirements of these diverse areas would not only provide a bridge between the areas, but also suggest further solutions of related problems. In physics, for instance, sr has been suggested as a means of generalizing and unifying theories, in metaphysics sr may provide an explication of natural kinds, and in the general philosophy of science sr may provide a new account of the semantics of scientific theories in which relations, not *relata*, are the primary referents of terms. We hope that this symposium will pave the way for such new results.

## Abstracts

### 1. Otávio Bueno: *Overcoming Newman's Objection*

Max Newman's (1928) objection to structuralism provides a formidable challenge—particularly to epistemological forms of the view, according to which all we can know about the world is structure. The objection threatens to turn structuralism into something trivial: as long as there are enough objects in the relevant domain, one can always obtain a structure suitable for that domain. But our knowledge of the world—even structural knowledge of it—is supposedly nontrivial. I consider two responses to this objection. One is provided by Rudolf Carnap's version of the objection that he considers in the *Aufbau* (Carnap [1928, sec. 154]; see Demopoulos and Friedman [1985] for an early discussion). Carnap suggests that the triviality can be avoided by requiring that the relations in question be *founded*. Surprisingly, however, he considers a founded relation as a basic concept of *logic*. I argue that, interpreted in this way, Carnap's solution fails. A more promising approach is to note that the relevant structures that have content about the world are ultimately *finite* (even though they may be, and typically are, embedded in infinite structures), and it is not a trivial matter to determine what is the appropriate structure for finite domains. This is a suggestion that

Russell briefly considers in his own response to Newman (Russell 1968, 176), but, unwisely, does not develop. It is a far more promising route than it may initially seem.

## **2. James Ladyman: *The Newman Problem and Ontic Structural Realism***

After reviewing the extant responses to the problem I argue that the ontic structural realist should not be troubled by the problem in the first place because the problem Newman raised for Russell's structuralism hangs on the identification of relations with sets. The question is whether there are generalisations of the problem for other accounts of relational structure. Various variants of the Newman problem are considered and it is argued that they do not apply to ontic versions of structural realism. However, the Newman problem is related to the problem van Fraassen poses for what he calls 'pure' structuralism (Ladyman et al. 2011, 439) and the options for replying to that problem are outlined.

## **3. Sebastian Lutz: *Newman's Objection is Dead, Long Live Newman's Objection!***

There are two ways of reading Newman's objection to Russell's structuralism. One assumes that according to Russell, our knowledge of a theory about the external world is captured by an existential generalization on all non-logical symbols of the theory. Under this reading, our knowledge amounts to a cardinality claim. Another reading assumes that our knowledge singles out a structure in Russell's (and Newman's) sense: a model theoretic structure up to isomorphism. Under this reading, our knowledge is far from trivial, for it amounts to knowledge of the structure of the relations between objects, but not their identity. In this sense, Newman's objection is but an expression of sr. In particular, classes of structures closed under isomorphism naturally describe the ontological commitments of epistemic sr and can be seen as a convenient circumscription of the ontological commitments of ontic sr. This entails that in sr theories are most naturally described syntactically and that the intuition behind the explication of sr by the Ramsey sentence is better explicated by the semantics of the logical empiricists' view on theories.

## **4. Thomas Meier: *A Carnapian Answer to Newman***

The Newman objection states that structural realism fails to specify a unique structure for the unobservable world, and hence it is ultimately compatible with an empiricist position. I propose a pragmatic version of structural realism, where the Newman objection does not hold. For this purpose, I will first discuss Carnap's notion of founded relations. According to Carnap, founded relations are real, experienceable, physical relations. If we rely on such founded relations when we specify a structural description of a given physical system, the threat of the Newman objection is avoided. As a second step, I will argue that the formal framework of the semantic conception of theories provides us with tools that allow us to overcome Newman's objection. This holds especially for the structuralist meta-theory and its notion of intended

application. However, pure structural realism has to be given up and in a Carnapian sense I propose a shift to a Pragmatic Structural Realism.

### 5. Thomas Pashby: *Russell's Response to Newman: Space-Time Structuralism*

Russell's response to Newman (by personal letter) has widely been taken to be a complete capitulation. While Russell does concede the central point, his fallback position is not so unprincipled or unworkable as it has often been assumed to be. While Russell begins by saying that he takes Newman's objection to have shown that his "statements to the effect that nothing is known about the physical world except its structure are either false or trivial," he goes on to point out those statements are even inconsistent with what he claims elsewhere in the *Analysis of Matter*. Russell's *Analysis of Matter* (1927) also contains an analysis of relativistic space-time in terms of spatio-temporal relations among an ontology of events, and it is these relations of which we have direct knowledge, he claims. Russell's *Problems of Philosophy* (1912) also contains the idea that we may be directly acquainted with universal relations that hold between both percepts and external objects. Furthermore, when the notion of structure reappears in *Human Knowledge* (1948) it is in particular spatio-temporal structure to which Russell commits. This conception of structure has never been properly addressed by the literature on structural realism since it cannot be captured by the notion of the Ramsey sentence nor allied model theoretic structures. Although Russell's attempted reconstruction of relativistic space-time from relations among events was not a success, recent developments in mathematical physics indicate how to successfully carry it out. The mathematical structures in question do not fall foul of Newman's objection.

### References

- Carnap, Rudolf. 1928. *Der logische Aufbau der Welt*. References are to the translation (Carnap 1967). Berlin-Schlachtensee: Weltkreis-Verlag.
- 1967. *The Logical Structure of the World. Pseudoproblems of Philosophy*. Translated by Rolf A. George. Berkeley and Los Angeles: University of California Press.
- Demopoulos, William, and Michael Friedman. 1985. "Bertrand Russell's *The Analysis of Matter*: Its Historical Context and Contemporary Interest." *Philosophy of Science* 52 (4): 621–639.
- Gower, Barry. 2000. "Cassirer, Schlick and 'Structural' Realism: The Philosophy of the Exact Sciences in the Background to Early Logical Empiricism." *British Journal for the History of Philosophy* 8 (1): 71–106. doi:10.1080/096087800360238.
- Ladyman, James, Otávio Bueno, Mauricio Suárez, and Bas C. van Fraassen. 2011. "Scientific Representation: A Long Journey From Pragmatics to Pragmatics." *Metascience* 20:417–442. doi:10.1007/s11016-010-9465-5.
- Maxwell, Grover. 1970. "Structural realism and the meaning of theoretical terms." In *Analyses of Theories and Methods of Physics and Psychology*, edited by Michael Radner and Stephen Winokur, 4:181–192. Minnesota Studies in the Philosophy of Science. Minneapolis, MN: University of Minnesota Press.
- Newman, M. H. A. 1928. "Mr. Russell's "Causal Theory of Perception"." *Mind* 37 (146): 137–148.
- Russel, Bertrand. 1912. *The Problems of Philosophy*. Oxford: Oxford University Press.
- Russell, Bertrand. 1927. *The Analysis of Matter*. London: Allen and Unwin.
- 1948. *Human Knowledge: Its Scope and Limits*. London: Allen and Unwin.
- 1968. *The Autobiography of Bertrand Russell*. Vol. 2. London: Allen and Unwin.

Worrall, John. 2007. "Miracles and Models: Why Reports of the Death of Structural Realism May be Exaggerated." *Royal Institute of Philosophy Supplement* 61:125–154. doi:10.1017/S1358246107000173.

**How is Reduction Achieved?**

**Symposium**

Organizer: Gergely Kertész

Chair: Vera Hoffmann-Kolss

Room 5F, Friday 09:30 – 11:30

---

*Reductive Explanation and Hypothetical Identities*

PETER FAZEKAS

University of Antwerp

fazekas.peter@gmail.com

---

*Mechanisms and Reduction in Chemistry*

ROBIN HENDRY

University of Durham

r.f.hendry@durham.ac.uk

---

*Mechanisms and Reduction in Psychiatry - An Interventionist Perspective*

LISE M. ANDERSEN

Aarhus University

lmandersen@cas.au.dk

---

*"Nothing-over-and-above-ness" without Reduction*

UMUT BAYSAN

University of Glasgow

e.baysan.1@research.gla.ac.uk

---

*Autonomy, Multiple Realization and the Way Reduction is Done*

GERGELY KERTÉSZ

University of Durham

gergely.kertesz@durham.ac.uk

**General Description**

Recent literature on reduction and reductive explanation has investigated how reduction is achieved in scientific practice. This reflects a shift of philosophical focus away from idealized reconstructions of reductive explanation, exemplified in Ernest Nagel's classical texts, to a more naturalistic philosophy of science that starts with the practice of science and tries to understand the role and structure of reductive explanation in science itself. In contrast to the classical reductionists, those working in reductionism's new wave try to understand how scientific research programmes at different scales or compositional levels co-evolve and guide each other in practice. The deductive ideal of theory reduction developed by the classical reductionists is replaced by mechanistic reduction that focuses on local relationships: the explanation of particular capacities of particular entities by the properties of their parts, and the interactions between them.

There are two kinds of reason for this shift of focus. Firstly there is the widespread failure of explanatory relationships between real sciences to fit the idealized model of classical reductionism. Fodor's and Putnam's examples of multiple realizability convinced many philosophers of science from the 70s onwards that higher-level sciences are typically autonomous. Contemporary reductionists in the philosophy of mind challenge multiple realizability (or at least its ontological significance), but the idea of autonomy is still with us in alternative, weaker formulations. Some mechanists, such as Carl Craver, are keen at least to allow for multiple realization in mechanistic reduction. Quite separately, detailed studies of theoretical explanations even within the physical sciences showed them to be messy and approximate, appealing to information from the higher-level sciences in ways that violate the ideal of Nagelian reduction. Once again, the higher levels seemed explanatorily autonomous.

Secondly, in the last three decades, following rapid advances in the brain sciences, the contemporary practice of the life sciences and psychology has become the main area of interest among philosophers. This area of scientific research is a strongly interdisciplinary endeavour, integrating biochemistry, biology, neuroscience, individual psychology and psychiatry, evolutionary and social psychology. How theoretical explanations are achieved in these sciences is quite different from how it is achieved in physics, which served as the ideal case for Nagel and other early reductionists. As William Bechtel argues, research in psychology is done by "looking down, around and up", which means that investigations at different levels and in different contexts are reinforcing each other and each is important in its own right because it might shed light on issues arising at other levels and in other contexts of investigation.

In our symposium we would like to focus on new wave reduction and the mechanistic idea, but we would also like to integrate into the discussion of reductive explanations the underlying metaphysical issues concerning higher-level autonomy and reduction. We are not only interested in what scientists are doing, but also in how can we make metaphysical sense of what they are doing.

The first and most general question we would like to tackle is the status of theoretical identities between different levels. What is involved in the act of reduction in the mechanistic context? What does it take to

establish links between entities, properties and processes at different levels? Can such links be made? Some models of reduction attribute a key role to a priori conceptual analysis in connecting the base and target levels, while others assume that reduction is an empirical activity, and that identities can be justified only by empirical evidence. On this second view, the identities themselves serve as premises in the deduction – the explanation of the target theory – and are justified by inference to the best explanation. In the view of Peter Fazekas, one of our contributors, none of these views are plausible descriptions of scientific practice. He argues for a view similar to what Bechtel and McCauley call heuristic identity theory. On this view, *hypothetical identities* are not justified prior to attempts at reduction, but only post hoc, by their success in guiding research practices. But after they became justified they serve as a sufficient reason for accepting real identities as underlying reductive explanation.

The new mechanistic philosophy is worked out mainly in the context of the life sciences: the first formulations from Darden concerned biochemistry, and from the 90s onwards – the decade of the brain – advocates like Craver and Bechtel started to concentrate on neuroscience and its relations to cognitive psychology. Surprisingly, no-one in this literature has examined mechanisms in chemistry, a field in which a series of Nobel prizes have been awarded for the identification of mechanisms. Robin Hendry will therefore scrutinize mechanistic reduction in this field of research, connecting it with important issues in the metaphysics of science. He will argue that, despite its direct employment of physical entities and processes, mechanistic explanation in chemistry is perfectly compatible with an emergentist interpretation of the relation between chemistry and physics. So, for Hendry, explanatory reduction is insufficient for accepting reduction in a metaphysical sense.

Lise Andersen will critically examine the methodology of a new reductive research program in psychiatry and cognitive neuroscience, pointing out that privileging the lowest levels of explanation in the practice of science can lead to category errors and the impoverishment of research in a field that is fundamentally inter-level, and where contextual factors are highly important in understanding and finding the causes and effects of mental illnesses. She will argue for a Bechtel-type, metaphysically less committing autonomy for higher level and contextual research.

The last two papers are more concerned with the metaphysics of science and reduction. The issue of multiple realizations still has a bearing on how reduction is understood, and related issues can only be tackled in the context of metaphysics. Umut Baysan's paper examines how to account for "nothing-over-and-above-ness" without relying on a metaphysical notion of reduction and reductive identities. According to some authors, the identification of different properties on different levels with each other has proved to be an impossible ideal because of the qualitative dissimilarities between them, others, like Craver, even though they are not committed to any metaphysical interpretation of mechanistic reductions, are willing to allow for this interpretation in the context of mechanistic explanations. Baysan will argue for a new formulation of the



realization relation that is capable of explaining metaphysical relationships between levels even in case metaphysical identification cannot deliver the job.

Gergely Kertész will argue that questions concerning the multiple realizability of higher level properties and of higher level autonomy cannot simply be decided by the investigation of scientific practice. Even if it is true, as Carl Gillett argues, that the qualitative differences between different levels of description are important in reductive explanations and by disregarding them we misinterpret the practice of science, the idea that there can be something in common to all the different realizers of the same higher level kind remains tenable and important in deciding the issue of higher level autonomy in a metaphysical sense.

## Abstracts

### 1. Peter Fazekas: *Reductive Explanation and Hypothetical Identities*

Reductive explanation advances our understanding of a target phenomenon by accounting for it in terms of some co-occurring lower level base phenomena. In this paper, I shall briefly review existing models of reductive explanation, and propose a novel approach that is designed to capture how reductive endeavours in actual scientific practice typically proceed. I shall argue that reductive explanations (contra one of the two major existing views) rely on explicit identities linking the target and the base domain. However, these identities (contra the other major existing view) are typically not evoked in order to explain correlations, but rather are *hypothesised* even before correlations could be recognised.

These, *hypothetical identities* are tools to anchor the target level to the base level. They are formulated on the basis of some initial similarities, and then they guide the mapping of target-level phenomena to base-level phenomena. They are justified if they are successful in this process of guiding mappings, i.e. if on the bases of them, one is able to uncover more and more connections, structural and functional similarities between the target and the base domains. Along these further structural and functional similarities, hypothetical identities make it possible to project the explanatory power of the base level onto the target level, and subsequently, to reductively explain the properties of target level phenomena on the basis of the properties of base level phenomena.

I conclude the paper by showing that this novel account of reductive explanation is compatible with both the modern reformulations of the classical Nagelian approach to reduction that re-introduce the role of bridge-laws into contemporary discourse, and with the so-called mechanistic model of explanation that has become the dominant view within the philosophy of life sciences in recent years.

### 2. Robin Hendry: *Mechanisms and reduction in chemistry*

The philosophical literature on mechanisms in science has concentrated almost exclusively on the biological and cognitive sciences. It has ignored chemistry, which offers the most extensive and significant body of mechanistic knowledge in science. Since the 1920s, chemists have developed a detailed, unified and

theoretically well-grounded body of theory about mechanisms, and sophisticated methods for testing mechanistic models experimentally. In this paper I will defend two general claims about mechanisms in chemistry.

Chemists understand reaction mechanisms in terms of structure at the molecular scale. They individuate substances in terms of their structures: to make a substance just is (from the chemical point of view) to bring its structure into existence. Hence to describe a mechanism is precisely to describe how the structures of the reagents interact to give rise to the structures of the products.

Since the 1920s, organic chemists have characterised mechanisms as involving the transfer of electrons. This is very natural, given two facts: (i) within organic chemistry, molecular structures are just atoms arranged in space, connected by chemical bonds; (ii) chemical bonds have long been understood in terms of the pairing or transfer of electrons between atoms.

In the light of these claims mechanistic explanation in chemistry seems to be a reductionist affair, in the broad sense that it involves the identification of entities and processes at higher compositional levels or larger scales with entities and processes at lower compositional levels or smaller scales. But I argue that it is not reductionist in the narrower sense that is of interest to philosophers of mind: it is perfectly consistent with the existence of genuine downwards causation, and hence strong emergence.

### **3. Lise Marie Andersen: *Mechanisms and Reduction in Psychiatry - an Interventionist Perspective***

The view that psychiatry should be elucidating the mechanisms behind mental phenomena is gaining momentum. This view, coupled with an intuition that such mechanisms must, by nature, be biological, has inspired the field to look to cognitive neuroscience for classification of mental illnesses. One example of this kind of reorientation can be seen in the Research Domain Criteria project (RDoC) introduced by the U.S National Institute of Mental Health. The RDoC project is an attempt to introduce a new classification system based on brain circuits. The central idea of this reorientation is that mental disorders can be understood in terms of brain disorders.

The problem with this kind of whole-scale reductionism is that multilevel models citing mental and social factors as part of the causal structures are rejected as non-scientific, or accepted only as provisional “stand-ins” for causal factors to be found at the biological level. However, it is precisely such multilevel models that are necessary for progress in this fundamentally interdisciplinary science. This paper analyses the reductive nature of the RDoC project and investigates the potential for an interventionist account of causation and mechanism to bridge the gap between mechanistic explanations and multilevel models of mental disorders.

### **4. Umut Baysan: *“Nothing-over-and-above-ness” without Reduction***

Are properties that are invoked in special sciences anything “over and above” physical properties? Some have thought that special science properties can be *reduced* to physical properties, hence that it can be shown

that, somehow, special science properties are “nothing over and above” physical properties. Different models of reduction have been put forward to account for how this could be done. One type of reduction, namely the *identification* of higher-level properties with lower-level properties, has been thought to be an impossible ideal, due to the observation that a certain higher-level property can be instantiated in virtue of the instantiation of different lower-level properties. Especially in philosophy of mind, this has been interpreted to be the demise of what has come to be known as *reductive physicalism*.

However, those who want to show that special science properties are “nothing over and above” physical properties do not have to appeal to reduction in the aforementioned sense. In fact, they do not have to appeal to *any* model of reduction. In this paper, I argue for the following schema: if a property  $P$  is “nothing over and above” a property  $Q$ , then there is some relation  $R$ , such that  $R(P, Q)$  explains that  $P$  is “nothing over and above”  $Q$ . Different relations can play the theoretical role that  $R$  is supposed to play. I argue that the *realization* relation is a plausible candidate to play this role. With some theoretical constraints as to how the “nothing-over-and-above-ness” of an entity relative to other entities can be shown, I will argue for a new formulation of the realization relation.

##### **5. Gergely Kertész: *Autonomy, multiple realization and the way reduction is done***

In the last decade we have witnessed a growing consensus that multiple realizability is a much less plausible view than it was supposed to be till the end of the last century, but the thesis still plays an important role in the discussion of reduction and reductive explanation.

The best arguments against autonomy and the antireductionist consensus are formulated in the works of Shapiro and Polger. One thought is that the original formulation implies only that a higher level property can be realized in various systems, but it does not guarantee that in these different systems the properties realizing the higher level property are not the same, which would be the only interesting case of multiple realization. Shapiro, relying on contemporary scientific practice, also provided detailed empirical arguments against the plausibility of the view that there exists such kind of multiple realization in our world. If these authors are right the autonomy thesis loses strength.

Other authors, like Carl Gillett, based on an alternative analysis of how reductive explanation is done in the sciences, are contending the former view. In case they are right, multiple realization is a frequent phenomenon which seems to support a stronger autonomy thesis.

In this paper I will compare the mentioned views of realization and I will argue that even if the more restrictive view is wrong and the permissive view on realization is tenable, applying the latter to the metaphysical problem of higher level autonomy is insufficient to secure that there are no properties in common on the lower levels and contrary to what Gillett argues, the question concerning multiple realization is a genuinely philosophical one, that cannot be decided solely by analysing scientific practice.

---

*Is Episodic Memory a Natural Kind?*

MARKUS WERNING

Ruhr University Bochum

markus.werning@rub.de

SEN CHENG

Ruhr University Bochum

sen.cheng@rub.de

---

Episodic memory is a critical part of the human mind and has frequently been claimed to be a cornerstone of personal identity. Yet, there is no universal consensus on what constitutes episodic memory. In many textbooks, the notion of episodic memory is introduced in a classical taxonomical manner – *per genus proximum et differentiam specificam*: In a first step, a distinction between declarative and non-declarative memory (Squire & Zola-Morgan, 1988) is made. In a second step, two subordinate categories are introduced within the superordinate category of declarative memory, namely, semantic memory and episodic memory. (Tulving, 1972) introduced the what-where-when (WWW) criterion to define the content of episodic memory. However, this criterion was found to be insufficient to distinguish semantic from episodic memories. As a result, Tulving later revised his definition of episodic memory based on auto-noetic consciousness, the conscious reliving of a past experience (Tulving, 1985). Suddendorf and Corballis (1997) went even further and suggested that episodic memory is linked to mental time travel into the past and facilitates mental time travel into the future.

Based on recent work ( , 2013, 2015), we digress from this classical taxonomical pathway. We will, instead, focus on the question of whether episodic memory is a natural kind and what implications this has for what episodic memory is best taken to be. The question whether memory, in general, is a natural kind has been addressed before by Michaelian (2010) who argues for a negative answer.

In the spirit of the homeostatic property cluster view, we will use the notion of a natural kind in the following sense:

A class C of entities is a natural kind if and only if there is a large set of properties that subserve relevant inductive and explanatory purposes such that C is the maximal class whose members are likely to share these properties because of some uniform causal mechanism.

We begin our investigation into whether episodic memory is a natural kind with the following Sequence Analysis of Episodic Memory (Cheng & Werning, 2015). The question will be whether there is a uniform neural mechanism that is a good candidate for the realization of episodic memory so characterized:

A subject  $S$  has episodic memory with content  $E$  at a time  $t_1$  if and only if the following conditions are fulfilled:

- (S1)  $E$  is an episode with  $E = \langle e_1, \dots, e_n \rangle$ .  $E$  is called the *mnemonic content*.
- (S2) At some time  $t_1$ ,  $S$  compositionally represents  $E$  as an episode of temporally succeeding events  $e_1, \dots, e_n$ .  $S$ 's representation of  $E$  at  $t_1$  is called the *mnemonic representation*.
- (S3) At a time  $t_0 < t_1$ ,  $S$  has a reliable experience of the temporally succeeding events  $e_1^*, \dots, e_m^*$ , which make up an episode  $E^* = \langle e_1^*, \dots, e_m^* \rangle$ .  $E^*$  is called the *experiential base*.
- (S4) The episode  $E^*$  occurs at or before  $t_0$  (*factivity*).
- (S5) The mnemonic content  $E$  is *ontologically grounded* in the experiential base  $E^*$  in the following sense of counterfactual dependence: Were  $E^*$  to occur at or before  $t_0$ ,  $E$  would also occur at that time.
- (S6)  $S$ 's representation with content  $E$  at  $t_1$  is *causally grounded* in  $S$ 's experience of  $E^*$  through a reliable memory trace.
- (S7) On the basis of its mnemonic representation with content  $E$ ,  $S$  is capable of generating a temporally explicit simulation with content  $E$  at some time  $t_2 \geq t_1$ . The generated simulation is called a *mnemonic simulation*.

These conditions can be related to the four major stages of memory processing: perception, encoding, storage and retrieval. (S3) and (S4) propose conditions on perception, (S5) and (S6) on encoding, (S1) and (S2) on storage, and (S7) on retrieval.

In the paper, we will argue that episodic memory as analyzed above indeed is a natural kind. Our argumentation will proceed along three cornerstones.

- (C1) The Sequence Analysis is both minimal and maximal with regard to its inductive and explanatory potential.
  - (C1.1) It is minimal because any violation of one of the conditions will lead to a deficiency in episodic memory.
  - (C1.2) It is maximal because other forms of memory do not satisfy the conditions nor do other cognitive processes.
- (C2) The principal anatomical substrate of episodic memory is the hippocampus.
  - (C2.1) The principal function of the hippocampus is episodic memory. That is, all processes hosted by the hippocampus contribute to episodic memory.

- (C2.2) Episodic memory is principally hosted by the hippocampus. That is: Even though episodic memory involves interactions with other cognitive processes, which are supported by a variety of brain regions, processes specific to episodic memory are hosted by the hippocampus.
- (C3) Neural processes in the hippocampus provide uniform causal mechanisms for the processing stages proposed by the Sequence Analysis.
- (C3.1) The hippocampus provides a uniform causal mechanism that aligns the sequential representation of mnemonic content with the sequential representation of the experiential base.
- (C3.2) The hippocampus provides a uniform causal mechanism for the compositional mnemonic representation of episodes and their mnemonic simulation in retrieval processes.
- (C3.3) Interventions in the memory trace warrant that mnemonic representations are causally grounded in experiences.

### References

- (2013). Composition and replay of mnemonic sequences : The contributions of REM and slow-wave sleep to episodic memory. *Behavioral and Brain Sciences*, 36(6), 610–611. doi:10.1017/S0140525X13001234
- (2015). What is Episodic Memory if it is a Natural Kind? *Synthese*, (January), 1–41. doi:10.1007/s11229-014-0628-6
- Michaelian, K. (2010). Is memory a natural kind? *Memory Studies*, 4(2), 170–189. doi:10.1177/1750698010374287
- Squire, L. R., & Zola-Morgan, S. (1988). Memory: brain systems and behavior. *Trends in Neurosciences*, 11(4), 170–5.
- Suddendorf, T., & Corballis, M. C. (1997). Mental time travel and the evolution of the human mind. *Genetic, Social, and General Psychology Monographs*, 123(2), 133–67. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/9204544>
- Tulving, E. (1972). Episodic and Semantic Memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–402). New York, NY, USA: Academic Press, Inc.
- Tulving, E. (1985). Memory and Consciousness. *Canadian Journal of Psychology*, 26, 1–26.

---

*Scientific Competition and Its Threat to a Neuroscience of Consciousness*

SASCHA BENJAMIN FINK

Otto-von-Guericke Universität Magdeburg

sfink@ovgu.de

---

In this talk, I explicate the notion of *scientific competition* (SC), give some historical examples for SCs, argue that SC is an essential feature of scientific research, and apply it to show that the search for neural correlates of consciousness (as currently understood) is not a science.

In most fields of scientific inquiry, there is a moment in time where there are several theories in a field of research concerning the same phenomena, sections of the world, or scientific subject matters. In all interesting cases, these theories are incompatible, i.e. they come to diverging and opposing predictions for certain cases. Thus, there will be some state of the world that is compatible with one theory but incompatible with the other. In these cases, how the world behaves gives “*instantia crucis*” — instantiations that help us to decide which theory we ought to accept and which one we ought to reject. Accepting both theories is then not an option, because this would lead to an incoherent view of the world. Call this incompatibility of scientific theories about the same subject matter “*scientific competition*”: If there is a science on some  $x$ , then there can be diverging theories  $T_1, T_2, T_3, \dots$  about the behavior of  $x$  which make opposing predictions for some circumstance  $c$ ; if we observe the prediction of one of these theories,  $T_i$  under  $c$ , this raises that  $T_i$ 's credibility while simultaneously lowering the credibility of its opponents.

SCs are common in all fields of science: Einsteinian and Newtonian physics both describe the behavior of light, but make different predictions how much rays of light will bend if they pass heavy stars; the “*aquatic-ape*”-theory (Westenhofer 1942, Hardy 1960, Roede 1997, Vaneechoutte et al. 2011) as well the “*monkey-fucked-a-pig*”-theory (McCarthy n.d.) of human evolution both explain our unique hairlessness and fat-deposition, but both make different predictions about possible fossil records and their location; both Beenken (1924), Hamann (1930), and Binding (1982) present theories about the age of the *Gero Kreuz*, but each makes different predictions about which historical and dendrochronological evidences may or may not be found. Lamme's Recursive Processing Hypothesis (RPH) (2006, 2004) and Tononi's Integrated Information Hypothesis (IIH) (2014, 2012a, 2012b, 2008, 2004) make predictions about which systems are conscious when, but they differ in some of their predictions.

SC is arguably an essential condition for fields of scientific research. We certainly can come up with diverging theories to explain phenomena, but we cannot assume that all of these diverging theories are adequate. Furthermore, SC is a common (if hidden) theme in prominent writings in the philosophy of science:

Popper (1935/1976: p. II.6) and Kuhn (2012/1962: 77f, 85) allude to it, and one can derive it from the basic presumptions of Bayesianism. I therefore suggest to see it as an essential feature of scientific endeavors.

Then, the notion of SC is potent enough to exclude certain fields of research as scientific. I suggest that the search for the neural correlate of consciousness (NCC) as it is currently performed is a candidate for exclusion because it does not fulfill the IC-condition.

Most scientist engaging in this research assert that there is one and only one neural correlate for any given type of experience, e.g. Crick and Koch (1998, 2003), Block (1998, 2005). This suggest that we ought to expect some intertheoretic conflict between NCC- hypotheses, e.g. between those of Lamme (2006, 2004) and Tononi (2014, 2012a, 2012b, 2008, 2004). Based on data from the visual cortex, Lamme suggests that experiences of any kind correlate with recursive processing. Tononi suggests that any kind of experience correlates with some amount of integrated information. But not every instance of recursive processing correlates with integrated information and vice versa. Thus, it seems that these hypotheses make different predictions.

However, both are NCC-hypotheses. Currently, the most prominent understanding of what an NCC is is due to Chalmers (2000): An NCC is that neural system or event which is minimally sufficient for some experience. Quite explicitly, the neural system or event is not necessary for that experience in order to allow for multiple realizations. This operationalization by Chalmers is widely accepted (see e.g. Crick and Koch (2003), Block (2005), Aru et al. (2012), Bayne and Hohwy (2013)).

If this operationalization is accepted and defines a field of research (namely NCC- research), then this has an unwelcome consequence: The hypotheses of Lamme and Tononi cannot stand in any competition, because each is – according to the NCC-operationalization only sufficient for experience. For logical reasons, no NCC-theory can stand in SC given the operationalization by Chalmers. (This may explain the fact that there is been a surge of “neuroscientific theories on consciousness” in the last decades, but not a single refutation. At most, some theories have been abandoned.)

My conclusion is then twofold: First, SC is a plausible criterion to demarcate science from pseudoscience; second, given SC as a necessary feature of scientific fields, then either NCC-research is unscientific or we cannot accept the NCC-operationalization by Chalmers.

## References

- Aru, J., Bachmann, T., Singer, W. and Melloni, L. (2012). Distilling the neural correlates of consciousness, *Neuroscience & Biobehavioral Reviews* **36**(2): 737–746.  
**URL:** <http://www.sciencedirect.com/science/article/pii/S0149763411002107>
- Bayne, T. and Hohwy, J. (2013). Consciousness: Theoretical approaches, in A. E. Cavanna (ed.), *Neuroimaging of Consciousness*, Berlin: Springer-Verlag, pp. 23–35.
- Beenken, H. (1924). *Romanische Skulptur in Deutschland (11. und 12. Jahrhundert)*, Leipzig: Klinkhardt & Biermann.
- Binding, G. (1982). Die datierung des sogenannten gero-kruzifixes im k“olner dom, *Archiv für Kulturgeschichte* **64**(1): 63–77.



- Block, N. (1998). How to find the neural correlate of consciousness, in S. R. Hameroff, A. W. Kaszniak and A. C. Scott (eds), *Royal Institute of Philosophy Supplement*, Cambridge, MA: MIT Press, pp. 23–34.
- Block, N. (2005). Two neural correlates of consciousness, *Trends in Cognitive Sciences* **9**(2): 46–52.
- Chalmers, D. J. (2000). What is a neural correlate of consciousness?, in T. Metzinger (ed.), *Neural Correlates of Consciousness: Empirical and Conceptual Questions*, Cambridge, MA: MIT Press, pp. 17–39.
- Crick, F. and Koch, C. (1998). Consciousness and neuroscience, *Cerebral Cortex* **8**: 97–107.
- Crick, F. and Koch, C. (2003). A framework for consciousness, *Nature Neuroscience* **6**(2): 119–126.
- Hamann, R. (1930). Studien zur ottonischen plastik, *Stadel-Jarhbuch* **1**: 5–19.
- Hardy, A. C. (1960). Was man more aquatic in the past?, *The New Scientist* **7**: 642–645.
- Kuhn, T. S. (2012/1962). *The Structure of Scientific Revolutions*, 50th anniversary edn, Chicago: University of Chicago Press.
- Lamme, V. A. F. (2004). Separate neural definitions of visual consciousness and visual attention: A case for phenomenal awareness, *Neural Networks* **17**: 861–872.
- Lamme, V. A. F. (2006). Towards a true neural stance on consciousness, *Trends in Cognitive Sciences* **10**(11): 494–501.
- McCarthy, E. M. (n.d.). Human origins: Are we hybrids?
- Oizumi, M., Albantakis, L. and Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0, *PLoS Computational Biology* **10**(5): 1–25.
- Popper, K. R. (1935/1976). *Die Logik der Forschung*, Tübingen: Mohr Siebeck.
- Roede, M. (ed.) (1997). *Aquatic Ape: Fact of Fiction: Proceedings from the Valkenburg Conference*, London: Souvenir Press.
- Tononi, G. (2004). An information integration theory of consciousness, *BMC Neuroscience* **5**(42): XXXX–XXXX.  
**URL:** <http://www.biomedcentral.com/1471-2202/5/42>
- Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto, *Biological Bulletin* **215**: 216–242.
- Tononi, G. (2012a). Integrated information theory of consciousness: an updated account, *Archives italiennes de biologie* **150**(2-3): 56–90.  
**URL:** <http://dx.doi.org/10.4449/aib.v149i5.1388>
- Tononi, G. (2012b). *PHI: A Voyage from the Brain to the Soul*, New York: Pantheon Books.
- Vanechoutte, M., Kuliukas, A. and Verhaeg, M. (eds) (2011). *Was Man More Aquatic in the Past? Fifty Years After Alister Hardy - Waterside Hypotheses of Human Evolution*, Bentham Science Publishers.
- Westenhöfer, M. (1942). *Der Eigenweg des Menschen. Dargestellt auf Grund von vergleichend morphologischen Untersuchungen u"ber die Artbildung und Menschwerdung*, Berlin: Verlag der Medizinischen Welt, W. Mannstaedt & Co.

---

*On the Plurality of Explanations in the Cognitive Sciences*

SILVANO ZIPOLI CAIANI

Università degli Studi di Milano

silvano.zipoli@unimi.it

---

According to a new generation of scholars, the study of cognition is at a crossroad (e.g.: Spivey 2007; Chemero 2009; Hutto & Myin 2012): either the cognitive sciences may continue to defend the assumption that cognition is an internal process based on the manipulation of symbolic states, or they may change their mindset, embracing a *dynamical* concept of mind and cognition. The dynamical revolution in the cognitive sciences currently strikes many people as the best available alternative to the informational mainstream. At the core of this view there is the assumption that cognitive phenomena should be explained using tools from the dynamical systems theory (DST), rather than from information theory. DST is a methodological framework imported from physics, and used in the cognitive sciences as an alternative to the information processing approach that has ruled the field since the dawn of the cognitive revolution.

DST views cognitive agents as dynamic physical systems which are best described as *non-linear* and *self-organizing*, rather than as linear and symbolic. Cognitive capacities are thus viewed like physical processes continuously evolving from chaotic to more stable trajectories in a theoretical state-space (van Gelder 1998; Beer 2000).

Notably, DST rests on two main assumptions. From a metaphysical point of view, DST states that the term “representation” should be expunged from the vocabulary of the cognitive sciences since there are not conclusive theories that provide a satisfactory naturalization of semantic properties. This takes the form of an *ontological* conclusion, according to which there is nothing in a cognitive system like a symbolic state. Moreover, from an *epistemological* point of view, DST states that our *best explanations* of cognitive phenomena do not involve reference to mental representations. Therefore, we should abandon an informational approach to cognition since it does not add any relevant explanation to the dynamical account of physical systems. The main difference between these two assumptions is that, whereas the first concerns the very identity of cognitive systems, the second concerns the methods of explanation in the cognitive sciences. While the former is a speculative hypothesis with virtually no impact on the methodology adopted in the study of cognition, the latter is in a much greater extent a hypothesis on how to carry out research in the cognitive sciences. On the basis of this consideration, the present study focuses on the epistemological claim, and takes aside metaphysical issues.

Contrary to DST, this paper maintains that to provide an explanation of cognitive phenomena in terms of chains of causal interactions between individual states may be more effective than framing cognition in terms

of continuous, non-linear processes. As a result, it will be shown that the a mechanistic approach to cognitive phenomena may provide the *best explanation* for addressing relevant issues in the cognitive sciences that cannot be addressed by DST. To support this claim, an abductive strategy is adopted.

The paper divides into two main sections. Section 1 outlines a series of epistemological desiderata which are required to set up an explanatory theory of cognition. Arriving at the best explanation involves establishing a set of criteria for assessing contending hypotheses. Solving the abductive puzzle, indeed, requires us to judge each of the contenders on the background of shared standards of explanation. Accordingly, any model of cognition should be able to fulfill the following constraints:

1. *Theoretical Unification*: a hypothesis best explains if integrates the explanandum with reliable background knowledge;
2. *Explanatory Consilience*: the best explanation is the hypothesis that covers the major number of empirical evidences;
3. *Epistemic Significance*: a hypothesis best explains if satisfies the relevant epistemic needs.

After having discussed the criteria, Section 2 gives rise to an assessment of DST. As it will be clear, DST fulfills constraints 1 and 2, but it is not able to satisfy the requirement 3. Indeed, DST has the merit to provide a theoretical unification between psychology and physics (1), and to cover all the relevant facts at the behavioral and neural level (2). However, as DST abandons the mechanistic methodology to embrace a purely nomological stance, it results unable to address questions concerning the individual causes of specific cognitive states and behaviors (3). As a consequence, since a dynamical model of cognition excludes the possibility of providing explanations in terms of individual causes and reasons to act, many of our cognitive practices are likely to remain without a satisfactory account.

In order to support this claim, the paper focuses on two paradigmatic case studies. On the one hand, it will be shown by means of empirical and conceptual analysis that DST is not able to deal with the explanation of *high level* cognitive phenomena such as the acquisition of memory and language. On the other hand, it will be shown that DST presents difficulties even when dealing with the explanation of *low level* cognitive skills, such as those concerning the perception for action. Notably, although DST could provide an account of *how* cognitive systems interact with the environment, it has no chances to explain *why* the environment prompts the agent to act in a certain way.

Finally, in Section 3, the take-home message is summarized: although DST integrates the cognitive phenomena with a corpus of well-established knowledge, it does not follow that it is always able to meet our explanatory needs. Accordingly, getting clear on the contextuality and plurality of cognitive explanations is not only at the heart of the theoretical debate, but also shapes our experimental goals in the cognitive sciences.

**References**

- Beer, R. (2000). Dynamical approaches to Cognitive Science. *Trends in Cognitive Science*, 4(3), 91–99.
- Chemero, A. (2009). *Radical Embodied Cognitive Science*. Cambridge (MA): MIT Press.
- Hutto, D., Myin, E. (2013). *Radicalizing Enactivism. Basic Minds Without Content*. London: Cambridge University Press.
- Spivey, M. (2007). *The Continuity of Mind*. New York: Oxford University Press.
- van Gelder, T. J. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, 21, 1–14.

---

*What are Phenomena in the Cognitive and Behavioral Sciences?*

ULJANA FEEST

Leibniz-Universität Hannover

feest@philos.uni-hannover.de

---

Authors in the mechanistic tradition sometimes use the term “phenomenon” to refer to macro-level behavioral regularities that can be “phenomenally” (functionally) characterized and decomposed (e.g., Bechtel & Richardson 1993; Bechtel 2008), or as a “repeatable event of product” (Craver & Darden 2013), but they can also sometimes be found to use the term as referring to a kind of object (e.g., spatial memory; e.g., Craver & Darden *ibid.*). It has also been a more or less unquestioned assumption in this literature that phenomena are not only the explananda of mechanistic explanations, but also that the discovery of mechanisms typically proceeds by way of constructing “higher-level” descriptions of the relevant explanandum phenomena. Some accounts even go so far as to suggest that such descriptions can coincide with the descriptions or explanations provided by psychology (e.g., Craver & Piccinini 2011). It is not my aim to question that neuroscientists are often engaged in the search for mechanisms. I do, however, wish to argue that the standard account of this search for mechanisms does not adequately capture the investigative process in cognitive and behavioral sciences, more broadly conceived. The aim of this process, I maintain, is to arrive at an understanding of particular *objects of research* (such as memory, attention, reasoning, emotions, etc.). Behavioral regularities, I will argue, play a vital role in this process, but they do so not by virtue of instantiating explanandum phenomena, but by virtue of providing evidence for specific *objects of research*, which are typically construed as behavioral capacities (see Cummins, 2000, for a similar point). Such objects of research are not helpfully characterized as phenomena. I will develop this point by drawing on (and developing) the more specific notion of a phenomenon as proposed by Bogen and Woodward (1988).

Bogen & Woodward’s notion of phenomenon has been very evocative and productive in the philosophical literature, but it also continues to pose questions. According to the authors, two distinctive features of phenomena are (1) that they are stable, general, and unobservable, features of the world, and (2) that phenomena differ from data, which are idiosyncratic to specific experimental contexts, are directly observable, and can function as evidence for phenomena. – Taking this distinction, some authors have argued that the behavioral regularities produced in the lab (=experimental effects) are mere data, which are used to make inferences about phenomena (e.g., Guala 2005; Craver & Darden 2013). However, this assumption is problematic for two reasons: First, it is not clear that experimental effects are “mere” data, but rather they share some features of Bogen and Woodward’s phenomena (Feest 2011). Second, what types of phenomena do scientists make inferences to on the basis of a given experimental effect? With respect to the second

question, several distinct possibilities are discussed in the literature. For example, scientists can make the inference (a) that a particular effect is indeed genuine (i.e., not an artifact of the experiment) (Guala 2005), or (b) that we can extrapolate from the existence of a particular effect inside the lab to the existence of phenomenon outside the laboratory (e.g., Cartwright 2007), or (c) that the effect really tells us something about a given object of research. In the first case, the inference is from the effect to a specific cause. In the second case, the inference is one that extrapolates from the cause-effect relationship in the lab to one that is also operative in the real world. If we want to use the language of phenomena here, we might say that in those cases the inferences are from one type of phenomenon (an experimentally produced behavioral regularity or effect) to a different type of phenomenon (a causal relationship or mechanism).

It is, however, the third kind of inference that I will specifically draw attention to in my talk, as it highlights the question of why scientists would be interested in experimental effects in the first place. I argue that in psychology the reason why scientists create effects is not (or not primarily) that they want to find an underlying mechanism. Nor do behavioral regularities (even when treated as phenomena) in and of themselves constitute interesting objects of research. Rather, they are investigated because they are taken to be indicative of, or instantiate, a given object of research that transcends the effect in question in several respects. To explain what I mean by this, I will (a) spell out my notion of an object of research in more detail and (b) discuss the question of how experimental effects, given that they are idiosyncratic to specific experiments, can become to be considered relevant to a particular object of research. These two tasks are primarily analytic: That is to say, I aim to fine-tune existing philosophical usages of “phenomena” in order to correct the skewed picture of discovery in cognitive science (particularly psychology) that has been created by an overly strong emphasis on mechanistic explanations. In this vein, I will argue that there is no good reason to suppose that objects of psychological research (such as memory) can be identified with one type of mechanism, but that they should rather be viewed as clusters of phenomena. Secondly, I will highlight the ways in which experimental effects (and the experimental paradigms that produce them) are conceptually tied to specific assumptions about a given object of research.

In addition to the analytic and descriptive project just outlined, I will also raise normative questions about criteria of adequacy for inferences from specific experimental effects to specific objects of research. I will argue that scientists do (and should) engage in an ongoing process of evaluation that I call “operational analysis.”

## References

- Bechtel, William (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. London: Routledge.
- Bechtel, William & Richardson, Robert (1993). *Discovering complexity: Decomposition and localization as strategies in scientific research*. Princeton: Princeton University Press.
- Craver, C. F. (2007). *Explaining the brain: The mosaic unity of the cognitive neurosciences*. Oxford: Clarendon.
- Bogen James & Woodward, James (1988). “Saving the phenomena.” *The Philosophical Review* 97(3), 303-

352.

Cartwright, Nancy (2007). *Hunting causes and using them. Approaches in philosophy and economics*. Cambridge University Press.

Craver, Carl & Darden, Lindley (2013). *In search of mechanisms. Discovery across the life sciences*. Chicago: The University of Chicago Press.

Cummins, R. (2000). "‘How does it work?’ versus ‘what are the laws?’: Two conceptions of psychological explanation." In F. Keil & R. Wilson (Eds.), *Explanation and cognition*. Cambridge, MA: The MIT Press, 117-144.

Feest, Uljana (2011). "What exactly is stabilized when phenomena are stabilized?" *Synthese* 182, 57- 71.

Guala, Francesco (2005). *The methodology of experimental economics*. Cambridge University Press.

Piccinini, Gualtiero & Craver, Carl (2011). "Integrating psychology and neuroscience: functional analyses as mechanism sketches." *Synthese* 183 (3) 283-311.

*Classical Limit of a (Macroscopic) Particle in a Box. A Suggested Solution to Einstein's Objection to Bohm's Theory (cancelled)*

DAVIDE ROMANO

University of Lausanne  
davide.romano@unil.ch

GUIDO BACCIAGALUPPI

University of Aberdeen  
g.bacciagaluppi@abdn.ac.uk

In 1953, Einstein raises an objection to the de Broglie-Bohm theory, arguing that it does not describe the real physical state of individual systems. For showing that, he considers a macroscopic object ("a bullet") of 1 mm in diameter moving back and forth between two perfectly reflecting and parallel walls, which are about 1 meter apart. The collisions between the bullet and the walls are supposed to be elastic. In the classical limit, the system under consideration is a particle that moves continuously and uniformly between the two walls with a definite position and momentum at any time. Einstein affirms that a realistic quantum theory, i.e., a theory that aims to describe the physical state of individual systems, should be able to account for this transition, under typical classicality assumptions.

At the quantum level, we can represent the system by a superposition of two plane waves with opposite momenta  $\pm p = \pm \hbar k$  and constant energy  $E$ :

$$\psi(x, t) = A e^{\frac{i}{\hbar}(Et+px)} + A e^{\frac{i}{\hbar}(Et-px)}$$

where  $A$  is a normalization constant. Therefore, the wave function of the system is a stationary wave:

$$\psi(x, t) = B \sin(kx) e^{\frac{i}{\hbar}Et}$$

where  $B$  is a normalization constant and  $k = \frac{n\pi}{L}$ , with  $n$  being a positive integer number and  $L$  the distance between the two walls.

Einstein's objection to the de Broglie-Bohm theory is the following: since the gradient of the phase of the wave is zero in each point between the walls:

$$\frac{\partial S}{\partial x} = \frac{\partial(Et)}{\partial x} = 0 \quad 0 < x < L$$

then the velocity of the particle is always zero ( $v = \frac{1}{m} \frac{\partial S}{\partial x}$ ), i.e., the particle remains at rest inside the box.

This is a very non-classical situation and, unfortunately, it does not change for  $\lambda = \frac{2\pi}{k} \ll L$ , considered by Einstein the typical condition for the classical limit, under which a (approximate) classical motion should emerge. Thus, he concludes, the de Broglie-Bohm theory does not really describe the physical state of individual systems. A first response to Einstein's objection is spelled out by Bohm himself and Hiley, in Bohm



and Hiley (1985) (1993, sections 8.2, 8.5). They observe that, even though having a particle at rest in a box (with finite energy) is a non-classical situation, nonetheless it is consistent with the static nodal structure that emerges from the solution of the stationary waves. The nodes are all the points in the box where the wavefunction is equal to zero: the formalism of quantum mechanics says that the probability to find the particle in one of these points is zero. The existence of such nodes, which increase in number in the classical limit (the condition  $\lambda \ll L$  is equivalent to  $n \gg 2$ ), is not consistent with any picture of an object moving back and forth between the walls, but it is perfectly consistent with a particle being at rest in a (non-nodal) point inside the box. Yet, the theory explains this apparently odd situation: for a particle in a box, the quantum potential  $Q$  is equal to the total energy  $E$ :

$$Q = \frac{\hbar^2}{2m} \frac{\nabla^2 R}{R} = \frac{k^2 \hbar^2}{2m} = \frac{p^2}{2m} = E_{tot}, \quad \text{where } \psi = Re^{\frac{i}{\hbar}S}$$

so the kinetic energy of the particle is 'absorbed' by the wave function *via* the quantum potential ( $E_k = \frac{p^2}{2m} = E_{tot} = Q$ ), and the particle remains at rest. Moreover, quantum mechanics has to be regarded as a universal theory, whose validity extends *prima facie* from the microscopic to the macroscopic range. There is no need to consider the condition  $\lambda \ll L$  as a universal condition of classicality: on the contrary, "it seems arbitrary to suppose that [...] quantum theory will fail beyond certain quantum numbers" (Bohm and Hiley 1985, p. 2513).

Indeed, according to Bohm and Hiley, we expect that the classical motion emerges from the de Broglie-Bohm theory only if the quantum potential is negligible, and this is not the case for a particle in a box, independently from the size of  $\lambda$ .

Following Bohm and Hiley, we do think that the weak point of Einstein's objection lies in the characterization of the classical limit he defines for the system. But, differently from Bohm and Hiley, we suggest that the classical motion of a quantum system should emerge not from the condition of closed system with a negligible quantum potential, but considering an open system interacting with its environment. Indeed, in a realistic framework for the classical limit, a macroscopic particle necessarily interacts with other quantum systems (being these an air molecule, a photon, a neutrino,...), getting entangled with them and, thus, leading to decoherence effects. The interactions between the system and its environment might play a crucial role for making clear the transition from a Bohmian particle at rest to a classical particle moving continuously and uniformly in the box. For this purpose, we shall propose a simple model of Bohmian decoherence, evaluating whether this new framework could finally lead to the "real description of an individual system" in the classical limit, i.e., the classical motion of a particle moving back and forth between the walls of a box.

## References

- D. Bohm (1952), A Suggested Interpretation of the Quantum Theory in Terms of "Hidden" Variables, I and II, *Physical Review* **85**, pp. 166---179 and 180---193.  
D. Bohm & B. J. Hiley (1985), Unbroken Quantum Realism, from Microscopic to Macroscopic Levels, *Physical Review Letters* **55**, pp. 2511---2514.

- Einstein (1953), *Elementare Überlegungen zur Interpretation der Grundlagen der Quantenmechanik*, in *Scientific Papers Presented to Max Born on his Retirement from the Tait Chair of Natural Philosophy in the University of Edinburgh* (Edinburgh: Oliver and Boyd), pp. 33–40. (English translation available on line by Dileep Karanth (2011): ‘Elementary Considerations on the Interpretation of the Foundations of Quantum Mechanics’ ).
- P. Holland, *The Quantum Theory of Motion. An Account of the de Broglie–Bohm Causal Interpretation of Quantum Mechanics*, Cambridge University Press, 1993.
- M. Schlosshauer, *Decoherence and the Quantum-to-Classical Transition*, Springer, 2007.

---

*On the Epistemic Interpretation of Quantum Mechanics*

FLORIAN BOGE

Universität zu Köln

boge@phil.hhu.de

---

Since the days of its very discovery, the status of state descriptions by wave functions in quantum mechanics (QM) has been the subject of heavy debates. These wave functions (or quantum states, more generally) have the curious feature that they do not assign definite values to all physical magnitudes at all times, and at any given time can only assign definite values to a limited subset of such magnitudes. Famously, a wave function which describes the system, say, as fairly localized in some region assigns no definite momentum to it, the precise numerical statement of which is Heisenberg's uncertainty relation  $\Delta x \Delta p_x \geq \frac{\hbar}{2}$ . Von Neumann's (1932) analysis of the measurement process made things even worse: for a whole host of ('nondemolition') measurements, the unitary time evolution of the state function would make the measurement apparatus become 'fuzzy' and indefinite in the above sense as well. But the indefiniteness could, arguably, be considered to spread further: Couple a sufficiently isolated cat to some deadly machinery which is engaged by a quantum mechanical process, and the cat should in principle evolve into a state of being dead and alive at the same time—or rather in a 'quantum superposition' of both (cf. Schrödinger 1935). On von Neumann's own account, only the act of observation should lead to a definite state of the cat, of being either dead or alive.

Some physicists, most notably Albert Einstein, have felt particularly uncomfortable about these consequences, and have hence urged to regard the wave function of a system merely as representative of our *knowledge* about its true, definite state, instead of a description of its actual physical constitution (cf. Einstein 1939). Despite its many difficulties, this view has become fashionable again in recent years. N. Harrigan and R. W. Spekkens (2010) have developed a formal framework for construing the quantum state of a system as something that merely reflects our knowledge about it.

This framework consists in building *ontological models* for QM, which basically means postulating a state space,  $\Lambda$ , that contains the 'ontic' (true, actual) states,  $\lambda$ , of physical systems, and is equipped with two (conditional) probability distributions  $p(\lambda | \hat{\rho}) =: p\hat{\rho}(\lambda)$  and  $\xi(k | \lambda, M) =: \xi_M^\lambda(k)$ . In contrast to the true states, the quantum state  $\psi$ , or more generally  $\hat{\rho}$  (allowing also for mixed quantum states), is interpreted merely as representative of a *preparation procedure* the system is subjected to. The distribution  $p\hat{\rho}(\lambda)$  is then called the 'epistemic state', and supposed to reflect the knowledge an observer has about the true state  $\lambda$  of the system, given that  $\hat{\rho}$  was prepared. The second one,  $M(k)$ , is called a 'response function', and is supposed to reflect uncertainties in the measurement procedure  $M$ , which leads to outcome  $k$ .

For the model to reproduce QM, it is required that

$$\int_{\Lambda} d\lambda p_{\hat{\rho}}(\lambda) \xi_{\lambda}(k) = \text{Tr}(\hat{E}_k \hat{\rho}), \quad (1)$$

with  $\hat{E}_k$  an element of a POVM (positive operator valued measure) and  $\hat{\rho}$  the statistical operator of the system. In words, summing up (integrating) all outcome probabilities  $M(k)$  for obtaining result  $k$  in measurement  $M$ , conditional on states  $\lambda$  which may possibly result from the preparation associated with quantum state  $\hat{\rho}$ , and weighted by the probability that a given  $\lambda$  results from this preparation, must reproduce the (generalized) Born probabilities for outcomes  $k$  in quantum state  $\hat{\rho}$ .

To incorporate the idea that quantum states are merely an indication of our knowledge, it is further required that there are cases where the supports of two distributions  $p_{\psi}(\lambda)$  and  $p_{\phi}(\lambda)$ , for two (pure) non-orthogonal quantum states  $\psi$  and  $\phi$ , overlap on a set  $\Delta$  of non-zero measure. The model is then called  *$\psi$ -epistemic* (Harrigan and Spekkens, 2010, p. 126). This makes the true states  $\lambda$  *ipso facto* hidden variables, which have bugged QM for many decades. Of course we know from Bell's (1964) theorem that such hidden variables must be *nonlocal*.

However, Spekkens (2007) as well as Bartlett et al. (2012) have developed actual models which—seemingly—reproduce many QM phenomena *locally*, based on epistemic restrictions any agent has to face w.r.t. the true states of systems. The most impressive successes of these models include the apparent reproduction of interferometric examples on a purely epistemic basis. Of course these ideas have also provoked a whole host of responses—most of them negative. The most influential response is certainly the PBR-theorem (Pusey et al., 2012) which puts clear restrictions on any possible  $\psi$ -epistemic model.

Following an article by Schlosshauer and Fine (2012), however, I am first going to argue that despite its popularity, the PBR-theorem is not necessarily the biggest threat to the  $\psi$ -epistemic approach. In the first place, it puts restrictions on *composite* quantum systems, instead of showing that  $\psi$ -epistemic models are generally flawed. However, further work inspired by the PBR-theorem aids to highlight deeper flaws in the reasoning of the  $\psi$ -epistemicist. Drawing on a theorem by Hardy (2013), and on the Reeh-Schlieder theorem (Reeh and Schlieder, 1961) from algebraic quantum field theory, I will then demonstrate that the  $\psi$ -epistemicist ultimately has to appeal to nonlocal elements of QM, in order for his models of interference to make sense, which, of course, undermines the apparent success. My argument is hence that the existing  $\psi$ -epistemic models do not even get off the ground, but falsely claim success in certain areas. My aim is to drive a further nail in the coffin of this kind of epistemic approach to QM, by showing that the  $\psi$ -epistemicist does not have as much of a leg to stand on as he claims.

## References

- Bartlett, S. D., Rudolph, T., and Spekkens, R. W. (2012). "Reconstruction of Gaussian quantum mechanics from Liouville mechanics with an epistemic restriction". *Physical Review A*, 86(1):012103 (25 pp).
- Bell, J. S. (1987 [1964]). "On the Einstein-Podolsky-Rosen paradox". In *Speakable and unspeakable quantum mechanics*, pages 14–21. Cambridge, New York: Cambridge University Press.

- Einstein, A. (2011 [1939]). Letter to Schrödinger. In Przibram, K., editor, *Letters on Wave Mechanics: Correspondence with H. A. Lorentz, Max Planck, and Erwin Schrödinger*. New York: Open Road Media.
- Hardy, L. (2013). “Are Quantum States Real?”. *International Journal of Modern Physics B*, 27(1-3):1345012 (19 pp).
- Harrigan, N. and Spekkens, R. W. (2010). “Einstein, Incompleteness, and the Epistemic View of Quantum States”. *Foundations of Physics*, 40:125–157.
- Pusey, M. F., Barrett, J., and Rudolph, T. (2012). “On the reality of the quantum state”. *Nature Physics*, 8:475–478.
- Reeh, H. and Schlieder, S. (1961). “Bemerkungen zur Unitäräquivalenz von Lorentzinvarianten Feldern”. *Nuovo Cimento*, 22:1051–1068.
- Schlosshauer, M. and Fine, A. (2012). “Implications of the Pusey-Barrett-Rudolph Quantum No-Go Theorem”. *Physical Review Letters*, 108:260404(1–4).
- Schrödinger, E. (1983 [1935]). “The Present Situation in Quantum Mechanics. A Translation of Schrödinger’s “Cat Paradox” Paper.”. In Wheeler, J. A. and Zurek, W. H., editors, *Quantum Theory and Measurement*. Princeton, New Jersey: Princeton University Press. Translated by John D. Trimmer.
- Spekkens, R. W. (2007). “Evidence for the epistemic view of quantum states: A toy theory”. *Physical Review A*, 75:032110.
- von Neumann, J. (1955 [1932]). *Mathematical Foundations of Quantum Mechanics*. Princeton: Princeton University Press. Translated by Robert T. Beyer.

---

*Cosmological Probabilities: General Relativity and Statistical Mechanics Writ Large*

C. D. McCoy

University of California San Diego

cdmccoy@ucsd.edu

---

Cosmologists and philosophers have occasionally advanced probabilistic arguments as a way of favoring or disfavoring particular cosmologies. In a similar vein, if some observational or theoretical feature can be shown to be “generic” among possible universes, it is felt that a satisfying explanation has been given for its presence; contrarily, if some such feature can be shown to be “special” among the possible universes, it is felt that a specific explanation must be offered for its presence. Examples of such special features of our universe, for which explanations have been urged, include its initial low entropy state (Price 2004), and its observed spatial isotropy and flatness (Linde 1990).

It would appear, however, that this notion of cosmological probabilities is an unscientific flight of fancy. How can there be an objective probability assignment to *possible universes*, as if its creation was by chance? Are we meant to picture the Creator, armed with a ‘pin’ which is used to pick out a universe at random in this space of possible universes (Penrose 1989)? As far as we know, there is just one universe, so we can certainly have no direct, empirical reason to suppose that the creation of the universe was a probabilistic trial. Thus one might suppose that we have no basis at all for taking probabilistic arguments in (single universe) cosmology seriously.

Yet this seeming flight has been taken by many earnest philosophers and physicists, on the grounds, one might plausibly suppose, that the justification for cosmological probabilities is provided by extrapolating justified applications of probability in physics to the universe at large. I attend to two prominent examples. First, on the well-known foundational proposal in statistical mechanics (SM) known now as the Past Hypothesis, following a tradition stretching back to Boltzmann, one supposes that the universe began in a low entropy state to which is associated a uniform probability distribution over the phase space region compatible with that state (Albert 2001). Second, Gibbons, Hawking, and Stewart (1987) (GHS hence) adapt the probabilistic apparatus of statistical mechanics to the general theory of relativity (GR) to derive a phase space measure of Big Bang type cosmologies as a means of determining the likelihood of cosmological inflation in the early universe. Both the Past Hypothesis and the GHS measure introduce a phase space measure of possible cosmologies, which, in the absence of technical problems, one uses to define cosmological probabilities.

I argue that the justification of cosmological probabilities fails in two cases associated with these examples. First, I deny that SM probabilities, as they are understood in the Past Hypothesis, are applicable to

the universe understood as a SM system. Second, I deny that GR probabilities are justified when they are introduced in a manner analogous to the successful introduction of probability in SM. Although other authors (Earman 2006, Curiel 2014, Schiff and Wald 2012) have already raised important technical and conceptual difficulties with these two approaches to cosmological probabilities, my arguments are independent of these and I believe more decisive. I summarize the arguments I make below.

On the Boltzmannian view SM systems possess microstates (points in phase space) and coarse-grained macrostates, the latter of which are represented by disjoint probability measures on the system's phase space. To interpret SM probabilities as probability assignments of possible universes, one must assume that the universe is a SM system and that microstates are "random variables." If the microstates are genuinely chancy in this way, then one must suppose that the "creation" of a SM system is itself a random trial, since the microstate dynamics is deterministic thereafter. The only way this can be empirically justified is if the statistics of the observables are in accord with the distribution of microstates consistent with its initial macrostate, which for familiar SM systems is reasonable. But this justification cannot be transferred to the universe as an SM system. It is not possible to know that microstates are distributed consistently with the universe's macrostate, since there is only a single universe.

Furthermore, I claim that it is not even clear that one should suppose that SM systems have a particular microstate, for all of the observable outcomes of the system are determined solely by the macrostate, i.e. the system's state is for all empirical purposes a probability measure. Indeed, the supposition that the state space of (classical) SM is phase space is dubious, since microstates are epistemically inaccessible and classical mechanics is false as a microphysical theory. If we suppose, then, that SM states are just probability measures, it is also not possible to consider the universe as a SM system, since the universe evidently does not have stochastic observables (or even stochastic dynamics) according to our best model of the universe. Thus SM probabilities are simply not applicable to the universe understood as a SM system.

I conclude by mentioning how analogous problems hold for GR probabilities, such as those introduced by GHS. One cannot hold that such GR probabilities have any empirical content, as neither the dynamics nor the observables of GR systems are stochastic. It is also not possible to hold, in the absence of a well-defined probability measure, that the phase space measure alone has empirical content. It strains credulity to suppose that the uniform phase space measure encodes distributions of GR subsystems of the universe, such as, for example, the distribution of black holes according to mass.

---

*Typicality in Multiverse Cosmology*

FERAZ AZHAR

University of Cambridge

feraz.azhar@gmail.com

---

A startling prediction of a range of modern cosmological theories is that there exist domains outside our observable horizon, where the fundamental constants of nature, and perhaps the effective laws of physics more generally, vary. These multi-domain universes (henceforth multiverses) can be described by inflationary theories (Vilenkin, 1983; Linde, 1983, 1986), and have also generated attention as a result of the discovery of the string theory landscape (Bousso and Polchinski, 2000; Kachru et al. 2003; Susskind, 2007).

A pressing question in this context is: how can we test the existence of such a multiverse? One indirect test is via a comparison between theory-generated probability distributions for observable parameters, and observations that we might make in our domain. Defining probability distributions in multiverse scenarios is, however, beset with difficulties (Aguirre, 2007). One must choose a measure to regulate infinities that arise in these scenarios, and even if one can solve this thorny issue, the presumed parsimony of any fundamental theory that describes a multiverse will likely render probabilities for our observations to be small (Hartle, 2007). To facilitate the required comparison between theory-generated probability distributions and our observations, it has been argued that anthropic conditionalization is needed. The means to achieve such conditionalization are problematic, however, as it is unclear who or what we should be conditionalizing on, as well as which physical parameters are needed in order to describe the object of the conditionalization.

Even if one manages to address these problems in a plausible way, a third stumbling block remains, which will constitute the focus of this paper. Namely, a suitable measure and conditionalization scheme might make our observations more likely, but how likely or *typical* should they be, before we can consider them to have provided support for the theory under consideration?

One means to address this question is through the principle of mediocrity (Vilenkin, 1995), which in more current terminology, argues that we should assume that we are typical of any reference class to which we believe we belong. A theory, measure, and suitable conditionalization scheme, which in combination, effectively define this reference class, will give rise to a probability distribution whose typical values constitute its predictions for what we might observe. The argument then goes that as long as our observations are typical according to the distribution, we can assume that our observations have been successfully predicted, and that these observations then provide support for the conjunction of the theory, measure, and conditionalization scheme under consideration.



In this paper, I analyze the efficacy of the principle of mediocrity in two complementary settings. First, I argue that under top-down conditionalization (Aguirre and Tegmark, 2005), namely, when we conditionalize our distributions by demanding consistency with (all relevant) experimental evidence, we cannot simply assume typicality (as argued by Garriga and Vilenkin (2008)), nor can we ignore typicality (Weinstein, 2006). I show in a concrete example related to dark matter, that typicality dramatically affects top-down predictions, exemplifying the sense in which errors in reasoning about typicality translate to errors in the assessment of predictive power. I thereby advocate a line of thinking promoted by Srednicki and Hartle (2010), who argue for the inclusion of a 'xerographic distribution' in the computation of probability distributions for observables, where these xerographic distributions encode a variety of assumptions regarding typicality. One then effectively tests the conjunction of a theory, measure, conditionalization scheme, and xerographic distribution, in comparing the resulting probability distribution with our observations.

In the second half of this paper, I test the principle of mediocrity, utilizing xerographic distributions in the context of a multiverse model that generalizes the cosmological model of Srednicki and Hartle (2010). I find that for a fixed theory, the assumption of typicality gives rise to higher likelihoods for our data. If, however, one allows the underlying theory and the assumption of typicality to vary, then the assumption of typicality does not always give rise to the highest likelihoods. Understood from a Bayesian perspective, these results show that when one has the freedom to vary both the underlying theory and the xerographic distribution, one should find the combination of the two that maximizes the posterior probability, and from this combination, one can then *infer* how typical we are.

In this way, I argue that not only are we not justified in assuming the principle of mediocrity in multiverse cosmological settings, but likelihoods generated by assuming the principle of mediocrity are not necessarily the most predictive ones. The principle of mediocrity therefore, is more questionable than has been recently claimed.

## References

- Aguirre, A. (2007). Making predictions in a multiverse: conundrums, dangers, coincidences. In *Universe or Multiverse?* ed. B. Carr (Cambridge: Cambridge University Press).
- Aguirre, A. and Tegmark, M. (2005). Multiple universes, cosmic coincidences, and other dark matters. *Journal of Cosmology and Astroparticle Physics* JCAP01(2005)003.
- Bousso, R. and Polchinski, J. (2000). Quantization of four-form fluxes and dynamical neutralization of the cosmological constant. *Journal of High Energy Physics* JHEP06(2000)006.
- Garriga, J. and Vilenkin, A. (2008). Prediction and explanation in the multiverse. *Physical Review D* 77, 043526.
- Hartle, J. B. (2007). Anthropic reasoning and quantum cosmology. In *Universe or Multiverse?* ed. B. Carr (Cambridge: Cambridge University Press).
- Kachru, S., Kallosh, R., Linde, A. and Trivedi, S. P. (2003). de Sitter vacua in string theory. *Physical Review D* 68, 046005.
- Linde, A. D. (1983). Chaotic inflation. *Physics Letters* 129B, 177–81.
- Linde, A. D. (1986). Eternally existing self-reproducing chaotic inflationary universe. *Physics Letters B* 175, 395–400.
- Srednicki, M. and Hartle, J. B. (2010). Science in a very large universe. *Physical Review D* 81, 123524.

- Susskind, L. (2007). The anthropic landscape of string theory. In *Universe or Multiverse?* ed. B. Carr (Cambridge: Cambridge University Press).
- Vilenkin, A. (1983). Birth of inflationary universes. *Physical Review D* 27, 2848–55. Vilenkin, A. (1995). Predictions from quantum cosmology. *Physical Review Letters* 74, 846–9.
- Weinstein, S. (2006). Anthropic reasoning and typicality in multiverse cosmology and string theory. *Classical and Quantum Gravity* 23, 4231–6.

**Historical, Social and Cultural Studies in****Contributed Papers****Philosophy of Science II**

Chair: Antonios Basoukos

Room 5H, Friday 09:30 – 11:30

---

*Continuities and Discontinuities Across Theory Change in Cassirer's Argument for a "Relativized" Conception of the A Priori*

FRANCESCA BIAGIOLI

University of Konstanz

francescabiagioli1983@gmail.com

---

Michael Friedman – who is one of the main proponents of a relativized conception of the a priori in the current debate – draws back the idea of a relativized a priori to Hans Reichenbach's (1920) distinction between two meanings of the notion of a priori in Kant's philosophy. The first is that a priori principles are valid for all time. The second meaning is that these principles are constitutive of experience, insofar as they provide nonempirical presuppositions for the definition of empirical concepts. This distinction suggests that, though the first meaning of the notion of a priori was disproved by Einstein's general relativity, the second meaning applies to Einstein's principle of equivalence as a coordinating principle linking Riemannian geometry to empirical reality.

Notwithstanding the importance of Cassirer's neo-Kantianism for the idea of a renewal of Kant's transcendental philosophy in its connection with the history of science, Friedman's objection is that Cassirer did not take into account discontinuities in the formulation of coordinating principles. For Friedman, Cassirer limited himself to point out that there is continuity across theory change, insofar as the mathematical structures of the previous theories can be included in that of the new theory. However, Cassirer's argument depends not so much on a formal-logical consideration of mathematical structures but on his analysis of the mathematical method in the works of mathematicians such as Bernhard Riemann, Richard Dedekind, and Felix Klein. These examples can be related to current variants of structuralism in mathematics, because they show that Cassirer's main interest lies in the transformation of nineteenth-century mathematics from a science of specific objects (i.e., of quantities) to the study of mathematical structures. At the same time, Cassirer observed that such abstract concepts as that of manifold and of group were introduced both in view of developments immanent to mathematics and in view of possible applications to physics. For example, Riemann and Klein expressed the conviction that a clarification of the foundations of geometry was necessary for a more comprehensive classification of the hypotheses that can occur in physics. Such a conviction was confirmed by Einstein's use of Riemannian geometry, which had hitherto been considered a purely mathematical speculation.

The historical aspect of Cassirer's approach sheds light on his argument for continuity across theory change because, in this case, Cassirer (1921) emphasized that classical mechanics and general relativity presuppose completely different geometrical hypotheses. He did not account for continuity in terms of inclusion of the structures of the former theories in that of the new one. Rather, continuity here depends on the symbolic character of mathematical thinking, which for Cassirer entails: 1) the possibility of correlating geometries with the empirical manifold of physical events; 2) the possibility of revising the principles of measurement in order for such a correlation to be univocal. My suggestion is that Cassirer's approach did imply a relativized conception of the a priori. Misunderstandings in later reconstructions of his argument depend on the broader scope of Cassirer's inquiry: the a priori role of the principles relative to special theories presupposes a comparison of hypotheses at the meta-scientific level of the formation of concepts. Therefore, Cassirer argued for continuity at this level. In doing so, he formulated an argument that is not subject to the same objections as the logical positivist argument. This is the argument that, even after such radical changes as general relativity, the consequences of the former theories can be derived as limiting special cases of the new theory. As pointed out by Thomas Kuhn, the problem of this view is that continuity can be showed only retrospectively, from the viewpoint of the latter theories. Although the laws derived in this way are special cases of relativistic physics, they are not Newtonian laws. "Or at least they are not unless those laws are reinterpreted in a way that would have been impossible until after Einstein's work" (Kuhn 1962, 101). Cassirer did not restrict his consideration to the retrospective view because he looked at the example of mathematics to explore the connection between symbolic thinking and future experience. It did not suffice to point out the limiting cases; Cassirer's goal was to show that even the new formulation of natural laws had been foreshadowed in the form of mathematical hypotheses.

---

*On the Role of Political Science Research in Philosophy of Science*

JAANA EIGI

University of Tartu

jaana.eigi@gmail.com

---

Recent philosophy of science is characterised by considerable interest towards the social aspects of science and some of these accounts also include proposals about changes to be made in the way science is organised and governed (see, e.g., Biddle 2013, Brown 2008, Solomon 2001). The aim of my presentation is to discuss what role research in political science could play in such a proposal.

Philip Kitcher (2001, 2011) has developed the model of “well-ordered science” as the basis for approaching the organisation of science in democratic society. This model requires considerable changes in the organisation of science. The most important of them is the radical rethinking of the process of setting the aims for scientific research: the priorities are to be assigned to research projects on the basis of understanding that the ultimate aim is to ensure equal chances to live a worthwhile life to all members of human race. With this aim in mind, deliberators representative of the human race make particular decisions, subject to tutoring by experts about epistemic significance of the research projects that are currently pursued. The decisions are made as a result of sympathetic mutual engagement in a conversation where deliberators learn about each other’s needs and decide upon the fair distribution of research effort to address them.

While “well-ordered science” is meant as an ideal, Kitcher’s recent discussion includes references to practical attempts to involve the public in deliberations on issues related to science and technology, such as deliberative polling and citizen juries. Discussing them, Kitcher presents a particular vision of the relation between philosophical proposals and research in political science: in Kitcher’s words, the political scientists show “How?” and his model indicates “Where?” to improve democratic deliberation (Kitcher 2011, 225).

In my presentation I argue against Kitcher’s view that the role of political science is best seen as offering some universal tools to be applied where the philosopher needs them. Instead, I suggest a more local and context-specific approach.

In my argument, I draw on Sheila Jasanoff’s (2005) comparative analyses in order to undermine Kitcher’s assumptions about the universal character of democracy and the context- independent and transferable character of forms of democratic deliberation. Jasanoff’s analysis of politics of biotechnology in Germany, Great Britain and the United States shows that countries with similarly well-established democratic traditions may have surprisingly different political cultures, with different standards for establishing the status of the expert and credibility of claims to knowledge made in the public sphere—what Jasanoff calls “a civic

epistemology”—as a part of these cultures. Thus, democracies may (and do, as Jasanoff shows on numerous examples) end up with remarkably different science and technology policies created and maintained as a result of different forms of deliberation and policy making, including different ways of engaging the public.

Jasanoff associates the belief in universal solutions with the early period of comparative studies—a belief that the growth of the field makes increasingly problematic:

With growing awareness of the culturally embedded character of both knowledge and policy, there are reasons to be sceptical of unproblematic learning from others' experiences. The insights gained from comparative analysis suggest, indeed, that neglecting cultural specificities in policymaking may be an invitation to failure within any political community's own terms of reference. (Jasanoff 2005, 15)

Thus, there are reasons to think that increasing attention to the forms that science policy in general and attempts to involve the public there in particular may take in a democratic society will not result in universally applicable instructions “How”. Learning from analyses of developments in science policy is not likely to result in the straightforward ability to direct and control. Instead, one could hope for an improved ability to understand the differences and to recognise the difficulties an attempt of change may face.

The appreciation of possibilities and obstacles of a particular local context may be used to support a different role for the philosopher of science that avoids another problem inherent in Kitcher's position. Kitcher seems to believe that philosophers are in position to point out “where” changes are to be initiated. One may wonder, however, whether there is political will—be it among policy makers, researchers or the public—to initiate and sustain such a philosophically motivated change.

I suggest that instead of trying to indicate where to begin, philosophers of science could draw on science policy analyses in order to understand where relevant developments in science policy are already happening, and become involved with them, attracting attention to the epistemic aspects of the changes that often have different—political and ethical—motivations. There already are some discussions of the possibility of such a local political engagement in philosophy of science (Kourany 2003) and accounts of a philosopher's experience doing that (Douglas 2010). I suggest that the engagement of philosophy of science with research in political science is best understood as developing further this emerging model of doing politically relevant philosophy of science and not as a search for context-independent tools to use in a reform of science from the ground up.

## References

- Biddle, Justin (2013). “Institutionalizing Dissent: A Proposal for an Adversarial System of Pharmaceutical Research”. *Kennedy Institute of Ethics Journal* 23(4), 325–353.
- Brown, James Robert (2008). “The Community of Science®”. In: Carrier, Martin, Don Howard and Janet A. Kourany (eds.) *The Challenge of the Social and the Pressure of Practice: Science and Values Revisited*. Pittsburgh: University of Pittsburgh Press, 189–216.
- Douglas, Heather (2010). “Engagement for Progress: Applied Philosophy of Science in Context”. *Synthese* 177(3): 317–335.
- Jasanoff, Sheila (2005). *Designs on Nature: Science and Democracy in Europe and the United States*.

Princeton: Princeton University Press.

Kitcher, Philip (2001). *Science, Truth, and Democracy*. Oxford: Oxford University Press. Kitcher, Philip (2011). *Science in a Democratic Society*. New York: Prometheus Books.

Kourany, Janet A. (2003). "A Philosophy of Science for the Twenty-First Century". *Philosophy of Science* 70(1): 1–14.

Solomon, Miriam (2001). *Social Empiricism*. Cambridge (Mass.) and London: The MIT Press.

---

*Well-being Intuitionism and Conceptual Adequacy in Well-being Science*

WILLEM VAN DER DEIJL

Erasmus University Rotterdam

vanderdeijl@ese.eur.nl

---

Different approaches to welfare measurement in economics and other social sciences do not only differ in terms of measurement procedures, but also employ different conceptions of well-being. How do we know which one of these has a better understanding of well-being? Anna Alexandrova (2012a, 2012b, 2014) has argued that the appropriate axiological standards for well-being science do not come from a theory in philosophy, because of the lack of a unifying philosophical account and because the current accounts are unhelpfully context-independent. I argue that the reason philosophy cannot play this role is not context-dependency, but different conceptual demands in science and philosophy. I argue that we can assess value adequacy on the shared pre-theoretical basis of intuitions of philosophical well-being theories. I spell out a methodology to arrive at a robust basis of pre-theoretical intuitions by means of a reflective equilibrium, and derive a number of practical criteria.

---

*The Law of Continuity, Determinateness, and the Mathematizability of Nature: Boscovich and his Contemporaries*

MARIJ VAN STRIEN

Max Planck institute for the history of science

marijvanstrien@gmail.com

---

The law of continuity is well-known as a fundamental principle in Leibniz' metaphysics. A version of Leibniz' law of continuity played a central role in eighteenth century physics, especially, but not exclusively, in the context of hard body collisions. In this paper, I discuss the role of the law of continuity in the work of Johann Bernoulli, Emilie Du Châtelet, and Roger Boscovich, around the mid- eighteenth century. I argue that for them, the law of continuity implied a correspondence between nature and mathematics, which made a mathematical treatment of nature possible. They provided arguments for the universal validity of the law of continuity in both natural processes and geometry. However, they ran into problems when confronted with developments within mathematics, specifically the possibility of discontinuous geometrical curves. This



undermined the argument of continuity in nature based on a correspondence with mathematics, and showed that the mathematizability of nature could not be based on such strict argumentation.

In the first part of the talk, I show that for Bernoulli, Du Châtelet and Boscovich, there are two versions of the law of continuity: (1) all geometrical curves are continuous and go through all intermediate values, and (2) all change in nature takes place gradually in time, and physical quantities can only change from one value to another by going through all intermediate values. These two versions of the law of continuity were so closely related that they were often not even distinguished. Thus, a core aspect of the law of continuity was a direct and unproblematic correspondence between nature and mathematics.

The law of continuity played a central role in the work in physics of Bernoulli, Du Châtelet and Boscovich. It ruled out the possibility of hard bodies, for a collision between hard bodies would involve a discontinuity in the direction of motion. It had implications for the laws of motion. And for Boscovich, it was a core assumption of his influential theory according to which matter consists of point particles, with a force acting between them which oscillates between attractive and repulsive as a function of distance. Furthermore, by implying a correspondence between nature and mathematics, the law of continuity ensured that physical processes could be represented by continuous mathematical functions, and that the rate of change of physical quantities could be described by differential equations. It was therefore essential for the mathematical treatment of nature.

The main part of my talk will be devoted to an examination of the basis of the law of continuity in the work of Bernoulli, Du Châtelet and Boscovich. I show that for Bernoulli and Du Châtelet (who explicitly based their ideas on Leibniz), the law of continuity was based on a metaphysics of causal determination. If there is a discontinuity between subsequent states of a system, this means that there is a break in causal connection, and the new state is not a continuation of previous states and therefore undetermined; this is metaphysically impossible.

Boscovich rejected this line of argument. However, he gave two other arguments for the law of continuity: one from induction, and one from metaphysics. According to the argument from induction, we see continuity everywhere in nature and in mathematics, and on this basis we are justified in inferring that there must always be continuity. I show that this argument is deeply problematic; in fact, the fact that the law of continuity leads him to a theory according to which matter is not continuous but consists of a discrete number of point particles undermines some of his examples of observed continuity. The argument from metaphysics goes roughly as follows: suppose that a physical quantity, e.g. the position of a body, changes as a function of time. The possible values of the quantity form a mathematical continuum, thus for any two different values, there is an interval in between. Now, suppose that at time  $t$ , the position changes discontinuously from A to B. Then there are two possibilities. Either A and B correspond to the same instant  $t$ . Then at  $t$ , the position has two different values, which is metaphysically impossible. Or A and B do not correspond to the same instant  $t$ . In that case, there must be an interval in between, in which the quantity either has no value

(if A comes before B) or two values (if B comes before A); this is also metaphysically impossible. Thus, no such discontinuous changes are possible. Ultimately, the argument is based on a denial of open intervals.

A relevant aspect of the arguments of Du Châtelet and Boscovich is that they are applied both to nature and to geometry. The arguments do not only show that natural processes are always continuous, they are also meant to be strict demonstrations of the continuity of all 'proper' curves in geometry. A 'proper' curve here is any curve that can be expressed through a single algebraic expression: a discontinuity in such a curve must be impossible. This applies not only to jump discontinuities but also to sharp bends, which represent discontinuous changes in the derivative of the curve.

Du Châtelet and Boscovich thus defended a strong claim of continuity in geometry; however, this claim turned out to be problematic. They were confronted with the fact that there are curves which are generated by a single algebraic expression but which do have sharp bends: an example is the curve  $x^2+y^3=0$ , which has a sharp bend in the origin. There seems to be no reason to invalidate this curve as a proper geometrical curve.

Du Châtelet and Boscovich both sought for ways to deal with this type of discontinuity in geometry, and to argue that it does not involve a violation of the law of continuity: Du Châtelet argued that the curve can be understood as a limit of continuous curves, and Boscovich argued that the tangent can be understood as changing continuously. However, their arguments were problematic. Ultimately, their attempts to base the mathematizability of nature on strict metaphysical argumentation therefore failed.

## Symposia &amp; Contributed Papers VI

Science within Metaphysics and Metaphysics within Science:

Symposium

Articulating the Relationship between Metaphysics of  
Science and Traditional Metaphysics

Organizer: Thomas Pradeu

Chair: Juha Saatsi

Room 5D, Friday 13:00 – 15:00

*Metaphysics and Science: Rationalism and Empiricism*

HELEN BEEBEE

University of Manchester

helen.beebee@manchester.ac.uk

*Building Bridges with the Right Tools: Modality and the Standard Model*

STEVEN FRENCH

University of Leeds

s.r.d.french@leeds.ac.uk

*Situating Metaphysics of Science: Back to Square One*

ALEXANDRE GUAY

Université catholique de Louvain

alexandre.guay@uclouvain.be

THOMAS PRADEU

CNRS, University of Bordeaux

thomas.pradeu@u-bordeaux.fr

*Are the Questions of Metaphysics More Fundamental than Those of Science?*

ALYSSA NEY

University of Rochester

alyssa.ney@rochester.edu

**General Description**

In recent years, “metaphysics of science” has become a highly dynamic subfield within philosophy of science. Yet it is not always easy to understand what this label means. The project of this symposium stemmed from

the realisation that the phrase “metaphysics of science” is used in quite diverse senses, and covers in fact quite different research projects. In an attempt to contribute to clarifying what the meaning and aims of “metaphysics of science” are, we suggest distinguishing at least three different projects, all defended in recent literature:

- i) “Metaphysics of science” as an attempt to replace traditional and analytic metaphysics by a “scientific metaphysics”, which rests on the conviction that our best metaphysical views should reflect directly our best scientific theories (e.g., Maudlin 2007; Ladyman and Ross 2007; Ross, Ladyman and Kincaid 2013). Here metaphysics of science is intrinsically very close to science, most often to one specific scientific field, such as physics, biology or the social sciences, and it is built mainly by philosophers of science or sometimes by philosophically-inclined scientists.
- ii) “Metaphysics of science” as an attempt to offer an account of highly general notions supposedly found in all the sciences, as for examples the notions of law or causation (e.g., Mumford and Tugby 2013; see also Lewis 1973 on causation, and Armstrong 1983 on laws). Here metaphysics of science is rather a general metaphysics *applied to* science. It is at a distance from science, and a fortiori at a distance from any specific science, and it is mainly built by metaphysicians, rather than philosophers of science or scientists.
- iii) “Metaphysics of science” as an attempt to build something like a “reflective equilibrium” between traditional metaphysics and metaphysical lessons taken from current science (e.g., Morganti 2013; Guay and Pradeu 2015). Here proponents of metaphysics of science always keep an eye on current sciences, but consider that metaphysics of science cannot be a direct reflection of those current sciences. Rather, they use both the concepts and the views built by a long metaphysical tradition and they confront them with current scientific worldviews (something that even some logical positivists did: see Ney 2012). One recent illustration of this attitude is the suggestion to use traditional metaphysics as a potential “toolbox” for the construction of an adequate metaphysics of science (French and McKenzie 2012; French 2014). An important challenge, at this stage, is to determine whether or not current science and metaphysics of science could in turn be used as “toolboxes” for traditional metaphysics.

Through a discussion of these three different research programs that have all used the label “metaphysics of science”, the aim of this session will be to clarify the nature of the relations between metaphysics and science, by showing how metaphysics can be used by science, and vice versa. Stepping back from the current metaphysical “battlefield” (in which “scientific metaphysicians” tend to argue that metaphysics is just too far from science to be legitimate, and traditional metaphysicians tend to argue that science-based metaphysics is nothing more than a naïve acceptance of current scientific worldviews), we deliberately adopt a pragmatic attitude that emphasises dialogue and fruitful interaction: first, we claim that metaphysics of science can be built only by people who know the details of both current science and traditional metaphysics; second, we suggest that the most productive results are likely to stem from the precise articulation of the inter-relationships between these domains. Key concepts such as causation or individuality, fundamentality and

dependence deployed in the context of key scientific fields such as quantum mechanics or cognitive sciences will serve to flesh out this articulation.

### References

- Armstrong D. (1983). *What Is a Law of Nature?*, Cambridge: Cambridge University Press.
- French S. (2014). *The Structure of the World: Metaphysics and Representation*. Oxford University Press.
- French S., & McKenzie K. (2012). Thinking outside the (tool)box: Toward a more productive engagement between metaphysics and philosophy of physics. *European Journal of Analytic Philosophy* 8 (1): 42-59.
- Guay A. and Pradeu T. (2015). Progressive steps towards a unified conception of individuality across the sciences, in Guay and Pradeu (eds.) *Individuals Across the Sciences*. New York: Oxford University Press.
- Ladyman J., & Ross D. (2007). *Every thing must go: Metaphysics naturalized*. Oxford: Oxford University Press.
- Lewis D. (1973). Causation. *Journal of Philosophy* 70: 556–567.
- Maudlin, T. (2007). *The Metaphysics within Physics*. Oxford: Oxford University Press.
- Morganti M. (2013). *Combining Science and Metaphysics: Contemporary Physics, Conceptual Revision and Common Sense*. Palgrave.
- Mumford S. and Tugby M. (2013, eds.), *Metaphysics and Science*. Oxford: Oxford University Press.
- Ney A. (2012). Neo-positivist metaphysics. *Philosophical Studies* 160 (1): 53-78.
- Ross D., Ladyman J. and Kincaid H. (2013, eds). *Scientific Metaphysics*. Oxford: Oxford University Press.

### Abstracts

#### 1. Helen Beebe: *Metaphysics and science: rationalism and empiricism*

This paper addresses the second conception of ‘metaphysics of science’ described in the symposium description, according to which metaphysics of science is ‘an attempt to offer an account of highly general notions supposedly found in all sciences’, e.g. laws and causation.

My starting-point is the conception of metaphysics and its evidential (or, perhaps better, quasi-evidential) base, as enshrined in much of contemporary ontology and described explicitly by Paul (2010, 2012). On this view, roughly, theory construction in ontology proceeds by appealing primarily to *a priori* forms of reasoning, deploying principles whose justification (at least in the domain of ontology) does not derive from empirical success, such as appeals to simplicity, parsimony and inference to the best explanation. And the ‘data’ that supports (or conflicts with) ontological theories comes from ordinary judgements (or ‘intuitions’) that are perceptual or ‘quasi-perceptual’ in nature.

The paper asks two questions: (a) to what extent does the above conception of metaphysics constitute a reversion to rationalism?, and – to the extent that it does constitute such a reversion – (b) is this rationalist conception of the nature of reality and our epistemic access to it really viable? My basic contention will be that the answers to these questions are ‘to quite a great extent’ and ‘no’: there are good (and familiar) reasons to deny that discovery of the nature of things can be achieved on the basis of ‘intuition’ and *a priori* reasoning. I argue that old-school conceptual analysis, while much maligned of late, has much to offer when it comes to a rapprochement between metaphysics and science.

**References**

- Paul, L.A. 2010. 'A new role for experimental work in metaphysics', *Review of Philosophy and Psychology*, 1: 461-76.
- 2012. 'Metaphysics as modeling: the handmaiden's tale', *Philosophical Studies*, 160: 1-29

**2. Steven French: *Building Bridges with the Right Tools: Modality and the Standard Model***

The relationship between metaphysics and the philosophy of science has often seemed to be antagonistic, with swathes of current metaphysics dismissed for being out of touch with developments in modern physics (see for example, Ladyman and Ross 2007; Callendar 2011). However, more recently it has been suggested that although there is some force to this last claim, metaphysics can nevertheless be taken to offer a kind of 'toolbox' from which philosophers of science can select various devices, manoeuvres and frameworks to help in the interpretation and understanding of scientific theories (French and McKenzie 2012 and forthcoming; French 2014). In the current paper I develop this approach further in the context of the metaphysics of modality: with Humean and dispositionalist approaches to the laws and symmetries of physics (as embodied in the so-called Standard Model) regarded as inadequate, conceptual space has opened up for alternative approaches to physical possibility in this context. Here I indicate how some recent considerations in the metaphysics of possibility might be drawn upon to help outline and explore such approaches. I conclude by discussing the concern that such a toolbox approach is inherently unstable insofar as it both rejects and makes use of current metaphysical moves and I use the example of physical possibility to help assuage such worries.

**References**

- Callender, C., 'Philosophy of Science and Metaphysics', in *The Bloomsbury Companion to the Philosophy of Science*, ed. S. French and J. Saatsi, 33-54. London: Bloomsbury, 2011.
- French, S. *The Structure of the World: Metaphysics and Representation*. Oxford: Oxford University Press, 2014.
- French, S. and McKenzie, K. 'Rethinking Outside the Toolbox: Reflecting Again on the Relationship Between Philosophy of Science and Metaphysics', forthcoming in T. Bigaj and C. Wuthrich (eds), *Metaphysics in Contemporary Physics*, Poznan Studies in the Philosophy of the Sciences and the Humanities.
- French, S. and McKenzie, K. 'Thinking Outside the (Tool)Box: Towards a More Productive Engagement Between Metaphysics and Philosophy of Physics', *The European Journal of Analytic Philosophy* 8, pp. 42-59, 2012.
- Ladyman, J. and Ross, D., *Everything Must Go*, Oxford University Press, 2007.

**3. Alexandre Guay and Thomas Pradeu: *Situating Metaphysics of Science: Back to Square One***

Several advocates of the active field of "metaphysics of science" have recently argued that a truly naturalistic metaphysics should be based solely on current science, and that it should replace more traditional, intuition-based, forms of metaphysics. In this paper, we aim at clarifying this claim. We describe the current metaphysical battlefield of metaphysics not as a dichotomy, but as a "square", with descriptive metaphysics, revisionary metaphysics, commonsense metaphysics and metaphysics of science as its four corners. We use

this square to better determine the project of today's "naturalistic" metaphysics of science, and to demonstrate that the target of this project is not defined with enough precision and clarity. This leads us to defend the view that metaphysics of science should be seen as a form of general metaphysics that needs other forms of metaphysical projects to achieve its aims. We conclude by endorsing a wider and more peaceful understanding of "metaphysics of science."

#### **4. Alyssa Ney: *Are the Questions of Metaphysics More Fundamental Than Those of Science?***

When pursued naturalistically, fundamental metaphysics may seem forced to navigate a narrow path. So that it may be a worthwhile enterprise, it must have claim to discovery of a distinctive set of objective truths. Yet it must also avoid potential competition or conflict with fundamental scientific theories. And so this seems to require that metaphysicians avoid those topics addressed by scientific theories. This would threaten to exclude most of the traditional areas of fundamental metaphysical research. In response to this problem, some (naturalistic) metaphysicians have argued that properly understood, the traditional problems of metaphysics are not aimed at discovering a realm of truths also studied by science. Instead metaphysicians investigate a realm of truths more fundamental than those of fundamental science. Laurie Paul (2013) has developed the best worked-out defense of this claim. This paper brings the author's recent work on fundamentality to bear on this issue, first examining what is required both in science and metaphysics for a theory to count as a fundamental theory. Some criteria are presented which suggest that metaphysics does not investigate a realm more fundamental than that of science. This then raises the question of how metaphysics can have a distinctive subject matter without stepping on the toes of science. I argue that metaphysics does not need to have a distinctive subject matter to be a worthwhile enterprise and sketch some problems in current research that require collaboration between metaphysics and physics.





**Imprecise Probabilities****Symposium**

Organizer: Gregory Wheeler

Chair: Thomas Müller

Room 5F, Friday 13:00 – 15:00

*The Epistemological Significance of Imprecise Probability*

JON WILLIAMSON

University of Kent

j.williamson@kent.ac.uk

*What do (Im)Precise Credences Represent?*

JENNIFER CARR

University of Leeds

j.carr@leeds.ac.uk

*Reply to Carr and Williamson*

SEAMUS BRADLEY

MCMP

seamus.bradley@lrz.uni-muenchen.de

GREGORY WHEELER

MCMP

gregory.r.wheeler@gmail.com

*Carr and Williamson's Response*

JENNIFER CARR

University of Leeds

j.carr@leeds.ac.uk

JON WILLIAMSON

University of Kent

j.williamson@kent.ac.uk

**General Description**

The theory of imprecise probabilities (IP) encompasses a family of models of uncertainty, but chief among the brood is **the theory of lower previsions** (Williams 1978; Williams 2007; Walley 1991; Troffaes and de Cooman 2014), which is a generalization of de Finetti's theory of previsions (1937, 1974). The aim of the theory of lower previsions is to model the type of uncertainty that arises from having insufficient information to identify a single probability, thereby distinguishing between what Frank Knight called decision-making under conditions of *uncertainty* and decision-making under conditions of *risk*.

The history of IP is long but controversial.<sup>38</sup> Savage (1972, Ch. 4) thought IP un-workable; de Finetti, ill-motivated (1974). Versions of these objections are heard from contemporary critics, and more. But one thing that has changed is that the mathematical foundations for a wide class of imprecise probability theories, which have been in place since (Walley 1991), recently have been greatly simplified and extended (Trofaes and de Cooman 2014). Thus, now is the time to give a critical reassessment of the theory of imprecise probabilities.

The proposed symposium will draw together two contemporary critics of imprecise probabilities, **Jennifer Carr** and **Jon Williamson**, and two defenders of imprecise probabilities, **Seamus Bradley** and **Gregory Wheeler**. There will be two 30 minute presentations (25 minute talk and 5 minute response) by Carr and Williamson, respectively, a 40 minute reply by Bradley and Wheeler, and a 20 minute discussion period.

## Abstracts

### 1. Jon Williamson: *The Epistemological Significance of Imprecise Probability*

It has been suggested that imprecise probability offers a more plausible theory of partial belief than Bayesian (precise) probability does. This is largely due to two considerations. First, because imprecise probability is apparently based on a more realistic betting model, where one is not obliged to bet for and against a proposition at the same rate as one is under the Bayesian account. Second, because it is thought that imprecise probabilities distinguish the weight of evidence that is pertinent to any particular proposition; it may not be clear whether a Bayesian probability of  $1/2$ , for instance, is based on excellent evidence of a chance process or on no evidence at all.

I argue against both these claims. First I explain that, while imprecise probability does indeed align with a more realistic betting scenario than the one which underpins Bayesian probability, the betting quotients under this account offer a less plausible explication of partial belief than do Bayesian betting quotients. Thus, what imprecise probability gains in terms of descriptive accuracy when betting comes at the expense of it providing a viable normative account of partial belief. Second, I show that the well known problem of dilation for imprecise probability puts paid to the idea that imprecise probability handles weight of evidence any better than Bayesian probability. I explain why it is a mistake to think that one's partial beliefs should be expected to capture the notion of weight of evidence and why in any case a Bayesian account of partial belief is already a part of a general epistemology that does adequately capture this notion.

I close by noting one way in which the standard Bayesian account offers a better model of partial belief than does imprecise probability. Belief is the primary basis for action. A model of partial belief should provide an account of the bearing of partial belief on rational action. The Bayesian account, which appeals to a

---

<sup>38</sup>Contemporary work on imprecise probabilities traces to ideas found in George Boole (1854) and John Maynard Keynes (1921), but especially to work by B. O. Koopman (1940), Alfreds Horn and Tarski (1948), Paul Halmos (1950), I. J. Good (1952), C. A. B. Smith (1961), Daniel Ellsberg (1961), Glenn Shafer (1976), Isaac Levi (1974, 1980), and Jay Kadane, Mark Schervish and Teddy Seidenfeld (1999). Recent critics include (White 2010; Elga 2010). See (Bradley 2014) for a review.

decision theory based on maximising expected utility, does this quite well. However, there is no single decision rule for imprecise probability - there are lots of rules, each with their own merits. Moreover, the more plausible general decision rules tend not to determine which action should be taken. Hence, imprecise probability underdetermines rational action in two key ways.

## **2. Jennifer Carr: *What Do (Im)precise Credences Represent?***

Some proponents of imprecise credences hold that precise credences are not merely psychologically unrealistic, but epistemically irrational. Faced with ambiguous or un-specific evidence, imprecise credences are rationally mandatory: precise credences are objectionably committal and evidentially arbitrary. I pose a challenge to the imprecise view. Briefly, the view faces many of the same worries that were meant to motivate it in the first place. (If a real-valued credence is arbitrary or overly committal, so is a set-valued credence.) The imprecise view faces this challenges because it incorporates a certain idealization: that imprecise credences are non-vague. But, I argue, doing away with that idealization effectively collapses the imprecise view into a particular form of precise view: one according to which ambiguous or unspecific evidence rationally requires uncertainty about evidential or objective probabilities.

Once uncertainty about evidential and objective probabilities are on the table, I argue that the imprecise view is unmotivated. The claim that precise credences are inappropriately informative or committal depends on an implausible interpretation of the what precise credences are informative about. The claim that indeterminate evidential support requires imprecise credences is also unmotivated: it tacitly depends on a hypothesis about the relation between indeterminacy and norms that is at best questionable, and clearly false in the practical case. Are there any reasons to go imprecise that don't equally support going precise with normative uncertainty? The answer, I argue, is no. Anything mushy can do, sharp can do better

## **3. Seamus Bradley and Gregory Wheeler: *Replies to Carr and Williamson***

Before considering the specific objections raised by Carr and Williamson, we will first give a brief overview of the theory of lower previsions, where we will demonstrate that this influential IP theory is an epistemic and behavioral model of rational belief that is firmly in the tradition of Ramsey, de Finetti, and Savage.

To be more specific, one way to model an agent's beliefs is to use risky transactions to identify her *sets of acceptable gambles*<sup>39</sup> that she will accept. Equivalently, one may instead model belief directly through eliciting her lower previsions, which represent her disposition to exchange risky transactions having some fixed and certain reward. The first general point we wish to make is that the framework of sets of acceptable gambles is a very general theory: several familiar models of belief, including many advanced by critics as alternatives to IP, appear as special cases, including the traditional strict Bayesian model of belief advocated

---

<sup>39</sup>Here we follow the terminology of acceptable (bounded / unbounded) gambles of (Troffaes and de Cooman 2014). Note that (Walley 1991; Walley 2000, §3.8) uses the term 'desirable gambles' instead.

by Carr. A second general point we wish to make is that, unsurprisingly, individuating properties of these special case theories of belief should not be mistaken for properties of the general theory; in particular, some forms of reasoning that are sound within particular special theories are invalid in the general theory (Pedersen and Wheeler 2014). This observation is crucial to keep in mind when evaluating alleged failings of IP in general, and the theory of lower previsions in particular.

Criticisms of IP fall into two main categories. There are epistemological or evidential criticisms and there are decision theoretic ones. We discuss both in turn.

The evidential criticisms concern IP's alleged advantage over orthodox Bayesianism with respect to representing agents with uncertain or incomplete evidence. Carr argues that there is no such advantage: precise models can do just as well. Williamson argues that the IP phenomenon of "dilation" undermines any such claim to superiority on behalf of IP. We respond that, given recent characterization results of dilation purely in terms of distance from stochastic independence (Pedersen and Wheeler 2014), several misconceptions about dilation can be put to rest.

Turning to the decision theoretic criticisms now: it is claimed that IP can't be given the same betting odds interpretation that orthodox Bayesianism can. We show that IP admits a very natural betting interpretation in terms of sets of acceptable gambles, and we argue that the standard "two-sided" bet interpretation favoured by Williamson suffers from a problem that doesn't affect IP. Williamson further argues that IP cannot always give advice as to which of two options is to be preferred. We argue that this is not a failure but a feature: one motivation for IP is to represent agents who fail to have complete preferences. Since incompleteness of preference is precisely failing to take either option to be preferred, IP may be seen as an inference mechanism to guide action in the face of such attitudes, when and if they should arise.

## References

- Boole, G. (1854). *An Investigation of the Laws of Thought*. New York: Dover.
- Bradley, S. (2014). Imprecise probabilities. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 ed.). CSLI, Stanford University.
- de Finetti, B. (1937). *La prévision: ses lois logiques, ses sources subjectives*. *Annales de l'Institut Henri Poincaré* 7(1), 1–68.
- de Finetti, B. (1974). *Theory of Probability* (1990 ed.), Volume I. New York: John Wiley.
- Elga, A. (2010, May). Subjective probabilities should be sharp. *Philosophers Imprint* 10(5).
- Ellsberg, D. (1961). Risk, ambiguity and the savage axioms. *Quarterly Journal of Economics* 75, 643–69.
- Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B* 14(1), 107–114.
- Halmos, P. R. (1950). *Measure Theory*. New York: Van Nostrand Reinhold Company.
- Horn, A. and A. Tarski (1948). Measures in boolean algebras. *Transactions of the AMS* 64(1), 467–497.
- Kadane, J. B., M. J. Schervish, and T. Seidenfeld (1999). *Rethinking the Foundations of Statistics*. Cambridge Studies in Probability, Induction, and Decision Theory. Cambridge University Press.
- Keynes, J. M. (1921). *A Treatise on Probability*. London: Macmillan.
- Koopman, B. O. (1940). The axioms and algebra of intuitive probability. *Annals of Mathematics* 41(2), 269–292.
- Levi, I. (1974). On indeterminate probabilities. *Journal of Philosophy* 71, 391–418.
- Levi, I. (1980). *The Enterprise of Knowledge*. Cambridge, MA: MIT Press.

- Pedersen, A. P. and G. Wheeler (2014). Demystifying dilation. *Erkenntnis* 79(6), 1305–1342.
- Savage, L. J. (1972). *Foundations of Statistics*. New York: Dover.
- Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton University Press.
- Smith, C. A. B. (1961). Consistency in statistical inference (with discussion). *Journal of the Royal Statistical Society* 23, 1–37.
- Troffaes, M. C. M. and G. de Cooman (2014). *Lower Previsions*. Chichester, West Sussex: Wiley and Sons.
- Walley, P. (1991). *Statistical Reasoning with Imprecise Probabilities*. London: Chapman and Hall.
- Walley, P. (2000). Towards a unified theory of imprecise probability. *International Journal of Approximate Reasoning* 24, 125–148.
- White, R. (2010). Evidential symmetry and mushy credence. In T. S. Gendler and J. Hawthorne (Eds.), *Oxford Studies in Epistemology*, Volume 3, pp. 161–186. Oxford University Press.
- Williams, P. (1978). On a new theory of epistemic probability. *The British Journal for the Philosophy of Science* 29, 375–387.
- Williams, P. (2007). Notes on conditional previsions. *International Journal of Approximate Reasoning* 44(3), 366–383.

---

*Explanation and Scientific Understanding*

SORIN BANGU

University of Bergen, Norway

sorin.bangu@gmail.com

---

Is there a relation between scientific explanation and scientific understanding? The question was probably first addressed by Hempel a good while ago in his work on the ‘covering-law’ model, and then revived by Friedman’s seminal 1974 paper ‘Explanation and Scientific Understanding’, which in turn elicited a series of reactions, both criticisms and refinements (most notably by Salmon, Kitcher and Lipton). In this paper, after I discuss some recent exchange between Trout and De Regt (see below), I propose to conceive of scientific understanding as featuring a ‘local’ and a ‘global’ component. Under this assumption, various types of theories of explanation (causal, unificationist, etc.) are assessed, in light of the fundamental idea that the main aim of scientific explanation is to yield understanding of the world.

As mentioned, the precise relation between explanation and understanding has been discussed in a recent debate between J. D. Trout and H. de Regt (see Trout 2002, 2005, 2007; De Regt 2004, 2009), who argued *con* and *pro* (respectively) the idea that understanding is important to the practice of providing scientific explanations. While acknowledging the importance of what is at stake in this exchange, in this paper I attempt to show that neither of the parties involved is right – and that their failures are both non-trivial and instructive, thus guiding us toward a potentially better understanding (of the role) of understanding in science. Thus, with regard to the central question, whether the ‘aha!’ “feeling of understanding” (or ‘FU’, de Regt’s abbreviation in his 2009, 588) truly plays a role in science, I submit that both Trout and de Regt not only misidentify their opponent’s position, and thus mostly speak past each other, but also – more importantly – fail to support their own positions.

More specifically, in my estimation the situation is rather entangled, as follows. Trout sets up initially to defend the genuinely interesting and provocative view that FU is not necessary in science, but he adduces (otherwise novel and illuminating) evidence (from psychology) to the effect that it is in fact not sufficient. So, although Trout’s initial goal remains unachieved (I contend), de Regt fails to point this out (actually, he did not seem to notice it), and instead veers the discussion somewhat orthogonally to Trout’s concerns, by building a case for the existence and significance of what he calls ‘pragmatic understanding’. As it turns out, however, this direction is actually right, I believe, and, in agreement with de Regt, I argue that understanding has a pragmatic dimension indeed – yet unfortunately his construal of this pragmatic function is misguided.

The more general upshot of my discussion is not merely a highlight of the shortcomings of these positions, but rather an attempt to take some steps toward a constructive goal, to begin the articulation of a novel conception of understanding – I will call it ‘pragmatic-motivational’, featuring a ‘local’ and a ‘global’ component) This alternative either has gone unnoticed (by Trout and others, most notably by Hempel himself) or has been mis-described (by de Regt). Along the way, I also connect the Trout-de Regt debate to related points raised by other authors, such as Grimm (2006, 2010) and Khalifa (2011, 2012).

---

*Abduction and Cultural Evolution*

ILKKA NIINILUOTO

University of Helsinki

ilkka.niiniluoto@helsinki.fi

---

Charles S. Peirce introduced in the 1860s his notion of *hypothesis* as “inference of a cause from its effect”. In the 1890s he coined the term *abduction* for such an “inference to an explanation”, where explanation can be deductive or probabilistic. This notion covers two types of cases. Singular abduction assumes a general law of the form “All Fs are G” and infers Fa from Ga. Theoretical abduction starts from surprising facts and seeks general laws and theories as their explanations.

Abduction became a hot topic in the philosophy of science after World War II when Hanson suggested that abduction is a logic of discovery and Harman argued that all types of inductive reasoning can be reduced to inference to the best explanation (IBE). At the same time Hempel and Rescher started to analyze “retroductions” which make inferences from a state of a system to its temporally previous states. But it was not recognized that Peirce had discussed such retroductions as special cases of abduction already in the 1890s.

*Retroduction* is a variant of singular abduction, where the general premise is law is a causal law of succession: if we know or assume that all Fs are followed by G, then from the event Ga we can reason backward in time to the conclusion Fa. Peirce himself illustrated retroduction by the inference from present documents to the historical existence of Napoleon Bonaparte.

This paper shows that similar – and more interesting – examples abound in biology and cultural sciences. Darwin’s theory of evolution involves abduction in three ways. First, evidence for this theory is largely abductive, based upon its explanatory power with respect to the fossil record, the geographical distribution of present species, and morphological and embryological facts. Darwin’s theory of gradual evolution by variation and natural selection gives a better explanation of these phenomena than creationism or Intelligent Design theories. Secondly, many current traits of organisms can be explained as adaptations, i.e. by the assumption that at some earlier time there was a selection or fitness advantage for them. Thirdly, on the basis of Darwinian evolution theory and palaeontology, the history of life on earth can be represented by a tree with a common descent and branches corresponding to speciation. Evolution is a process which goes forward in time, and the reconstruction of the evolutionary trees on the basis of present evidence (contemporary forms of life and fossil records of extinct species) is an abductive task in Peirce’s sense. The method for constructing such trees of is known as cladistics: similarity of species is measured by the number



of their common traits or characters, and a “cladogram” or a “dendrogram” is constructed so that more similar species have more recent ancestors.

Starting from the 1860s, the idea of evolution was applied in the study of culture (sociology, philology, ethnology, anthropology, folkloristic, epistemology, ethics). Today there are research programmes for reducing cultural evolution to biology (evolutionary psychology, sociobiology) or showing it to be analogous to biological evolution (memes as cultural genes), but the success of such attempts is limited by the fact that cultural evolution is more Lamarckian than Darwinian: variation is not random (cultural entities are intentionally produced with a goal), transmission involves inheritance of acquired characteristics (by learning and symbolic languages), and fitness is measured by cultural utilities or social values rather than by success in reproduction. In spite of these differences, evolutionism in the humanities has led to abductive inferential patterns which are similar to cladistics in biological taxonomy.

In Peirce’s example about Napoleon, the physical objects constituting the documents and monuments about this emperor have not changed over time, but in his discussion about the evolution of science Peirce considered the possibility that judgments are modified at every recall to the mind. Such gradual changes are typical in cultural evolution. For example, Schleicher assumed that, while languages adopt new words, they gradually develop toward greater structural simplicity. He concluded that the evolution of languages can be represented by family trees which grow from a basic or original form to several branches.

Textual criticism studies existing variants of written texts (e.g. manuscripts, poems) which are modified when they have been copied: hand written copies may include errors which then repeat themselves in later copies. The problem is to construct a tree or “stemma” which identifies the original text and shows the historical development of its variants. The application of this method to texts in the oral tradition (such as epic and lyric folk songs, tales, and jokes) was called “the Finnish method”, since it was developed by Julius and Kaarle Krohn, professors of folkloristic at the University of Helsinki. Such cultural items are changed when they are orally transmitted from one generation to another or from one village to another. The reconstruction of the evolutionary trees of such texts is again an abductive task in Peirce’s sense.

The inferential structure of textual criticism and stemmatology in cultural sciences is similar to cladistics in biological taxonomy. This was shown in 1991 by Robinson and O’Hara who applied a software package PAUP, developed originally for phylogenetic analysis, as a computer method in stemmatology. New effective computer methods have recently been developed by Heikkilä and Roos for finding parsimonious solutions to problems in stemmatology with large data sets.

These observations reconfirm Peirce’s insight that abduction, both as a method of discovery and justification, is an important form of reasoning in many different scientific disciplines.

---

*Causal Probability and Scientific Practice*

MARSHALL ABRAMS

University of Alabama at Birmingham

mabrams@uab.edu

---

I discuss and refine the concept of “causal probability” that Abrams introduced in 2012. Roughly, causal probabilities are objective probabilities for which frequencies reflect probabilities, and for which manipulating probabilities manipulates probabilistic patterns. Philosophers often assume that only propensities and closely related kinds of probability have this characteristic. Even if propensities are defensible, it’s doubtful that they can play all the roles required by causal probability. Moreover, other causal probability interpretations have been proposed in recent years by Rosenthal, Strevens, and Abrams, among others. Clarifying a concept of causal probability would therefore illuminate the roles of probabilities in science.

First note that many events associated with probabilities exhibit a kind of very loosely systematic pattern that Strevens calls a probabilistic pattern: As the numbers of trials of the same kind are increased, frequencies usually (but not always) stabilize near certain values. Abrams also emphasized that certain processes (roulette wheels, societies, biological populations) often generate stable frequencies, i.e. frequencies that don’t usually change in extreme ways from one period of time to another (e.g. from one year to another). Such concepts have a certain amount of inherent vagueness, but nevertheless capture intuitions about the world with which we’re all familiar, and which scientists routinely take for granted.

A causal probability exists when:

1. Objects and properties in terms of which an objective interpretation of probability is defined are realized in repeated instances of a chance setup, and produce probabilistic patterns.
2. Frequencies in these patterns and probabilities can be manipulated by manipulating the objects and properties so that outcome frequencies are usually close to outcome probabilities in large numbers of trials.

For example, a roulette wheel generates probabilistic patterns of outcomes, and these outcomes can generally be manipulated by altering the sizes or colors of wedges on the wheel. Many objective interpretations of probability will define probabilities for roulette outcomes that are usually close to the frequencies of roulette outcomes in large numbers of trials. Moreover, one can manipulate the chance setup involving a roulette wheel—for example, by manipulating wedge sizes—so as to change both probabilities of outcomes and frequencies of outcomes in the same way. These probabilities are therefore causal probabilities. Interpretations of probability that define probabilities that are always causal probabilities can be called “causal probability interpretations”.

It appears that scientific practice often depends on an implicit distinction between causal probabilities and other objective probabilities. For example, when elected officials consider changing a policy about what sort of early-childhood education to offer, they don't do so because they believe that the new policy will guarantee outcomes for particular children, or even that particular frequencies of outcomes will be guaranteed. Rather, they think that changing the policy is likely to increase the frequencies of certain desirable outcomes. On the other hand, when a social scientist calculates the mean and variance of an outcome in a population, the probabilities involved in the calculation are themselves finite frequencies. Manipulating those probabilities can't cause changes in those frequencies with which they are identical.

Some probabilities in models represent causal probabilities and some do not. Probabilities in a model may fail to represent causal probabilities because they're intended to represent non-causal probabilities in the world, or because probabilities are introduced into the model solely for the purpose of getting parts of it to behave in a specified manner, without regard to whether processes in the physical world work in a similar way. In this last case, probabilities in the model might not represent probabilities of any kind.

I'll provide a more precise characterization of "causal probability" by extending Woodward's account of manipulation to probabilistic patterns.

Assuming that one is willing to go along with the idea that there are probabilistic patterns and that they vary, we can represent such variation by values of a variable  $Y$ . However, since the notion of a probabilistic pattern is loose and informal, we should be clear that this notion of "variable" is also loose and informal. (One benefit of my more precise characterization of causal probability is that it sharpens questions about the notion of a probabilistic pattern.)

Let  $Y$  range over probabilistic patterns, and let variables  $X_1, \dots, X_m$  range over variations in properties of a chance setup consistent with a general kind of probability. (Example: Many variations in the physical structure of dice serve merely to bias outcomes of dice tosses; these variations would be represented by the variables  $X_i$ . Alterations in the dice's structure that turn them into uniformly-colored, spherical marbles would fall outside the range of variation represented by those variables; whatever probabilities might apply to tosses of marbles, they are of a different kind.) Then we can define, or at least sketch the notion of causal probability as follows:

Suppose that there is a possible intervention on  $X_1$  that would change  $Y$ . That is, suppose that there are some values of  $X_2, \dots, X_m$  for which, if there is a variable  $I$  with a value that (1) would cause  $X_1$  to take on a specific value, which in turn would cause  $Y$  to take on a particular value, in such a way that (2) the value of  $X_1$  no longer depends on the values of other variables  $X_i$ , where (3)  $I$  affects  $Y$  only via  $I$ 's effect on  $X_1$ , and (4)  $I$  is probabilistically independent of any variable that could affect  $Y$  by causal paths not passing through  $X_1$ .

Suppose, in addition, that the same intervention  $I$  "affects" probabilities via  $X_1$ 's values' roles in the chance setup, and does so for exactly the same values of  $X_2, \dots, X_m$  in which  $I$  affects  $Y$  via  $X_1$ .

Suppose, finally, that there is a fit, in sense appropriate to probabilistic patterns between the probabilities determined by the chance setup exhibiting certain values of the  $X_i$ 's, and the values of  $Y$  determined by those same values of the  $X_i$ 's. (I'll describe the appropriate sense of "fit" in my talk.)

Then the probabilities are causal probabilities.

---

*On the Limits of Causal Modeling: Spatially-Structurally Complex Phenomena*

MARIE I. KAISER

Universität zu Köln, Philosophisches Seminar

kaiser.m@uni-koeln.de

---

In recent decades major advances have been made in formalizing causation and causal inference (Spirtes et al. 2000; Pearl 2000) and in using these formalisms to address traditional philosophical issues such as scientific discovery and the nature of scientific explanations (Woodward 2003, Woodward and Hitchcock 2003). At the heart of these formal theories lie causal models that involve elements such as causal graphs, probability distributions, Bayesian nets, and structural equations which satisfy certain conditions, most prominently the Causal Markov Condition. Causal models are appreciated because they allow for inferring causal relations from observed probabilistic correlations, for predicting the effects of manipulations and interventions, and because they can be used for representing and explaining causal relationships in very general, formal terms.

Proponents of the causal modeling approach usually emphasize and exemplify the wide scope of their approach. In recent years several authors have, for instance, shown that the causal modeling approach can also be applied to mechanistic explanations in biology and medicine (Casini et al. 2011; Clarke et al. 2014; Gebharder and Kaiser 2014). Also Woodward's interventionist theory of causation and causal explanation that makes extensive use of causal graphs is supposed to be applicable to a very wide range of causal relationships, including those in the biological sciences (2010, 2011, 2013).

I agree that causal modeling is central to scientific practice and that it is a powerful approach to formally represent, explain, and discover causal relations. However, I also think that its scope should not be overestimated and that it is important to recognize also the limits of the causal modeling approach. In this paper I use an example from molecular biology to reveal one of these limits: the explanation of spatially and structurally complex biological phenomena. According to my line of criticism, formal causal models fail to offer adequate causal explanations of biological phenomena that essentially involve complex spatial and chemical-structural relations. This failure is due to the fact that causal graphs only provide causal difference-making information of the sort: A change in the value of X would under suitable conditions change the value (or probability distribution) of Y. The explanations of some biological phenomena, however, seem to be richer than this: these explanations do not only represent causal relations but also and in particular spatial relations and biochemical structures (such as the conformation and chemical structure of macromolecules, the spatial orientation and fitting of macromolecules to each other, and the complementarity of chemical structures). Based on the analysis of a case study from molecular biology, DNA recognition and binding by gene regulatory

proteins (cf. Pavletich and Pabo 1991; Luisi 1991; Somers and Phillips 1992; Klemm et al. 1994), I show that the formal tools of causal graph theory are too impoverished to model biological processes that involve complex spatial-structural relations.

Interestingly, Woodward (2011) has basically conceded this point, but he does not see this as a limitation of his causal modeling or interventionist approach to scientific explanation. Taking a relaxed stance, he argues that complex spatiotemporal information can just be added back to the backbone of causal difference-making information and can be used to “organize” (2011, 423) or “fine-tune” (2013, 55) causal difference-making information. This paper shows that things are not that easy. Some biological processes involve complex spatial and chemical-structural relations that are central to explaining these processes but that cannot be adequately represented in causal graph models. My central argument will be that the formal tools of causal graph theory are inappropriate to model and to explain spatially and structurally complex biological phenomena because they result in causal models which either ignore the importance of spatial and structural relations altogether or which try to include the relevant spatial and structural information but, in so doing, render the causal graph models non-explanatory, unmanageable, or inadequate because they conflict with basic assumptions of causal graph theory. This argument does not erode the significance of formal approaches to causal modeling, but it demonstrates that their scope is limited.

## References

- Casini, L., P. M. Illary, F. Russo, and J. Williamson 2011. "Models for Prediction, Explanation and Control: Recursive Bayesian Networks." *Theoria. An International Journal for Theory, History and Foundations of Science* 70(1): 5-33.
- Clarke, B., B. Leuridan, and J. Williamson 2014. "Modelling Mechanisms with Causal Cycles." *Synthese* 191(8): 1-31.
- Gebharder, A. and M. I. Kaiser 2014. "Causal Graphs and Biological Mechanisms." In *Explanation in the Special Sciences. The Case of Biology and History* by M. I. Kaiser, O. Scholz, D. Plenge and A. Hüttemann. Berlin: Springer, 55-85.
- Klemm, Juli D., Mark A. Rould, Rajeev Aurora, Winship Herr, and Carl O. Pabo 1994. "Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding modules". *Cell* 77: 21-32.
- Luisi, B.F., W. X. Xu, Z. Otwinowski, L. P. Freedman, K. R. Yamamoto, P. B. Sigler 1991. "Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA". *Nature* 352:497-505.
- Pavletich, Nikola P. and Carl O. Pabo 1991. "Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å". *Science, New Series* 252:809-817.
- Pearl, J. 2000. *Causality. Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
- Somers, W. S. and S. E. Phillips 1992. "Crystal structure of the met repressor-operator complex at 2.8 Å resolution reveals DNA recognition by beta-strands". *Nature* 359:387-393.
- Spirtes, P., Glymour, C. and R. Scheines. 2000. *Causation, Prediction, and Search*. Cambridge, Mass.: MIT Press.
- Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.
- Woodward, J. 2010. "Causation in biology: stability, specificity, and the choice of levels of explanation." *Biology & Philosophy* 25: 287-318.
- Woodward, J. 2011. "Mechanisms Revisited." *Synthese* 183(3): 409-427.
- Woodward, J. 2013. "Mechanistic Explanation: Its Scope and Limits." *Proceedings of the Aristotelian Society*

Supplementary Volume 87(1): 39-65.

Woodward, J. and C. Hitchcock 2003. "Explanatory Generalizations, Part I: A Counterfactual Account." *Noûs* 37(1): 1-24.

---

*Einstein's Physical Strategy, Energy Conservation, Symmetries and Stability*

J. BRIAN PITTS

University of Cambridge

jbp25@cam.ac.uk

---

While Einstein's developed General Relativity ahead of his time, the development of relativistic quantum mechanics and quantum field theory involved a more systematic exploration by many people over several decades: particle physics. It turns out to have a massive, largely positive impact on the manifest rationality of gravitational physics *vis-a-vis* the mystical aura of Einstein's work on General Relativity noticed by Norton and Renn. From the early 1970s until the 2000s, particle physics seemed to provide an eliminative induction for General Relativity. From the 1920s to 1970 and again in the 2000s, particle physics suggests what the most interesting competition for General Relativity might be, in terms of prior plausibility, fit to data, and philosophical payoff.

One major consequence of the work on quantum mechanics in the 1920s-30s was the completion of the exploration of (classical) *Special* Relativity by finding all possible wave equations covariant under the Poincaré group. The result was Wigner's taxonomy in terms of 'mass' (a new fundamental inverse length scale for a particle/field) and 'spin' (intrinsic angular momentum, related to the scalar vs. vector vs. matrix form of the potential). Spin 1/2 particles/fields and the Dirac equation are the most spectacular novelty. Systematic comparison of massless particles/fields (for which waves travel at  $c$ , the 'speed of light' and point sources produce a  $1/r$  solution) and massive particles/fields (for which wave speed varies with frequency and point sources produce a solution  $\exp(-mr)/r$ ) is comparably important. Massless particles/fields tend to have larger symmetry groups, sometimes even infinite-parameter gauge symmetries.

Thus one could situate Einstein's theory within the full range of relativistic wave equations and Lorentz group representations: classical field theories in disguise. In 1939 Pauli and Fierz recognized the equation for a massless spin 2 field as the source-free linearized Einstein equations. Rosen wondered about deriving General Relativity's nonlinearities from flat space-time. Kraichnan, Gupta, Feynman, Weinberg, Deser *et al.* eventually filled in the gaps, showing that, on pain of instability due to negative energies, Einstein's theory (massless spin 2) is basically the only option except perhaps a massive spin 2, if even that: an argument style that philosophers of science recognize as eliminative induction. One can show that the particle physics derivations of Einstein's equations work by a sort of converse Noether theorem from 'improper' conservation laws to gauge symmetries. Recognizing General Relativity as a massless spin 2 theory provided a physical



meaning for the earlier proposed modification for Newtonian gravity by Seeliger and Neumann (1890s) and Einstein (1917 cosmological constant paper) in relation to the divergent Newtonian gravitational potential: they had proposed in effect a non-relativistic massive spin-0 gravity theory. By diagnosing many possible theories (including Einstein's 1913-15 Entwurf theory, I observe) as non-viable due to negative-energy instability, while suggesting massive counterparts to massless particle/field theories, particle physics channels theorizing into directions likely to be viable. Massive spin 2 gravity continued to be developed by, among others, Marie-Antoinette Tonnelat (encouraged by de Broglie, inventor of the massive photon in his youth) and Ogievetsky & Polubarinov (who pre-invented in the 1960s theories recently discovered in 2010 with a key feature).

In the early 1970s massive spin 2 gravity apparently foundered on a dilemma: either the theory has a spin 0 partner with negative energy ("ghost") and so is unstable, or it is pure spin 2 and has a discontinuous massless limit (van Dam-Veltman-Zakharov discontinuity) bringing empirical falsification. An auxiliary argument concluded that a linear spin 2 theory becomes a spin 2-spin 0 theory nonlinearly ("Boulware-Deser ghost"), so a full nonlinear (interacting) spin 2 theory isn't even possible. Thus General Relativity had a compelling basis in particle physics --a more compelling basis than the more famous motivation by Einstein's Principles (equivalence, general relativity, general covariance, Mach). It has been partly shaken, however, since 1999 by the doubts that dark energy cast on the long-range behavior of General Relativity. Progress in achieving a smooth massless limit nonperturbatively (the Vainshtein mechanism) in the 2000s and the avoidance of the spin 0 ghost nonlinearly in 2010-11 have made massive spin 2 gravity a very active topic of research in current physics.

Recent work on the history of General Relativity by Renn, Janssen *et al.* shows that Einstein found his field equations partly by a physical strategy including the Newtonian limit, the electromagnetic analogy, and energy conservation. Such themes are similar to those later used by particle physicists, I note. How do Einstein's physical strategy and the particle physics derivations compare? What energy-momentum complex(es) did he use and why? Given that Lagrange and Jacobi linked symmetries and conservation, did Einstein? If so, to which? How did his work relate to emerging knowledge (1911-14) of the canonical energy-momentum tensor and its translation-induced conservation in Herglotz, Mie and Born? After initially using energy-momentum tensors hand-crafted from the gravitational field equations, Einstein used an identity from his assumed linear coordinate covariance  $x^\mu = A^\mu_\nu x^\nu$  to relate them to the canonical tensor. Whereas Mie and Born were concerned about the canonical tensor's asymmetry, Einstein did not need to worry because his Entwurf Lagrangian is modelled not so much on Maxwell's theory (which avoids negative-energies at the cost of asymmetric canonical energy-momentum) as on a scalar theory (the Newtonian limit) with a symmetric canonical tensor. But as a result, Einstein's Entwurf theory has 3 negative-energy field degrees of freedom (ghosts). Thus it fails a 1920s-30s *a priori* particle physics stability test with antecedents in Lagrange's and Dirichlet's stability work. This critique of the Entwurf theory is much easier

and more compelling than Einstein's 1915 critique of Entwurf for not admitting rotating coordinates and not getting Mercury's perihelion right.

Particle physics thus can be useful in the study of gravity both in assessing the growth of objective knowledge and in suggesting novel lines of inquiry to see whether and how Einstein faced similar issues.

---

*Renormalization and Relativity*

JAMES FRASER

University of Leeds

jamesf09@hotmail.co.uk

---

There has recently been heated debate amongst philosophers of physics about which formulation of quantum field theory (QFT) ought to be subjected to philosophical interpretation. Doreen Fraser (2009, 2011) has argued that the approach to QFT found in physics textbooks is an inadequate basis for foundational study and instead advocates the philosophical significance of axiomatic formulations of the theory, such as the algebraic approach originating in Haag and Kaiser (1964). David Wallace (2006, 2011) on the other hand defends the legitimacy of the physicist formulation of the theory and argues that philosophers who are interested in what we should believe about the world given the successes of high energy physics ought to be engaging with this version of QFT.

There are, I think, many factors underlying this dispute between Fraser and Wallace, but in this paper I focus on one important issue over which they clearly diverge, namely the status of relativity in QFT. Wallace adopts an understanding of the physicist's formulation of QFT in which the process of regularising a model, for instance by imposing a cutoff on the energy, is interpreted as explicitly throwing away information about high energy degrees of freedom. On this view the standard model of particle physics is a so called effective field theory which describes low energy degrees of freedom below the cutoff but says nothing about the physics at higher energies where quantum gravity effects are expected to become dominant. Fraser and Wallace both point out that imposing a high energy cutoff amounts to putting a lattice on space-time, which violates Poincaré covariance. For Wallace compromising relativity in this way is consistent with the effective field theory approach. Relativistic symmetries will, according to Wallace, remain approximate symmetries of the low energy degrees of freedom described by the standard model but as an effective field theory it makes no claims about the nature of space-time at arbitrarily small length scales. By contrast, Fraser takes QFT to be a unification of quantum and relativistic physics which therefore incorporates a commitment to Minkowski space-time structure. In the context of this unificatory project imposing a cutoff to make the theory well defined is, according to Fraser, an ad hoc manoeuvre.

In the first part of this paper I assess the plausibility of Wallace's response to the apparent clash with relativity engendered by the effective field theory approach. Putting aside the technical question of whether we can make sense of regularisation in a Poincaré covariant way the key issue is how we should understand the notion of approximate or effective Poincaré symmetry once a cutoff has been imposed. I suggest that there is scope for more debate about how the notion of approximate symmetry ought to be explicated, but

claim that independently of these more general issues, proponents of the effective field theory approach can appeal to renormalization group arguments to the effect that the low energy phenomenology of a cutoff model will differ only negligibly from a Poincaré covariant completion of its high energy degrees of freedom. Indeed, Fraser grants this point when she characterises the choice between axiomatic and cutoff formulations of QFT as a case of underdetermination of theory by evidence. This result provides a minimal notion of effective Poincaré covariance that is, I suggest, sufficient for Wallace's purposes.

Fraser's objection to imposing a cutoff is ultimately an extra-empirical one then. Cutoff models can reproduce relativistic empirical results; the problem is that they are not fundamentally Poincaré covariant. There are parallels here with a common complaint raised against hidden variable reformulations of quantum mechanics. Bohmian theories might be able to recover relativistic space-time symmetries at the phenomenological level, critics say, but they have given up on fundamental relativity, in this case by adding a preferred foliation to space-time. This kind of argument is only as strong as the reasons for demanding fundamental Poincaré covariance in the first place however.

In the latter half of this paper I discuss whether there are legitimate reasons for making this demand in the context of the debate over the formulation of QFT. It is useful here to distinguish internal questions about a theoretical framework from external questions about how theories relate to the world. One can ask, for instance, whether a consistent Newtonian cosmological theory can be formulated independently of whether such a theory has anything to do with the physics of the actual world. Similarly, there are excellent reasons, from both a philosophical and first order physics perspective, to ask whether quantum theories can be consistently combined with fundamental Poincaré covariance. But this project is distinct from the question of what we ought to believe about the world given the successes of high energy physics. It seems to me that both of these questions are of legitimate philosophical interest and that it may be that different theoretical approaches to QFT are the appropriate frameworks for answering each.

In particular, it is not obvious why fundamental Poincaré covariance should be demanded if we are interested in what the standard model tells us about the world, since there is no known formulations of the standard model which satisfies this requirement yet there are explicit regularised models that are effectively Poincaré covariant at low energies in the sense defended in the first part of the paper. One reason that might be put forward for requiring fundamental Poincaré covariance in this context is that special relativistic space-time structure plays an indispensable role in explaining phenomena like time dilation. Some philosophers do seem to talk in this way but I point out that this view has some problematic consequences. First it seems to struggle with possibilities raised by some approaches to quantum gravity: that space-time is discrete or ultimately a non-fundamental emergent structure for instance. Furthermore, we have already said that cutoff models are capable of reproducing relativistic empirical results so the sense in which they are supposed to be explanatorily deficient is at best unclear. I conclude that further arguments are needed to use relativity to rule out realist readings of effective field theory formulations of the standard model.

---

*What Explains the Spin-Statistics Connection?*

JONATHAN BAIN  
New York University  
jbain@duke.poly.edu

---

The spin-statistics connection (SSC) plays an essential role in explanations of a wide range of non-relativistic quantum phenomena such as the electronic structure of solids and the behavior of Einstein-Bose condensates, superconductors, and white dwarf stars, among other things. However, it is only derivable in the context of relativistic quantum field theories (RQFTs) in the form of the Spin-Statistics theorem; and there are mutually incompatible ways of deriving it. This essay considers the sense in which SSC is an essential property in RQFTs, and how it is that an essential property in one type of theory can appear in fundamental explanations offered by other, inherently distinct theories.

The first part of the essay argues that an explanation of SSC based on an appeal to the Spin-Statistics theorem is problematic because (a) proofs of the theorem can be formulated in different conceptually distinct ways that disagree on the principles essential to the derivation; (b) the theorem does not hold for realistic interacting RQFTs; and (c) the theorem does not hold for realistic interacting non-relativistic quantum theories. Given that the majority of evidence for SSC comes from physical systems best described by the latter type of theories, an appeal to the Spin-Statistics theorem alone fails to provide a complete understanding of SSC. I drive this point home by arguing that such an appeal cannot be viewed as providing any of the standard types of explanation discussed in the philosophy of science literature: deductive-nomological, unifying, causal, or structural.

These considerations suggest that a full account of SSC must go beyond the Spin-Statistics theorem. A full account should explain by virtue of both a derivation from non-fundamental principles in RQFTs, and an explication of intertheoretic relations between RQFTs on the one hand, and non-relativistic quantum theories on the other. The second part of the essay characterizes this type of explanation and compares it with a similar account given by Weatherall (2011). In Weatherall's example, a general observational feature of the world (the equality of the gravitational and inertial mass of any body) that is expressed in one theory, Newtonian gravity, is explained by appealing to another, presumably more fundamental theory, general relativity (GR), in which the explanandum cannot be expressed. The explanatory work is done by means of a translation between GR and Newtonian gravity which demonstrates how the explanandum arises in a limiting process that goes from GR to Newtonian gravity. In the present essay, a similar explanation can be constructed for SSC in realistic interacting non-relativistic quantum theories. This explanation demonstrates how SSC arises in a limiting process that goes from presumably more fundamental realistic interacting RQFTs

(in which SSC cannot be expressed), to non-interacting RQFTs (in which SSC can be expressed and derived), and thence to realistic interacting non-relativistic quantum theories (in which SSC appears as a brute empirical fact). On the other hand, I argue that the presence of SSC in realistic interacting RQFTs cannot be explained by even this type of explanation, and that currently, a problem at the heart of foundational issues associated with RQFTs (the "Existence Problem") precludes a full understanding of SSC in this context.

**Reference**

Weatherall, J. (2011) 'On (Some) Explanations in Physics', *Philosophy of Science* 78, 421–447.

---

*Macroscopic Oil Droplets Mimicking Quantum Behavior: How Far can we Push an Analogy?*

LOUIS VERVOORT

University of Quebec at Montreal

louisvervoort@hotmail.com

Yves Gingras

University of Quebec at Montreal

gingras.yves@uqam.ca

---

**Abstract.** We describe here a series of experimental analogies between fluid mechanics and quantum mechanics recently discovered by a team of physicists. We argue that these experimental facts put ancient theoretical work by Madelung on the analogy between fluid and quantum mechanics into new light. We place these analogies in their historic and philosophical context, relating them to the de Broglie-Bohm interpretation of quantum mechanics. Finally we point out a distinctive advantage of the ‘fluid-mechanical’ interpretation of quantum mechanics over the Bohm interpretation: Madelung’s interpretation may rid Bohm’s theory of its strongly non-local character.

### 1. Introduction

Historically analogies have played an important role in understanding or deriving new scientific results. They are generally employed to make a new phenomenon easier to understand by comparing it to a better known one. Since the beginning of modern science different kinds of analogies have been used in physics as well as in natural history and biology (for a recent historical study, see Gingras and Guay 2011). With the growing mathematization of physics, mathematical (or formal) analogies have become more frequent as a tool for understanding new phenomena; but also for proposing new interpretations and theories for such new discoveries. Einstein, for example, used formal analogies in several of his papers to reveal the corpuscular nature of light (Gingras 2005, Norton 2006) and the wave-particle duality (Gingras 2011). Also, there is an important philosophical literature devoted to discussing the general validity of analogical inferences (cf. e.g. Hesse 1966, Bartha 2010, Norton 2014). One general conclusion of the latter works is that although analogies between phenomena are rarely perfect, this mode of inference has, to the least, an essential heuristic value.

In this article we analyze a striking case of experimental analogies, that may shed new light on the ancient problem of the interpretation of quantum mechanics, in particular the wave-particle duality. Over the last ten years, a group of French physicists led by Yves Couder has shown, through a series of original experiments, that many properties typical of quantum systems can also be observed in classical systems. The team uses a fluid-dynamical experimental system essentially composed of a thin film of fluid (a special oil), made to vertically vibrate, on which oil droplets are deposited; the dynamics of the system is such that under specific conditions such droplets can horizontally ‘walk’ over the oil surface for indefinite time. The Paris group showed in particular that walking droplets can exhibit double-slit interference, quantization of angular

momentum, and the analogue of tunneling and Zeeman splitting (Couder et al. 2005, 2006, Fort et al. 2010, Eddi et al. 2011, 2012). Other researchers have already confirmed and extended these results (Molacek and Bush 2013a, 2013b). These analogies are striking because macroscopic fluid mechanics and microscopic quantum mechanics are usually thought to be quite disjoint. At the same time they suggest that, contrary to what is generally believed, an intuitive understanding of quantum mechanics is maybe not beyond reach. These *experimental* analogies also point to the possibility that *formal* analogies between hydrodynamics and quantum mechanics could exist and be further revealed.

The first objective of this article is to present these experiments in such a way as to make apparent the foundational issues they raise. The description of the most relevant experimental results, revealing the analogies, is given in Section 2. In Section 3 we will recall that a formal (mathematical) analogy between fluid and quantum mechanics had already been proposed by the German physicist Erwin Madelung, right at the birth of quantum mechanics (Madelung 1927). Since this theoretical result is not well known, and since we believe it gains new import in the context of the Paris experiments, we will re-derive it in Section 3. We will do so in a somewhat more detailed manner than in (Madelung 1927), paying attention to all hypotheses made. In Section 4 we will show that Madelung's 'fluid-mechanical' interpretation of quantum mechanics is linked to the better-known interpretation proposed by David Bohm (1952); and that an upgraded version of the latter (presented in Bohm and Vigier 1954) allows to make a connection with the droplet-experiments. It will appear very tempting, again in view of the intriguing analogical results of the Paris group, to identify some advantages of the Madelung interpretation over Bohm's interpretation, which we will do in Section 4.

By suggesting relations between different processes or object domains, analogies have the potential to contribute to the unification of what appear as radically distinct phenomena. Based on the experimental and formal analogies presented in Sections 2 and 3, one may infer, as a kind of maximal 'working hypothesis' that *quantum mechanics might, in the end, be nothing else than a fluid-dynamical theory*<sup>40</sup>. The latter inference from the experimental results of Couder et al. and the theoretical results of Madelung could be called the 'maximal induction' – it brings the inference to its epistemic focal point. Whether one finds this maximal induction a convincing hypothesis is, at present, largely a matter of taste. Clearly, at this point many researchers will not adhere to it.

## 5. Conclusion

The goal of this article was to present the analogies between fluid and quantum mechanics recently revealed by experiments by a French research team; to show that they can be linked to Madelung's theoretical analogy

---

<sup>40</sup> That this inference is *not* disproven by Bell's theorem, the most stringent mathematical argument that might a priori prohibit it, is argued in (Vervoort 2014).



between quantum and fluid mechanics; and to advance a few arguments for why the fluid-mechanical interpretation of quantum mechanics deserves renewed interest. First we described the analogies in sufficient detail so as to make them suitable for a foundational analysis. Then we derived Madelung's result starting from the Navier-Stokes equation and precisely stated all assumptions under which the Navier-Stokes equation and the continuity equation transform into the Schrödinger equation. We recalled that a strong link exists between the Madelung or fluid-mechanical interpretation of quantum mechanics and the Bohm interpretation (both frameworks start from the same basic equations (10b) and (13)). We emphasized that Bohm and Vigier (1954) integrated concepts from Madelung into their interpretation, leading them to see quantum particles as *singularities in a Madelung fluid*. We argued that it is this hypothesis that allows to connect the droplet-experiments to the Madelung–Bohm theory. Finally, we argued that the Madelung interpretation of the 1-particle Schrödinger equation may represent an essential improvement over the Bohm interpretation: it may rid the latter of its nonlocal character.

Thus we hope that the remarkable experiments by Couder et al., showing that local (fluid) systems can mimic quantum systems, will give a new impetus to the search for a local theory for quantum mechanics – i.e. a theory that is not in blatant contradiction with relativity theory.

### References.

- BARTHA, P. (2010), *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*, Oxford University Press, Oxford.
- BATCHELOR, K. (2000), *An Introduction to Fluid Mechanics*, Cambridge University Press, Cambridge
- BELL J. S. (1964), *Physics* 1, 195-200.
- BERRY M.V., et al. (1980), 'Wavefront dislocations In the Aharonov-Bohm effect and its water wave analogue', *Eur. J. Phys.* 1:154–162.
- BOHM D. (1952), 'A suggested interpretation of the quantum theory in terms of "hidden" variables. Part I', *Phys. Rev.* 85, 166
- BOHM D., VIGIER, J.P. (1954), 'Model of the causal interpretation of quantum theory in terms of a fluid with irregular fluctuations', *Phys. Rev.* 96, 208.
- BRADY, R., and ANDERSON, R. (2014), arXiv preprint <http://arxiv.org/abs/1401.4356>
- COUDER Y., S. PROTIÈRE, E. FORT, and A. BOUDAOU (2005), 'Dynamical phenomena: Walking and orbiting droplets', *Nature* 437 (7056), 208-208.
- COUDER Y., and E. FORT (2006), 'Single-particle diffraction and interference at a macroscopic scale', *Phys. Rev. Lett.* 97(15), 154101.
- EDDI A., E. SULTAN, J. MOUKHTAR, E. FORT, M. ROSSI, and Y. COUDER (2011), 'Information stored in faraday waves: the origin of a path memory', *Journal of Fluid Mechanics* 674, 433.
- EDDI A., J. MOUKHTAR, S. PERRARD, E. FORT, and Y. COUDER (2012), 'Level-Splitting at a macroscopic scale', *Phys. Rev. Lett.* 108, 264503
- FORT E., A. EDDI, J. MOUKHTAR, A. BOUDAOU, and Y. COUDER (2010), 'Path-memory induced quantization of classical orbits', *Proc. Natl. Acad. Sci.*, 107, 17515.
- GINGRAS, Y. (2005), 'How the photon emerged through the prism of formal analogies', *Photons*, 3, 13.
- GINGRAS, Y. (2011), 'La valeur inductive des analogies : comment Einstein a vu la lumière à travers le prisme des analogies formelles', dans Hugues Chabot, Sophie Roux, *La mathématisation comme problème*, pp. 88-108, Éditions des archives contemporaines, Paris.
- GINGRAS, Y. and GUAY, A. (2011), 'The uses of analogies in Seventeenth and Eighteenth Century Science', *Perspectives on Science*, vol. 19, no 2, 154-191.

- HESSE, M. (1966), *Models and Analogies in Science*, University of Notre Dame Press, Notre Dame.
- HOLLAND, P. (1993), *The Quantum Theory of Motion: An Account of the De Broglie-Bohm Causal Interpretation of Quantum Mechanics*, Cambridge University Press, Cambridge
- KOCSIS S., B. BRAVERMAN, S. RAVETS, M.J. STEVENS, R.P. MIRIN, L.K. SHALM, and A.M. STEINBERG (2011), *Science* 332, 1170
- KUNDU, P. and COHEN, I. (2008), *Fluid Mechanics*, Elsevier, Amsterdam
- MADELUNG, E. (1927), 'Quantentheorie in hydrodynamischer Form'. *Z. Phys.* 40 (3–4): 322–326
- MOLACEK, J. & BUSH, J. W. M. (2013a), 'Drops bouncing on a vibrating bath', *J. Fluid Mech.* 727, 582–611.
- MOLACEK, J. & BUSH, J. W. M. (2013b), 'Drops walking on a vibrating bath: towards a hydrodynamic pilot-wave theory', *J. Fluid Mech.* 727, 612–647.
- NORTON, J. (2006), 'Atoms, entropy, quanta: Einstein's miraculous argument of 1905', *Studies in Hist. Phil. Modern Phys.* 37, 71-100.
- NORTON, J. (2014), 'A material dissolution of the problem of induction', *Synthese* 191:671–690.
- SBITNEV, V. (2014), arXiv preprint <http://arxiv.org/abs/1403.3900> TSEKOV, R. (2009), *Int. J. Theor. Phys.* 48, 630
- SANZ, A. S. et al. (2002), *J. Phys.: Condens. Matter* 14, 6109
- VERVOORT, L. (2014), arXiv preprint, <http://arxiv-web3.library.cornell.edu/abs/1406.0901>
- WILHELM, H. (1970), 'Hydrodynamic model of quantum mechanics', *Phys. Rev. D* 1, 2278
- WYATT, R. E. (2002), *J. Chem. Phys.* 117, 9569

---

*Species Concepts as Tools*

JUSTIN BZOVY

Western

jbzovy@uwo.ca

---

I sharpen modern accounts of species pluralism by exploring the ‘species concepts as tools’ metaphor on the basis of a case study in yeast systematics. A pluralist may rule out certain species concepts as unsuitable tools for systematic work by the lights of biological theory, but my aim is to be more discerning. Pluralists understand that different concepts work well for different areas of biology, but what this remains under-explored. Rather than ask on what grounds a concept is legitimate by the light of biological theory, one ought to ask how concepts are applied. Philosophers need to develop a notion of how a species concept can be used well or misused. For example, the biological species concept (BSC) would be used appropriately if it were applied to sexual organisms, would not be used at all with respect to asexual organisms, but how might it be used for groups of organisms that straddle the divide between sexual and asexual organisms? Application questions remain unanswered by current versions of species pluralism.

Notably, pluralism implies a one-to-one correspondence between successful applications of a species concept and a species, but the way concepts are applied suggests otherwise. Identifying new species involves a diverse array of data, molecular, biogeographical, morphological, etc., (e.g., Lachance & Fedor, 2014). My work shows how multiple species concepts are used together as *tools* to extract information from such data. Given this sort of research, I develop a rationale to determine whether or not a species concept is being used well or not. In order to elucidate the species as tools metaphor, I provide a case study on how concepts are used by yeast taxonomists.

Identifying new species of yeast is a complicated procedure. Though species concepts play a role, it isn’t through a straightforward application of a concept that a species is identified. In yeast systematics multiple concepts are used in particular situations, because not all are separately adequate. In some cases the BSC is not relevant because there are asexual forms of yeast (Freitas *et al.*, 2013). But the BSC can be relevant because, though strains of yeast are often haploid (one set of chromosomes), they can be heterothallic (have sexes that reside in different individuals), so species ought to be delineated on the basis of reproductive discontinuity (Lachance & Fedor, 2014, 541). But in practice, species assignment can be based on the BSC *and* DNA sequencing (Lachance & Fedor, 2014, 542). Often mating success is determined by the sort of spores produced, which can be interpreted as a morphological concept tailored to a particular group, but also as a

phylogenetic species concept (PSC) based on synapomorphies (shared derived traits). With yeast there can also be quite a bit of interspecific mating that can lead to introgression, which can obscure how we interpret genetic information. An ecological species concept (ESC) is also relevant, because what sorts of flowers or beetles, for example, a strain is commonly found on is important.

Yeast systematics provides many examples where multiple species concepts are used to identify one species. This is important because species pluralists assume that each species concept, when appropriately applied, will yield only one species. That is, that there is still a one-to-one correspondence between successful applications of species concepts and species. I argue that yeast systematics shows that there can be a many-to-one correspondence between species concepts and species, and that this doesn't imply species monism. Generalizing from this case, I explain how the 'species as tools' metaphor can be used to develop a rationale for determining when a concept is being used appropriately within a pluralistic framework.

One way of interpreting the 'species as tools' metaphor imports Waters' (2011) account of 'Toolbox Theorizing.' Waters develops this in the context of evolutionary theory with regard to different formulations of the Price Equation, a covariance equation that determines the change in allele frequency of a population. According to Waters, different versions of the Price equation are needed, and neither is fundamentally correct. They are to be thought of as different tools in a box. Depending on the question, a particular version will be appropriate. If biologists want to know what would happen if selection were eliminated by changing the environment, one version will suffice. If biologists want to know what would happen if transmission bias were increased, another version will suffice. Both versions provide informative, partial causal decompositions. Applying Waters strategy to species concepts is not a straightforward procedure, and I consider and respond to the following two problems.

One problem concerns the notion of a 'parsing of causes.' There are three reasons why this is inadequate. First, taxonomists distinguish between different stages of taxonomy:  $\alpha$ ,  $\beta$ , and  $\gamma$ -taxonomy. The first stage,  $\alpha$ -taxonomy, does not involve parsing any causes (though  $\beta$  and  $\gamma$ -taxonomy do), but requires the use of a species concept (Mayr, 1969, 15). Second, non-causal species concepts can play an important role in yeast systematics, especially when used to corroborate the hypotheses of other species concepts. Third, adopting Dupré's (1993) species pluralism, which denies a fundamental position to evolutionary theory, allows for many non-causal species concepts.

A second problem comes from a general criticism of species pluralism. Given pluralism, it does not seem correct to say, when species concepts yield equi-inclusive groups, that we *still* have two species. As my case study shows, this is not how species concepts are applied. Further, Waters' 'Toolbox Theorizing' does not even account for this because there is a one-to-one correspondence between different formulations of the Price Equation and different causal parsings.

I conclude by considering the following worry about my approach. One reason multiple species concepts may be used in an area of biology such as this is pragmatic. We want to have a stable classification system.

Methods of identifying new species change, and one must be careful when claiming to have discovered a new species. If a researcher has more than one way of justifying their identification of a new species, then their justification will be more likely to remain stable.

## References

- Dupré, John. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Harvard University Press.
- Freitas, Larissa F. D., Barriga, Enrique Javier Carvajal, Barahona, Patricia Portero, Lachance, Marc-André, & Rosa, Carlos A. 2013. *Kodamaea transpacific* f.a., sp. nov., a yeast species isolated from ephemeral flowers and insects in the Galapagos Islands and Malaysia: further evidence for ancient human transpacific contacts. *International Journal of Systematic and Evolutionary Microbiology*, **63**, 4324–4329.
- Lachance, Marc-André, & Fedor, Alexandra N. 2014. Catching speciation in the act: *Metschnikowia bowlesiae* sp. nov., a yeast species found in nitidulid beetles of Hawaii and Belize. *Antonie van Leeuwenhoek*, **105**, 541–550.
- Mayr, Ernst. 1969. *Principles of Systematic Zoology*. McGraw-Hill.
- Waters, C. Kenneth. 2011. Okasha's Unintended Argument for Toolbox Theorizing. *Philosophy and Phenomeno- logical Research*, **82**(1), 232–240.

---

*Squaring the Circle? Assessing Mechanistic Constitution With Interventions*

LENA KÄSTNER

Mind &amp; Brain School Berlin

mail@lenakaestner.de

BEATE KRICKEL

RUB Bochum

beate.krickel@gmx.de

---

**1. Introduction**

Proponents of mechanistic explanations suggest there to be a *constitutive relevance relation* between the phenomenon to be explained and the components of its implementing mechanism (Bechtel 2008; Bechtel & Abrahamsen 2005; Craver 2007; Craver & Darden 2014; Machamer, Darden, Craver 2000; Illari & Williamson 2011). According to the standard view, this constitutive relation consists in a *part---whole relation* and *mutual manipulability* between the mechanism's components and the phenomenon as a whole (Craver 2007) which can be assessed by means of *interventions* (Woodward 2003, 2011).

While this view captures certain aspects of scientific practice, it creates a *conceptual problem*: Woodward's interventions are designed for detecting *causal* relations while constitutive relevance is explicitly described as *non-causal*. So can we use interventions at all to assess constitutive relations? Even if this can be done, an *empirical challenge* remains: How can we experimentally distinguish between causal and constitutive relations?

Recently, Gebharder & Baumgartner (2015) have suggested a solution to the conceptual problem. However, their treatment of the empirical challenge remains unsatisfying. In this paper, we suggest a different way to meet the empirical challenge that is based on different possible solutions to the conceptual problem.

**2. The Empirical Challenge**

We take the empirical challenge to be the following: Suppose you intervene into a variable P and detect a change in some other variable C; what can you infer from that? There are three straightforward interpretations:

- (1) *Causation*: P is a cause of C
- (2) *Common Cause*: The intervention is a common cause of both P and C
- (3) *Supervenience/Constitutive Relevance*: P supervenes on/is constituted by C (such that C changes in virtue of the changes our intervention induces in P)

How can we empirically distinguish between these three interpretations? 3 Excluding *Causation*

### 3. Excluding Causation

To allow interventionism to capture the difference between causation and constitution, Baumgartner & Gebharter (2013) introduce time into the interventionist definition of a cause: if we are dealing with causation, changes in the cause variable have to occur prior to changes in the effect variable. In contrast, *Common Cause* allows for changes at the same time, while *Supervenience/Constitutive Relevance* requires simultaneous changes. Hence, we have a way of solving the conceptual problem as well as a criterion of how to empirically identify *Causation*.

To grant compatibility between interventionism and the mechanistic view, BLINDED suggests a liberal reading of interventions according to which any kind of dependence relation can be picked out by interventions. Thus, according to BLINDED, interventions alone cannot disambiguate between (1), (2), and (3). Instead, we rely on non-interventionist manipulations — so-called *mere interactions* — to detect spatial and temporal parthood relations. If C and P are spatio-temporally related as part and whole, we have evidence against *Causation*.

### 4. Excluding *Common Cause*

#### 4.1 Fat-handedness

If we can exclude a causal relation between P and C, say, because we found that they occur simultaneously, we still need to disambiguate between interpretations (2) and (3).

Baumgartner & Gebharter (2015) argue that if C is part of the supervenience base of P, and C is constitutively relevant for P, necessarily, an intervention into P is a (direct) common cause of both P and C. Whether changes in C and P are *merely* due to a common cause or whether *additionally* there is supervenience relation involved, can be settled by testing whether interventions into P are always *fat-handed*. That is, by testing whether *every* intervention into P affects any of the mechanism's components. If this is the case, we have evidence for *Supervenience/Constitutive Relevance* rather than just a *Common Cause*.

This account is problematic. First, in order for C to be constitutively relevant for P, the event represented by C has to occupy a part of the *spatial* region occupied by P. Baumgartner & Gebharter provide no resources to account for that. Second, their approach is essentially based on induction. This is empirically problematic. Third, the fat-handedness criterion only requires every intervention into P to affect *one* of the components. It is thus insufficient to rule out that P and any particular component C only share a common cause while not being otherwise related. Finally, Baumgartner & Gebharter do not capture how empirical scientists actually assess different dependency relations.

#### 4.2 Mutual Manipulability as a Causal Relation

BLINDED2 answers the empirical challenge in a different way: although *constitutive relevance* is a non-causal relation, *mutual manipulability* occurs due to a causal relation between *temporal parts* of the phenomenon (represented by variables  $P_1$ - $P_n$ ) and the mechanism's components. On this view, we can empirically disambiguate between (2) and (3) by performing additional experiments. We may, e.g., interfere with the original intervention while keeping  $P_i$  fixed. If C changes in this experimental setup, this cannot be mediated through  $P_i$ ; our original intervention must be a common cause of  $P_i$  and C.

This approach has several advantages over the fat-handedness approach. It does not rely on induction, it can be applied to individual components, and it does not rely on major modifications of the interventionist view.

### 4.3 Non-Interventionist Strategies

One problem of the strategy suggested by BLINDED2 might be that focusing on interventions alone might not be empirically realistic. It may not be practically possible to hold  $P_i$  fixed while carrying out an intervention on the original intervention. In such cases, scientists may e.g. refer to *group comparisons* instead. And even if it is empirically possible to hold  $P_i$  fixed, we need to make sure that this fixation was successful. Such *manipulation checks*, like *group comparisons*, are instances of the non-interventionist strategies discussed by BLINDED. Therefore, we suggest addressing the empirical challenge in two steps: using causal mutual manipulation studies and employing additional, partly non-intervention experiments. This way we can render interventionism and mechanistic explanations compatible and gain an empirically adequate understanding of how different interpretations of observed manipulability can be disambiguated.

### References

- Bechtel, W. (2008). *Mental Mechanisms*, London and New York: Routledge. Bechtel, W., Abrahamsen, A. (2005). "Explanation. A mechanist alternative", *Stud Hist Philos Biol Biomed Sci*, 36, 421-441.
- Craver, C. F. (2007). *Explaining the Brain*, New York: OUP.
- Craver, C. F. & Darden, L. (2014). *In Search of Mechanisms*, Chicago: University of Chicago Press.
- Baumgartner M., & Gebharter, A. (2015). Constitutive relevance, mutual manipulability, and fat-handedness, *British Journal for the Philosophy of Science*, doi: 10.1093/bjps/axv003.
- Illari, P., M., Williamson, J. (2011). "Mechanisms are real and local", in: P. M. Illari, F. Russo, J. Williamson (eds.), *Causality in the Sciences*, OUP, 818-844.
- Machamer, P. K., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1–25.
- Woodward, J. (2003). *Making Things Happen*, Oxford: OUP.
- Woodward, J. (2011). Interventionism and causal exclusion. Manuscript. Available at: <http://philsci-archive.pitt.edu/8651/>. Establishing Constitutional Relations, in Theory and in Practice



---

*Establishing Constitutional Relations in Theory and in Practice*

MICHAEL BAUMGARTNER

University of Geneva

michael.baumgartner@unige.ch

LORENZO CASINI

University of Geneva

lorenzodotcasini@gmail.comIt

---

It is a popular maxim in recent debates about mechanistic explanation that a powerful strategy to explain the upper level behavior  $\Psi$  of some system  $S$  consists in pinpointing the lower level mechanism that constitutes  $\Psi(S)$  (Glennan 2002; Bechtel and Abrahamsen 2005; Craver 2007). This raises the methodological follow-up question as to how mechanisms are best identified, i.e. how those of  $S$ 's spatiotemporal parts  $\mathbf{X} = X_1, \dots, X_n$  are singled out, whose activities  $\Phi(\mathbf{X}) = \Phi_1(X_1), \dots, \Phi_n(X_n)$  are constitutively relevant to  $\Psi(S)$ . According to a prominent answer due to Craver (2007), constitutional relations are experimentally uncovered along roughly the same lines as causal relations—notwithstanding the fact that constitution and causation are very different relations (Craver and Bechtel 2007).

Since the time of Mill (1843), one of the dominant experimental approaches to uncovering causal relations, influentially systematized by Woodward (2003), consists in intervening on causes (in controlled environments) to change their effects. Craver (2007) argues that the same basic idea—with a mutuality tweak—applies to discovering constitutional relations. Subject to his *mutual manipulability* account of constitution (MM), the behavior  $\Phi_i(X_i)$  of a spatiotemporal part  $X_i$  of  $S$  is a constituent of  $\Psi(S)$  iff it is possible to 'ideally' intervene — in the sense of Woodward (2003, 98) — on  $\Phi_i(X_i)$  such that  $\Psi(S)$  changes, and on  $\Psi(S)$  such that  $\Phi_i(X_i)$  changes (Craver 2007, 153). Identifying constitutional relations along the lines of MM, for Craver, is not only a theoretical proposal but a faithful reconstruction of scientific practice.

By drawing on a recent result of Baumgartner and Gebharder (2015), the first part of this paper shows that MM does not ground an adequate methodology for constitutional discovery. In short, the reason is that the idealized experiments required by MM are unrealizable *in principle*, for upper level phenomena and their constituent mechanisms are so tightly coupled that they can only be manipulated with a fat-hand, i.e. via common causes. Furthermore, less rigorous but realizable experimental set-ups systematically underdetermine the inference to constitutional relations, due to the (non-ideal) fat-handed nature of relevant manipulations. In sum, while there exist experimental designs that, given compliance with required assumptions about unmeasured background influences, conclusively establish the existence of causal relations, no such experimental designs can possibly exist for the inference to constitutional relations. Therefore, the inference to constitutional relations cannot proceed along the lines of the inference to causal relations. If scientists were to follow MM's prescriptions, their reasoning would be fallacious. Hence, if we grant that their reasoning is not fallacious, it must be reconstructed differently.

Inspired by suggestions from Simon (1962) and Wimsatt (1997), the second part of the paper draws on recent research in neuroscience to develop an ‘abductivist’ alternative to MM. Neuroscientists increasingly rely on network theory (Newman 2006) to split networks of unit activations in the brain, and patterns of co-activation between them, into distinct component modules, in order to map different cognitive/behavioral phenomena onto the different modules (Nelson et al. 2010; Meunier et al. 2009). Resulting decompositions are considered optimal if they account for the phenomena under investigation to a high degree of accuracy without introducing unnecessary detail, e.g. without exposing the internal composition of each module. This procedure, we argue, is best reconstructed as an attempt to give a redundancy-free and empirically adequate account of the phenomena—that is, to provide a maximally powerful explanation. Once an optimal decomposition has been recovered, its robustness is tested by varying the number or size of the parcellated units whose activation is being analyzed, by adding or deleting some of their putative causal connections, or by modifying the descriptive grain. When the optimality of the decomposition is shown to be robust across such modifications, constitution is taken to be established. In the end, evidence for constitution consists in the repeated failure of improving the explanatory power of the decomposition by modifying the set of components or activities.

We propose an approach to constitutional discovery that generalizes this pattern of reasoning. Constitutional relations are established by way of *abductive* inferences. More concretely, the constituents of a mechanism for an upper level behavior  $\Psi(S)$  are recovered by decomposing the corresponding system  $S$  into a set of proper spatiotemporal parts  $X$  whose behaviors  $\Phi(X)$  provide the (or a) best explanation of  $\Psi(S)$ . This goal is accomplished, we contend, if the decomposition satisfies the following constraints:

- 1 **Minimality.** The set of components  $X$  and the description of their activities  $\Phi(X)$  are minimally sufficient to deduce  $\Psi(S)$ .
- 2 **Coupling.** The behavior  $\Psi(S)$  and the elements of  $\Phi(X)$  are so tightly coupled that: (i) all causes of  $\Psi(S)$  are common causes of  $\Psi(S)$  and some  $\Phi_i(X_i)$ ; (ii) every  $\Phi_i(X_i)$  has at least one cause that is a common cause of  $\Phi_i(X_i)$  and  $\Psi(S)$ .
- 3 **No de-coupling.**  $\Psi(S)$  and  $\Phi(X)$  resist de-coupling across all expansions of the variable sets  $\Psi(S)$  and  $\Phi(X)$ .

Against the background of our proposal, the role of top-down and bottom-up manipulations by fat-handed interventions on  $\Psi(S)$  and  $\Phi(X)$  is not the one depicted in MM. Contrary to MM, successful combinations of top-down and bottom-up experiments are never sufficient to warrant the inference to constitutional relations. Rather, they are a means to establish the coupling of  $\Psi(S)$  and  $\Phi(X)$ —in line with (2)—and to test whether this coupling can be broken—in line with (3). Persistent failure of attempts at de-coupling are best explained by introducing a constitutional dependency between  $\Psi(S)$  and  $\Phi(X)$ . In sum, experimental manipulations of  $\Psi(S)$  and  $\Phi(X)$  provide evidence for the robustness of the entanglement between  $\Psi(S)$

and its putative decomposition  $\Phi(\mathbf{X})$ . However, there may exist multiple decompositions that equally comply with constraints (2) and (3), i.e. decompositions among which experimental manipulations alone cannot discriminate. The deducibility of  $\Psi(S)$  from sets  $\mathbf{X}$  and  $\Phi(\mathbf{X})$  that are minimal—in line with (1)— then identifies a redundancy-free decomposition among the ones complying with (2) and (3). The joint satisfaction of constraints (1) to (3) justifies an abductive inference to  $\Phi(\mathbf{X})$  constituting  $\Psi(S)$ , namely an inference to a maximally simple and robust decomposition—an inference to the (or a) best mechanistic explanation of  $\Psi(S)$ . Ultimately, the constraints form a justificatory basis for the inference to constitution: the higher our confidence that (1) to (3) are satisfied, the more the inference is justified.

### References

- Baumgartner, M. and A. Gebharder (2015). Constitutive relevance, mutual manipulability, and fat-handedness. *The British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axv003.
- Bechtel, W. and A. Abrahamsen (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Biological and Biomedical Sciences* 36(2), 421–441.
- Craver, C. and W. Bechtel (2007). Top-down causation without top-down causes. *Biology & Philosophy* 22, 547–563.
- Craver, C. F. (2007). *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.
- Glennan, S. (2002). Rethinking mechanistic explanation. *Proceedings of the Philosophy of Science Association* 2002(3), 342–353.
- Meunier, D., R. Lambiotte, A. Fornito, K. D. Ersche, and E. T. Bullmore (2009). Hierarchical modularity in human brain functional networks. *Frontiers in Neuroinformatics* 3(37), 1–12.
- Mill, J. S. (1843). *A system of logic*. London: John W. Parker.
- Nelson, S. M., A. L. Cohen, J. D. Power, G. S. Wig, F. M. Miezin, M. E. Wheeler, K. Velanova, D. I. Donaldson, J. S. Phillips, B. L. Schlaggar, and S. E. Petersen (2010). A parcellation scheme for human left lateral parietal cortex. *Neuron* 67, 156–170.
- Newman, M. E. J. (2006). Modularity and community structure in networks. *PNAS* 103(23), 8577–8582.
- Simon, H. A. (1962). The architecture of complexity. In *Proceedings of the American Philosophical Society*, Volume 106, pp. 462–482.
- Wimsatt, W. (1997). Aggregativity: reductive heuristics for finding emergence. *Philosophy of Science* 64(4), 372–384.
- Woodward, J. (2003). *Making things happen. A theory of causal explanation*. Oxford: Oxford University Press.

---

*Against the Grain: An Investigative Model for the Ancestral Health Movement*

RICK MORRIS

UC Davis

jemorr@ucdavis.edu

---

The ancestral health movement (AHM) is a family of attempts to integrate concepts from evolutionary biology into the human and health sciences, in e.g. evolutionary medicine or the “Paleo diet”. The AHM looks to evolutionary biology not only to develop hypotheses about the causes of human health outcomes, but also to develop recommendations for health interventions. Needless to say, the AHM has been extremely controversial. Critics have characterized the AHM as making vague and inaccurate claims about human behavior and health outcomes. Advocates of the AHM’s approach charge critics with misunderstanding the nature of the claims made. As a consequence, much of the discussion over the issues involved has been unprofitable.

In this project, I discuss six questions which the AHM must answer to achieve its explanatory and interventionist goals in relation to one of its central claims: many negative human health outcomes are the consequence of a “mismatch” (Gluckman 2009, Nesse 2012) between contemporary human environments and “the” evolved physiology of the human organism.

I call the first question the mapping question: the AHM advocate, in suggesting that human populations are better-suited to some environments than others, must identify on principled grounds which populations she will examine. Advocates tend to compare “ancestral” populations to “contemporary” populations. I will highlight some of the challenges involved in doing so.

Second, the epidemiological question, which requires the advocate to delineate the differing health outcomes between the “ancestral” and “contemporary” populations. If health outcomes do not differ between populations in different environments, then the pathogenic (or non-pathogenic) nature of those environments is cast into doubt and the plausibility of the AHM is undermined.

Third, the evolutionary question, which requires the advocate to show the relevant genetic similarities between the populations selected for study. Human evolution may well have continued to the point that the conditions which promoted health and longevity in “ancestral” populations would be pathogenic to “contemporary” populations. If so, this would undermine the usefulness of “ancestral” populations in developing explanations and interventions in the health and biological sciences.

Fourth, the causal question, which requires the advocate to provide causal mechanisms for the identified outcomes. The advocate will, at the very least, have to give a biologically plausible account for how

environmental changes have led to changed health outcomes. This requires pointing to specific environmental changes and the effects of those changes on the human organism.

Fifth, the intervention question, wherein the advocates propose their interventions. A defining feature of the AHM is that it seeks not only to explain human health outcomes in evolutionary terms, but to use evolutionary biology to alter those outcomes. It is this question which, if answered successfully, will give the AHM utility in clinical practice, public policy, and individual life choices.

Finally, the evidential question, which requires that the advocate demonstrate empirical support for her claims. I do not intend to argue for particular criteria of adequate scientific evidence in this project, but I include the question to ensure that the evidential challenges are tackled explicitly and to avoid some of the vulnerabilities of the evidence used to support e.g. evolutionary psychology.

I argue that these questions, taken together, constitute an adequate investigative model for the different branches of the AHM. If these questions are answered clearly and well-supported evidentially, then the advocate should be taken to have made his case for the claims evaluated. If, on the other hand, the answers are incomplete or unsatisfactory, the critic will have clearly-identified vulnerabilities to which she can point. Developing a mutually-agreed-upon rubric of adequacy is critically important for evaluating the AHM effectively.

This model need not be taken as an all-or-nothing approach in evaluating the AHM: perhaps evolutionary medicine will stand up well under this scrutiny, but the “Paleo Diet” will be incapable of answering these questions satisfactorily. Such a failure need not undermine the AHM in its entirety, of course; instead, a failure in one branch of the AHM will show that evolutionary biology is not a useful tool for developing interventions in the particular area of human health.

My ultimate intent here is not to make a positive claim about how well (or poorly) different branches of the AHM make their case. Instead, I propose a model for investigating their claims which will simplify the discussion for the advocate and the skeptic alike by identifying and clarifying the disputed claims and providing a rubric by which each can evaluate the various claims of the AHM.

## Symposia &amp; Contributed Papers VII

**Physics and the Nature of Computation****Symposium**

Organizer: Chris Timpson &amp; Owen Maroney

Chair: Karim Thebault

Room 5G, Friday 15:30 – 17:30

*Is Information Physical?*

CHRIS TIMPSON

University of Oxford

christopher.timpson@bnc.ox.ac.uk

OWEN MARONEY

University of Oxford

owen.maroney@philosophy.ox.ac.  
uk*When does a Physical System Compute?*

VIV KENDON

University of Durham

viv.kendon@durham.ac.uk

CLARE HORSMAN

University of Oxford

clare.horsman@cs.ox.ac.uk

*The Mechanistic View of Computation and Quantum Computers*

ARMOND DUWELL

University of Montana

armond.duwell@umontana.edu

**General Description**

This symposium will bring together some of the leading researchers in physics, and in philosophy of physics, who are trying to understand the nature of computation at the fundamental level. There are a number of reasons to be interested in this question, ranging from pressing practical and technical physics issues in the theory and experiment of quantum information processing devices, via questions of what the ultimate limits of physical computation might be, to the assessment of novel bold metaphysical claims about the underlying computational or informational nature of physical reality.

The last three decades have seen a great flowering in physics and in computer science around the concept of quantum computation, and more broadly, around the question of what the ultimate limits of computation allowed by the physical world might be. This followed Deutsch's 1985 liberalisation of the concept of the

Turing machine, and his and Feynman's (1982) emphasis on the great possibilities that might be opened-up once we free our thinking about computational models from the pre- conceptions of classical computation and of classical physics - thereby seeking to harness and to exploit the non-classical features which exist in the actual world.

Great progress has been made both in understanding the general capacities of quantum computers theoretically, and in building prototypes or small-scale systems in the lab. (See e.g. Nielsen and Chuang 2010, and Ladd *et al* 2010, respectively.) However, a range of fundamental questions remain unanswered. Amongst these is the simple - or simple-seeming - question: what does it take for a system to count as instantiating a particular computation in the first place?

Whilst this question is made vivid by the search for viable (and scalable) physical implementations of quantum computers - since one needs criteria against which to assess whether one's device would count as doing what it is supposed to - its scope is in fact more general than this, since one can – and indeed, should - ask this question with regard to whatever background computational model one might have in mind: whether quantum, classical (both analogue and digital), or putative hyper-computational (or other non-standard) model.<sup>1</sup> Moreover, even if one has established a set of conditions (whether necessary or sufficient) for a device to count as being a computer of a particular kind (e.g., DiVincenzo 2001 gave an early well-known set of conditions for quantum computers) there remains the further very important question of what kind of *facts* it takes (e.g., whether wholly naturalistic, or whether in some sense mind- or representation-dependent) for these conditions to count as being satisfied. Finally there remain the tasks of i) explaining why *these* conditions are the right ones, for a given computational model; ii) explaining what the connections are, if any, between the requisite conditions for different *kinds* of computational models; and iii) establishing whether there is a completely *general* set of conditions which holds for *any* possible computational model, conceived or unconceived.

This question – what does it take for a physical system to count as instantiating a particular computation – is the key topic of our proposed symposium.

As we have said, the question is of practical and theoretical importance internally to physics. On the one hand, it needs to be addressed in order to allow one to assess whether various proposals of physical implementations would actually count as instantiating a given computational process: this problem arises not only for proposed quantum computers in the lab, but for the question of computation in naturally occurring processes - such as biological ones – and for the question of hypercomputation. (These matters have at times been very controversial, see e.g. discussions in Teuscher 2004, Cooper *et al.* 2007, and Hagar and Korolev 2007.) On the other hand, we need clarity on the matter in order to settle fundamental questions

---

<sup>1</sup> Hyper-computation is the computation of functions which are not Turing-computable. Quantum computers, whilst being more powerful than classical computers in that they can evaluate certain important functions exponentially more quickly, can only compute Turing-computable functions – the same class as classical computers.

about the ultimate energy costs in terms of heat and work of computational processes, since these latter questions cannot be addressed without an adequate understanding of the relations between abstract characterisations of computation, and their physical implementation. (See Ladyman et al (2007) and Maroney (2009) for some important discussion of these topics. The latest results (Bedingham and Maroney, forthcoming), which are prefaced on a great deal of care about what physical processes must actually be considered, indicate that the ultimate thermal costs of quantum computation are in fact greater than in the classical case, contrary to current orthodoxy.)

More broadly or conceptually, addressing this key question of when a physical system computes - and what it computes when it does so - allows us to gain a clearer view of what the fundamental status of computation is in the physical world. One important epistemological question is what ultimate limits the laws of physics might place on what can be computed. An alternative question which has intrigued some physicists pushes in the other direction and asks instead whether what can be computed might place constraints on the fundamental laws of physics (Landauer 1996; Jozsa 2004; Davies 2010; Aaronson 2013). A still more extreme view urges in response to developments in quantum information science that we should see the world as fundamentally *constituted* of information, as opposed to being constituted of material things; urging the view that the world literally is nothing but a giant (quantum) computer (Wheeler 1990; Lloyd 2006; Vedral 2012; Aaronson 2013). However, it is clear that attempting to assess such claims in a useful way requires us to know first what it *means* to say something is a computer, or is processing information.

Each of these broader questions adumbrated in the last paragraph touches on a further question which sometimes goes by the name of *realism about computation*. This is the question of whether or not facts about what, if any, computation is being performed by a given physical system are mind- or representation-independent. This topic has had some airing in debates in philosophy of mind and cognitive science, as it bears on the question of the viability of naturalistic computational theories of mind (see e.g., Piccinini 2008; Sprevak 2010). But it arises independently in the physics context too, for here also responses must be made to the trivialisation arguments of Searle (1992) and Putnam (1988) that without some significant external constraints, any physical system would count as computing every function, which is a *reductio ad absurdum*.

These general questions clustering around the core question of what it takes for a physical system to compute will be addressed in each of the contributions to this symposium. In each, a slightly different answer to the core question will be offered, and overall the respective merits of each answer assessed, along with the corresponding consequences for the broader considerations of the fundamental scope and nature of computation in the physical world. The format of the symposium will be three talks, followed by a general round-table discussion.



**Abstracts****1. Chris Timpson and Owen Maroney: *Is information physical?***

We have a conundrum. The physical basis of information is clearly a highly active research area. Yet the success of information theory comes precisely from separating it from the problems of building physical systems to perform information processing tasks. It is primarily developments in quantum information over the last two to three decades that have undermined this separation, leading to suggestions that information itself is a physical entity and must be part of our physical theories. We will argue to the contrary: rejecting the claims that information is physical provides a better basis for understanding the fertile relationship between information theory and physics. In developing this claim we will pay particular attention to how (quantum) computational processes should be characterised – as distinct from any other physical processes – and argue for the claim that for a process to count as an information processing (computational) one, there must be someone who would be *informed* by the result of the process.

**2. Viv Kendon and Clare Horsman: *When Does a Physical System Compute?***

Computing is a high-level process of a physical system. Recent interest in non-standard computing systems, including quantum and biological computers, has brought this physical basis of computing to the forefront. There has been, however, no consensus on how to tell if a given physical system is acting as a computer or not; leading to confusion over novel computational devices, and even claims that every physical event is a computation. In [Proc. Roy. Soc. A 470 20140182] we have introduced a formal framework - Abstract/Representation (AR) theory - that can be used to determine whether a physical system is performing a computation. We demonstrate how the abstract computational level interacts with the physical device level, in comparison to the use of mathematical models in experimental science. This powerful formulation allows a precise description of experiments, technology, computation, and simulation, giving our central conclusion: physical computing is the use of a physical system to predict the outcome of an abstract evolution. We give conditions for computing, illustrated using a range of non-standard computing scenarios. The AR theory framework also covers broader computing contexts, where there is no obvious human computer user. To address these cases, we introduce the notion of a 'computational entity', and its critical role in defining when computing is taking place in physical systems.

**3. Armond Duwell: *The mechanistic view of computation and quantum computers.***

The discovery and development of quantum computers has raised a number of interesting philosophical issues. Perhaps the most important issue is how to explain why quantum computers are faster than classical computers. In the literature, one can find, arguably, two polar opposite positions espoused. One position is that the source of quantum speedup was due to the fact that quantum computers could compute many values of a function in a single step. The alternative position is that quantum computers don't do this at all,

and in fact perform the requisite computational task by doing fewer computations than classical computers performing the same task. In this paper, I want to utilize the mechanistic view of computers to shed light on the debate. I will argue that the mechanistic view helps us understand that the polar opposite positions described above can be seen as consequences of conflicting intuitions about the appropriate computational description of quantum systems that perform computational tasks, and not a disagreement about the fundamental feature of quantum systems that allows for quantum speedup. The mechanistic view provides us with a principled means of understanding the usefulness and limits of computational descriptions of phenomena, and also helps us understand the essential differences between classical and quantum computers.

### References

- Aaronson (2013) *Quantum Computing Since Democritus* (CUP) Cooper et al eds. (2007) *Computation and Logic in the Real World* (Springer)
- Davies (2010) in Davies and Gregersen (eds.) *Information and the Nature of Reality* (CUP) Deutsch (1985) Proc. R. Soc. Lond. A, 400(1818):97–117.
- DiVincenzo (2000) *Fortschr. Phys.* 48:771–783. Feynman (1982) *Int J Theor Phys*, 21(6/7): 467-488. Hagar and Korolev (2007) *Phil Sci* 74(3): 347–363.
- Jozsa (2004) *IBM Journal of Research and Development*, 4(1): 79-85. Ladd et al (2010) *Nature*, 464(7285):45–53.
- Ladyman et al (2007) *Stud Hist Phil Mod Phys* 38(1): 58–79. Landauer (1996) *Physics Letters A*, 217: 188-193.
- Lloyd (2006) *Programming the Universe* (Jonathan Cape) Maroney (2009) *Phys Rev E* 79:031105.
- Nielsen and Chuang (2010) *Quantum Computation and Quantum Information* (CUP) Piccinini (2008) *Phil Stud* 137: 205-241.
- Putnam (1988) *Representation and Reality* (MIT Press) Searle (1992) *The Rediscovery of Mind* (MIT Press)
- Sprevak (2010) *Stud Hist Phil Sci* 41(3): 260-70.
- Teuscher ed. (2004) *Alan Turing: Life and Legacy of a Great Thinker* (Springer-Verlag) Vedral (2012) *Decoding the Universe* (OUP)
- Wheeler (1990) in Zurek (ed.) *Complexity, Entropy and the Physics of Information* (Addison-Wesley)

**Situated Cognition and Scientific Practice**

**Symposium**

Organizer: Adam Toon & Sabina Leonelli

Chair: Markus Werning

Room 5F, Friday 15:30 – 17:30

---

*Empiricism for Cyborgs*

ADAM TOON

University of Exeter

a.toon@exeter.ac.uk

---

*Building Computational Representations for Scientific Discovery: A Distributed Cognition Account*

MILES MACLEOD

University of Helsinki Georgia

miles.macleod@helsinki.fi

NANCY NERSESIAN

Institute of Technology, Harvard University

nancyn@cc.gatech.edu

---

*Distributed Reasoning in Data-Centric Science*

SABINA LEONELLI

University of Exeter

s.leonelli@exeter.ac.uk

---

*Hardwig's Dilemma and a Hidden Individualism in Social Theories of Scientific Knowledge (cancelled)*

AXEL GELFERT

National University of Singapore

axel@gelfert.net

---

*Active Externalism, Virtue Reliabilism and Scientific Knowledge*

ORESTIS PALERMOS

University of Edinburgh

spalermo@exseed.ed.ac.uk

---

**General Discription**

*Situated cognition* is a growing movement in cognitive science that stresses the importance of interaction between the brain, body and environment in carrying out cognitive tasks. The term “situated cognition” encompasses a number of different approaches – such as *embodied cognition* and *distributed cognition* – across a wide range of disciplines, from anthropology to robotics (for an overview, see Robbins and Aydede 2009). Although extremely diverse, much of this work shares an opposition to a traditional view of cognition as a matter of internal, “disembodied” computational processes involving symbolic representations. These developments in cognitive science have inspired a number of philosophers of mind to propose a radical shift in our view of the nature of mind and cognition, arguing that we should no longer think of cognition as something that happens inside the head. Instead, according to the *extended mind thesis*, cognitive processes – and even mental states - sometimes extend outside our brains and bodies into the environment. These views are most often associated with Andy Clark and David Chalmers (1998), but have also been defended by many others, including Richard Menary (2007), Mark Rowlands (1999), Mike Wheeler (2005) and Robert Wilson (2004).

In recent years, a number of authors have argued that work on situated cognition might provide a fruitful framework for studying scientific reasoning. Ronald Giere (2006) has suggested that distributed cognition can be used to analyse many different aspects of scientific practice that are grounded on collaborative efforts, such as the development and use of models and diagrams, and that this approach may allow philosophers to build upon well-known studies by sociologists of science such as Bruno Latour and Karin Knorr-Cetina. In a similar vein, and drawing on work by cognitive scientist Ed Hutchins (1995), Nancy Nersessian and her colleagues have claimed that scientific laboratories can be viewed as “evolving distributed cognitive systems”, where scientific judgement and expertise need to be evaluated at the collective level rather than at the level of each participating individual (e.g. Nersessian et al. 2003). More recently, Sabina Leonelli (2014) has suggested that distributed reasoning may provide a fruitful framework to understand how biologists make sense of large data collections assembled from a variety of sources – a view also defended, under the heading of ‘extendedness of scientific evidence’, by Eric Kerr and Alex Gelfert (2014).

In addition to these applications of work in situated cognitive science to scientific practice, there is a burgeoning literature on the implications of the extended mind thesis within epistemology. Authors in this area have already begun to explore the consequences of the notion of extended mental states for a range of issues, such as the relationship between knowledge and cognitive ability (Clark et al. 2012; Pritchard 2010; Vaesen 2011). Focusing in particular on scientific knowledge, Adam Toon (2014) has argued that the extended mind thesis offers scientific realists a new argument against constructive empiricists concerning the possibility of gaining knowledge using instruments.

Such work suggests that situated cognition offers valuable resources for understanding scientific practice, while also raising a range of further questions for philosophers of science. For example, how might situated

cognition contribute to existing work on instruments and material culture in science (e.g. Hacking 1983, Galison 1997)? Does it offer new theoretical or methodological resources for understanding these aspects of scientific practice?

Distributed cognition claims that cognition can be “distributed” across social groups, as well as tools and the environment. How does this idea relate to existing work in social epistemology and sociology of science? Does it offer new resources for understanding changes in the organisation of scientific research, such as what the European Union is calling “Science 2.0”? Can it provide us with a way to reconcile cognitive and social approaches to science (Nersessian 2005; Giere 2006)? Does understanding science in terms of situated cognition require us to endorse controversial claims about extended mental states, or even group minds? As the brief literature survey above makes clear, a key difficulty in using the notion of situated cognition to unravel fundamental issues within the philosophy of science is the disparate range of disciplinary approaches involved in developing and fostering this approach. Very few philosophers of science have a comprehensive grasp of contemporary cognitive science and philosophy of mind; even fewer can supplement such knowledge with an understanding of contemporary literature in social epistemology as well as the social studies of scientific practice. In such a situation, it seems that the very effort to develop a philosophy of science informed by the notion of situated cognition involves bringing together these different fields and starting an interdisciplinary discussion around the ways in which philosophers from different traditions are interpreting the various claims associated to situated cognition, and the implications that can be drawn from such interpretations.

To this aim, this symposium brings together philosophers working on situated cognition from the perspective of general philosophy of science, philosophy of science in practice, philosophy of the cognitive sciences, philosophy of mind and social epistemology. Their presentations, and the concluding discussion, will explore the potential of situated cognition for understanding scientific practice and the implications that such an approach might have for philosophy of science.

## **Abstracts**

### **1. Adam Toon: *Empiricism for cyborgs***

One important debate between scientific realists and constructive empiricists concerns whether we observe things using instruments. Realists argue that we do and that the development of scientific instruments has enabled us to observe new realms of phenomena previously beyond the reach of our senses. In contrast, constructive empiricists argue that the use of instruments does not count as observation. Instead,

observation remains limited to the use of our unaided senses and, as a result, for the constructive empiricist, so too does scientific knowledge.

Realists often speak of instruments as ‘extensions’ to our normal cognitive capacities. For example, in his book on instruments and computational science, revealingly entitled *Extending Ourselves*, Paul Humphreys argues that “[o]ne of science’s most important epistemological and metaphysical achievements has been its success in enlarging the range of our natural human abilities” (2004, pp. 3-4). In this paper, I will ask whether the realist may flesh out her view of instruments by looking to recent work in philosophy of mind and cognitive science and, in particular, the extended mind thesis.

Drawing on a range of recent movements in cognitive science – such as situated, distributed and embodied cognition - proponents of the extended mind thesis claim that cognitive processes sometimes extend beyond our brains and bodies into the environment. Although some have begun to explore the consequences of the extended mind thesis for epistemology, its implications for the philosophy of science have yet to be properly explored. In this paper, I will suggest that the extended mind thesis offers a way to make sense of realists’ talk of instruments as extensions to the senses and that it provides the realist with a new argument against the constructive empiricist.

## **2. Miles MacLeod and Nancy Nersessian: *Building computational representations for scientific discovery: a distributed cognition account***

How is it possible that an engineer with little knowledge of biology and only a few months of modeling biological systems can make significant biological discoveries with highly complex subject matter? The cases we will develop here might seem exceptional, but over the course of our 5 year ethnographic investigation into the problem-solving practices of integrative systems biologists we have witnessed many of these engineer-modelers making important contributions to the understanding and control of biological systems. We will argue in this presentation that the explanation requires a distributed account of cognitive processing in which the modeler and the model become a coupled system in the processes of building the model and running the simulations. Thus, inferences to novel discoveries are made by distributing model based reasoning across this system.

Much of the work on external representations within DC has focused on capturing detailed descriptions of the way external representations are used in highly structured task environments, such as ship navigation and landing of aircraft, and the way these representations change the nature/cost of cognitive tasks. Less understood are the processes of generating/building external representations to alter task environments (Kirsh, 1996) and the role played by this building process in cognitive processes while problem- solving (Chandrasekharan & Stewart, 2007). As Schwartz & Martin (2006) observe, “most cognitive research has been silent about the signature capacity of humans for altering the structure of their social and physical environment.” However, a central premise of the DC perspective is precisely this, as Hutchins has succinctly

stated, "Humans create their cognitive powers by creating the environments in which they exercise those powers" (1995b, p. 169). However, how humans go about creating their environments has not received much attention in the DC literature.

Since building problem-solving environments is a major component of scientific research (Nersessian et al. 2003, Nersessian, 2012), scientific practices provide an especially good locus for examining the human capability to extend and create cognitive powers, particularly through building new external representations. In this paper, we focus on an exemplar of the building of a computational model – a complex external representation – and examine the role this external representation, specifically the process of building it, plays in structuring, as well as altering, the task of making scientific discoveries collaboratively in a systems biology laboratory.

The central metaphor of DC is that of offloading cognitive processing to external representations. However, we will argue that offloading is not the right metaphor to understand the imagination process developed through the building of novel computational representations. Rather, the metaphor should be that of coupling between internal and external representations (Chandrasekharan & Stewart, 2007; Nersessian et al., 2003; Osbeck & Nersessian, 2006; Nersessian, 2009). This extension of DC requires moving the analysis from the use of external structures to lower cognitive load – the focus of DC till now – to the processes of building external representations to create coupled systems for reasoning and inference. As we will show such a DC analysis is necessary for understanding how researchers handle complex subject matter and manage discovery in integrative systems biology. By understanding the operations required of the human agent within this discovery system we can better organize research and train researchers for this kind of work.

### **3. Sabina Leonelli: *Distributed Reasoning in Data-Centric Science***

This paper discusses the patterns of reasoning involved in data-centric research, and particularly (1) the extent to which it may be regarded as distributed across vast networks of scientists with different backgrounds and interests; and (2) the epistemic importance acquired by material and social scaffolds for research as a consequence of this view. In previous work, I have claimed that the generative power of data-centric biology derives from the opportunity to assess the evidential value of data from a variety of viewpoints, including diverse theoretical backgrounds, experimental traditions and disciplinary training. In this talk, I will argue that data-centric biology should be viewed as a collective endeavour, and its results as an achievement of the complex network of expertises underlying data journeys and re-use. Researchers make important choices at all stages of data handling.

Individuals are called upon to decide how to set up experiments and calibrate instruments that produce the data in the first place; how data should be formatted, mined and visualised; how data should be interpreted and which evidential value they acquire in different research contexts. For any given datasets,

several individuals, sometimes hundreds of them, are involved in making those decisions. Thanks to the integrative platforms provided by computers and internet access, as well as specific regulatory and institutional structures enabling data dissemination, those individuals often will not know each other, they might have very different expertises and priorities, and they might be working within different epistemic cultures. Most importantly, each of those individuals might possess a different form of embodied knowledge, and thus make use of different skills and commitments when handling data. In such a situation, the intuition of scientific understanding as resulting from the heroic efforts of a lone genius needs to be abandoned. The ability to assign evidential value to biological data is not generated through an overarching synthesis, but rather through the fragmented efforts of several different groups of researchers, which offers unique opportunities for integration and cross-pollination. I interpret this situation as a case of extensive distributed reasoning, where cognitive achievements are earned by scientific collectives. As I shall discuss, this view has strong connections to the idea of distributed cognition famously championed by Edwin Hutchins (1995) in his study of collective agency in ship navigation, an idea that Andy Clark extends to all cases where “computational power and expertise is spread across a heterogeneous assembly of brains, bodies, artifacts, and other external structures” (1997, 77).

**4. Axel Gelfert: *Hardwig’s Dilemma and a Hidden Individualism in Social Theories of Scientific Knowledge***  
(cancelled)

One of the fundamental insights of the social epistemology of science over the past couple of decades has been the recognition that our dependence on others for knowledge and information runs deeper than has traditionally been acknowledged. The kinds of claims we typically regard as scientific knowledge – that ‘DNA encodes genetic information’, that ‘modern humans first evolved in East Africa’, etc. – are only the end product of a socially distributed process of inquiry and information-sharing. When we, as individuals, hold such claims to be true, we implicitly rely on the cognitive processes of other epistemic agents for the formation, sustainment, or reliability of our beliefs. This gives rise to a tension between the resolutely social nature of science as an activity and the seemingly unproblematic way in which we credit ourselves with knowledge of – sometimes arcane – scientific facts. For, no individual knower could realistically hope to secure any significant portion of scientific knowledge all on her own. Ironically, it is the fact that there is simply too much relevant scientific knowledge and evidence for a single human cognizer to process that precludes fully crediting any one person with it. One can put this in the form of a dilemma, according to which, to quote John Hardwig, ‘there can no longer be knowledge in many scientific disciplines because there is now too much available evidence’. Social theories of scientific knowledge have attempted to defuse this dilemma by emphasizing communitarian ideals of vigorous intra-group discussion, along with the various ‘processes by which a scientific community collects, sifts, and weighs evidence’ (Susan Haack). Yet, as I argue in the present paper, such social theories of scientific knowledge are still marked by a strong – albeit hidden



– individualism, in that they construe the gathering and assessment of evidence as essentially the result of a collective exercise of individual cognitive and epistemic capacities. To be sure, no individual could reasonably hope to achieve much all on their own, which is why vigorous intra-group discussion and ‘having several people make the same observation’ (Haack) is needed, but in the final analysis, on these accounts, individual reasoning processes still do all the epistemic work. This assumption, I argue in the present paper, can be questioned from two angles: first, by recognizing the pressures of contemporary scientific practice, which increasingly rely on automated processes (e.g. numerical analysis, filtering, etc.) that leave less room for the exercise (even collectively) of individual cognitive capacities; and second, by taking inspiration from externalist theories of knowledge, which drive a wedge between the individual first-person assessment of knowledge claims and the justification-conferring reliability of the actual processes of knowledge-generation.

### 5. Orestis Palermos: *Active Externalism, Virtue Reliabilism and Scientific Knowledge*

Combining active externalism in the form of the extended and distributed cognition hypotheses with virtue reliabilism can provide the long sought after link between mainstream epistemology and philosophy of science. Specifically, by reading virtue reliabilism along the lines suggested by the hypothesis of extended cognition, we can account for scientific knowledge produced on the basis of both hardware and software scientific artifacts (i.e., scientific instruments and theories). Additionally, by bringing the distributed cognition hypothesis within the picture, we can introduce the notion of epistemic group agents, in order to further account for collective knowledge produced on the basis of scientific research teams.

### References

- Clark, A. & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58(1): 7–19.
- Clark, A., Pritchard, D. & Vaesen, K. (Eds.). (2012). Extended cognition and epistemology [Special issue]. *Philosophical Explorations*, 15(2).
- Galison, P. (1997). *Image and Logic*. Chicago: University of Chicago Press.
- Giere, R. (2006). *Scientific Perspectivism*. University of Chicago Press.
- Hacking, I. (1983). *Representing and Intervening*. Cambridge: Cambridge University Press.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA: MIT Press.
- Kerr, E. and Gelfert, A. (2014) The “Extendedness” of Scientific Evidence. *Philosophical Issues* 24.
- Leonelli, S. (2014). Data Interpretation in the Digital Age. *Perspectives on Science* 22(3): 397- 417.
- Menary, R. (2007). *Cognitive integration: mind and cognition unbounded*. Basingstoke: Palgrave Macmillan.
- Nersessian, N. J. (2005). Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. In M. Gorman, R. Tweney, D. Gooding, & A. Kincannon, (Eds.) *Scientific and technological thinking* (pp. 17-56). New Jersey: Erlbaum.
- Nersessian, N. J., Kurz-Milcke, E., Newstetter, W. C., & Davies, J. (2003a). Research laboratories as evolving distributed cognitive systems. *Proceedings of the 25th Annual Conference of the Cognitive Science Society*: 857-862.
- Pritchard, D. (2010). Cognitive ability and the extended cognition thesis. *Synthese*, 175(1): 133–151.
- Robbins, P. and Aydede, M. (eds.) (2009). *The Cambridge Handbook of Situated Cognition*. Cambridge: Cambridge University Press.
- Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes*. Cambridge: Cambridge University Press.

- Toon, A. (2014). Empiricism for cyborgs. *Philosophical Issues*, 24(1): 409-425.
- Vaesen, K. (2011). Knowledge without credit, exhibit 4: extended cognition. *Synthese*, 181(3): 515–529.
- Wheeler, M. (2005). *Reconstructing the Cognitive World*. Cambridge, MA: MIT Press.
- Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences: Cognition*. Cambridge: Cambridge University Press.

---

*Relative Modalities and Chance*

QUINN HARR

University of Maryland

qharr@umd.edu

---

Recently there has been a relatively sustained debate about whether the existence of objective, non-trivial chances is compatible with determinism.<sup>1</sup> To those with clear incompatibilist intuitions, there may seem to be nothing to debate about: espousing compatibilism about determinism and chance merely bespeaks conceptual confusion. But there are good reasons to worry that such a simple incompatibilist conclusion is premature, reasons ranging from the scientific (e.g., the existence of deterministic theories such as classical statistical mechanics which have recourse to what are presumably objective chances) to the more pedestrian (e.g., the intuition that even in deterministic worlds there can still be fair coin flips). My concern in this paper, however, is not with this debate *per se* but rather with a specific way in which one might try to resolve it. The method I have in mind—an appeal to a semantic analysis of ‘chance’—was notably employed in a recent paper by Antony Eagle (2011). I argue that this methodology represents a fundamentally flawed approach to the debate about the compatibility of determinism and objective chance. Though my arguments in this paper are directed specifically against this methodology as instantiated in Eagle (2011), they generalize to others who would adopt a similar approach: whatever the correct semantic analysis of ‘chance’, the compatibilist debate remains a separate and independent issue. Thus it should be noted that although I am arguing against one sort of compatibilist position a person might adopt, I am not thereby arguing in favor of incompatibilism about determinism and chance.

The basic argument Eagle offers in favor of compatibilism and determinism and chance can be reconstructed as follows:

- E1) “x has a (non-trivial) chance of  $\varphi$ -ing” is equivalent to “x can  $\varphi$ ”.
- E2) ‘Can’ is a relative modality with contextually-fixed parameters: the set of possibilities relative to which a ‘can’ statement is assessed for truth is determined contextually.
- E3) Given E1) and E2), ‘has some (non-trivial) chance’ is also a relative modality with contextually-fixed parameters.

---

<sup>1</sup> Representative contributions to this debate include Loewer (2001), and Eagle (2011), both compatibilists, Schaffer (2007), an incompatibilist, and Lyon (2011), who argues that the debate rests upon a false bifurcation.

E4) There are sets of possibilities relative to which a ‘can’ statement can be truthfully uttered in a deterministic world (as well as sets of possibilities relative to which it cannot).

E5) Given E1) - E4), there are sets of possibilities relative to which a ‘has some (non-trivial) chance’ statement can be truthfully uttered in a deterministic world (as well as sets of possibilities relative to which it cannot).

E6) The statement incompatibilists use to express their position, (IC),<sup>2</sup> contains a relative modality with contextually-fixed parameters: ‘has some (non-trivial) chance’.

E7) Given E5) and E6), there are sets of possibilities relative to which IC can truthfully be uttered in a deterministic world (as well as sets of possibilities relative to which it cannot).

E8) IC does not always express a true proposition, and so  $\sim$ IC (a statement compatibilists might use to express their position) sometimes will express a true proposition.

The weak points in this argument are, I believe, the inferential premises. In this paper, I focus on the third of these inferences (i.e., the inference from E5) and E6) to E7)). The basic problem with this inference is that its plausibility crucially relies on no account having been given of the truth conditions of conditional sentences containing context-sensitive relative modalities like ‘has some chance’. I outline two options for maintaining that such conditionals are themselves context sensitive and show that the one option, while yielding the context sensitivity of IC, is not itself plausible while the other option, while itself plausible, does not yield the context sensitivity of IC without further, substantive argumentation.

Briefly, these two options are a) that IC is falsified in certain contexts in a deterministic world, if it is falsified at all, because true ‘has some (non-trivial) chance’ statements can be uttered in those contexts or b) that it is falsified in certain contexts, again if falsified at all, because  $\sim$ IC can truthfully be uttered in those contexts. These two options are not the same if we make the plausible assumption that the antecedent of a conditional contributes to the contextual set of propositions against which a statement containing a context-sensitive relative modality such as ‘has some chance’ is assessed for truth. Given this assumption, we can easily imagine a conversational context  $c$  at time  $t$  such that at time  $t+1$  either IC or a ‘has some (non-trivial) chance statement could truthfully be uttered (though not both).<sup>3</sup> Thus the two ways in which IC could be falsified differ.

Either of these two ways of maintaining the context sensitivity of IC is problematic for the compatibilist. Briefly, on the first of the two views, the compatibilist must argue that a plausible metaphysical view is simply in principle inexpressible in certain contexts. On the second of the two views, the compatibilist has to argue that the antecedent of IC does not fix the relativization of the modal expression in the sentence’s consequent (or at least not restrictively enough to guarantee IC’s truth). This latter route is the only plausible one to

<sup>2</sup> “IC: If a world is deterministic, then no possible outcome in that world has any chance there other than 1 or 0” (Eagle 2011, 286).

<sup>3</sup> This results from there being no antecedent constraints in the semantic theory on the possible contextual relativizations of modal expressions. Any relativization at all is permissible.

pursue, but the sort of arguments that would be necessary to show its viability are precisely metaphysical arguments about the compatibility of determinism and chance, and not semantic ones about the context sensitivity of 'chance'.

**References**

- Eagle, Anthony. 2011. Deterministic chance. *Noûs* 45, no. 2:269-99.
- Loewer, Barry. 2001. Determinism and chance. *Studies in History and Philosophy of Modern Physics* 32, no. 4:609- 20.
- Lyon, Aidan. 2011. Deterministic probability: neither chance nor credence. *Synthese* 182:413-32.
- Schaffer, Jonathan. 2007. Deterministic chance? *British Journal for the Philosophy of Science* 58:113-40.

---

*Communism and the Incentive to Share in Science*

REMCO HEESEN

Carnegie Mellon University

rheesen@andrew.cmu.edu

---

The social value of scientific work is highest when it is widely shared. It is fortunate, then, that there exists among scientists a norm that requires them to widely share their work. Merton (1942) has called this the *communist norm* of science.

In a recent paper, Strevens (forthcoming) argues that the communist norm (like many other social norms) has the structure of a Prisoner's Dilemma: a self-interested scientist would prefer not to comply with it, even though such a scientist would be better off in a world where everyone complied with it than in a world where no one did. As a result, the persistence of (compliance with) the communist norm requires some explanation. Strevens offers the outlines of such an explanation by casting the communist norm as an (implicit) social contract.

In contrast, I argue that no special explanation is necessary, because a self-interested scientist should already prefer to widely share her research even if there was no communist norm and no social contract or other arrangement to enforce it.

What do I mean by a self-interested scientist? Like Strevens, I have in mind a scientist who aims to maximize the credit she receives for her scientific work. Credit (prestige, resources) is awarded for scientific discoveries in accordance with the *priority rule* (Merton 1957, Strevens 2003). The scientist (thus) has no bias for or against sharing her work, except insofar as it gives her credit.

Clearly, scientific discoveries must be shared in order to establish priority and so get credit for them; on this Strevens and I agree. But "the same considerations give you a powerful incentive not to share your results before you have extracted every last publication from them" (Strevens forthcoming, p. 2).

Against this, I argue that the scientist has an incentive to share her results even if she has reason to believe that more publications can be extracted from them. Using a game-theoretic model, Boyer (2014) has shown that even in the most extreme case where future publications depend directly on the results the scientist shares now (and thus are maximally helpful to competing scientists), it is in the credit-maximizing scientist's best interests to share.

However, Boyer's model involves a number of restrictive assumptions that make it unclear whether his result can support the general conclusion that I want to argue for: that it is always in the scientist's best interest to share her results as soon as she can. First, Boyer assumes that there are only two competing scientists. Second, he uses discrete time units, which means that the priority rule cannot always be applied

unambiguously. Third, the scientists in Boyer's model have equal average productivity. Fourth, he compares only the two strategies "always share everything" and "never share anything until the very end". Finally, by using backwards induction, he relies on strong assumptions on the rationality of the scientists.

To address these concerns, I consider a continuous-time model in which an arbitrary number of scientists compete, whose average productivity may be the same or different. I allow the scientists to decide for each result separately whether to share it or not, and I consider both high-rationality and boundedly rational scientists aiming to maximize their credit. In this much more general model, "always share everything" still comes out as the best strategy (where the notion of "best" depends on the rationality assumptions).

Strevens (forthcoming, p. 5) disputes the claim that a model like Boyer's or mine can show that the communist norm is incentive-compatible for individual self-interested scientists. First, the communist norm requires that any and all results should be shared, whereas it has been shown only that it is in the interest of scientists to share results that have accumulated to publishable size. Second, the communist norm requires that any and all results should be shared, whereas it has been shown only that it is in the interest of scientists to share results that are required for a journal publication, potentially hiding crucial details in, e.g., the experimental setup or the raw data.

I reply that it is not clear to me that the communist norm makes such strong requirements. Short of what is required to be publishable, scientists' actual compliance with a putative norm of sharing drops off steeply (Louis et al. 2002). If Strevens' aim is to explain a norm of sharing for these cases, he may be trying to explain something that does not exist.

Strevens may reply to this that it would be nice if scientists did share results even before they had achieved publishable size and without hiding crucial details. And perhaps his social contract approach can help science get to such an improved norm.

Fair enough, but the results from Boyer's model and mine can do the same. They suggest a clear way to make it incentive-compatible for scientists to share work below publishable size: allow smaller publications. And sharing crucial details can similarly be made incentive-compatible just by giving credit for it. If getting scientists to share these minor results or crucial details is a goal that scientists and policy makers consider important, the models give fairly clear directions on how to get there.

I conclude that the communist norm is not like a Prisoner's Dilemma. Each scientist, acting only on their narrow self-interest, has reason to comply with the norm. No special explanation, using social contracts or something else, needs to be invoked to make sense of scientists' habit to share their work widely.

## References

- Thomas Boyer. Is a bird in the hand worth two in the bush? Or, whether scientists should publish intermediate results. *Synthese*, 191(1): 17–35, 2014. ISSN 0039-7857. doi: 10.1007/s11229-012-0242-4. URL <http://dx.doi.org/10.1007/s11229-012-0242-4>.
- Karen Seashore Louis, Lisa M. Jones, and Eric G. Campbell. Macroscopic: Sharing in science. *American Scientist*, 90(4):304–307, 2002. ISSN 00030996. URL <http://www.jstor.org/stable/27857685>.

- Robert K. Merton. A note on science and democracy. *Journal of Legal and Political Sociology*, 1(1–2):115–126, 1942. Reprinted in Merton (1973, chapter 13).
- Robert K. Merton. Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review*, 22(6):635–659, 1957. ISSN 00031224. URL <http://www.jstor.org/stable/2089193>. Reprinted in Merton (1973, chapter 14).
- Robert K. Merton. *The Sociology of Science: Theoretical and Empirical Investigations*. The University of Chicago Press, Chicago, 1973.
- Michael Strevens. The role of the priority rule in science. *The Journal of Philosophy*, 100(2):55–79, 2003. ISSN 0022362X. URL <http://www.jstor.org/stable/3655792>.
- Michael Strevens. Scientific sharing: Communism and the social contract. In Thomas Boyer-Kassem, Conor Mayo-Wilson, and Michael Weisberg, editors, *Scientific Collaboration and Collective Knowledge*. Oxford University Press, Oxford, forthcoming.



---

*Models, Postulates, and Generalized Nomic Truth Approximation*

THEO KUIPERS

University of Groningen

t.a.f.kuipers@rug.nl

---

The qualitative theory of nomic truth approximation, presented by Kuipers (2000), in which ‘the truth’ concerns the distinction between nomic, e.g. physical, possibilities and *impossibilities*, rests on a very restrictive assumption, viz. that ‘theories in the making’ are maximal in the sense that they claim to precisely characterize the boundary between nomic possibilities and *impossibilities*. Fully recognizing two different functions of theories, viz. excluding and representing, this paper drops this assumption by conceiving theories in the making as non-maximal, that is, as tuples of postulates and models, where the postulates claim to exclude nomic *impossibilities* and the models claim to represent nomic possibilities, leaving ample room for temporary undecided conceptual possibilities.

Revising theories becomes then a matter of adding or revising models and/or postulates in the light of increasing evidence. This not only leads to a conceptually elegant generalized qualitative theory of nomic truth approximation, but also to a nicely fitting quantitative version. Both are built upon plausible explications of ‘truth-content’ and ‘falsity-content’.

The assumption of non-maximal theories not only differs from Kuipers (2000), but more generally from the model-theoretic and structuralist (or semantic) views, in all of which the theories are assumed to be maximal, explicitly or implicitly, respectively. Hence, formally, the two-sided approach to theories provides, relative to the most relevant standard formal views on theories, an enormous widening of the ways in which theory revision can serve the (intended or unintended) purpose of nomic truth approximation. Substantively, it does justice to three different views on theorizing in philosophy of science, viz. theorizing is mainly a matter of 1) formulating and revising postulates that intend to exclude nomic possibilities, or 2) designing and redesigning models that intend to represent nomic possibilities, or 3) the two-sided combination of them. The analysis to be presented will provide considerable support for the two-sided view

It is important to note that the two-sided view on theories is in perfect agreement with the hypothetical-deductive (HD) and deductive-nomological (DN) views on prediction and explanation, respectively. For the prediction and explanation of an event we start with representing (modelling) the situation in the relevant terms, as far as possible, but without the crucial event. This is derived by applying the relevant postulates, which amounts to completing or closing the model as far as required. Hence, prediction and explanation naturally appear as co-productions of (partial) models and postulates, that is, of representation and exclusion.

The generalized theory of nomic truth approximation technically reduces to that of Kuipers (2000) by assuming maximality of theories throughout. The generalization is conceptually strongly stimulated by a paper of Gustavo Cevolani, Vincenzo Crupi and Roberto Festa (2011). Though not dealing with *nomic* truth approximation, it inspired me how to construe nomic truth approximation in terms of increasing truth-content and decreasing falsity-content by adding or revising models and/or postulates in the light of increasing evidence.

The generalized theory follows Kuipers (2000) in reconstructing evidence as a special kind of theories, viz. ‘data-theories’, based on realized, hence nomic, possibilities and inductive generalizations (implying induced *impossibilities*). The evidence will guide the comparative assessment of the successes and failures of theories and the subsequent planning of new experiments to be performed leading to new evidence. Ultimately, the comparative success assessment may give good reasons not only for the inductive conclusion that empirical progress has been made by a revised version of the theory relative to the original but even for the abductive conclusion that it is closer to the truth than the original and hence that nomic truth approximation has been achieved.

The presentation will focus on a systematic presentation of the qualitative and quantitative notions for both sides, starting with the notions related to *one* theory, notably, their claims, their truth- and falsity-content, and their accepted and rejected content on the basis of empirical evidence. Then the main comparative notions are presented, viz. ‘closer to the truth’ and ‘more successful’, where the latter amounts to ‘empirical progress’ if it appears to be sustainable in the light of new experiments. Finally, it will be indicated how and why empirical progress provides abductive support for the conclusion that nomic truth approximation has been achieved.

Cevolani, G., Crupi, V., and Festa, R., (2011), “Verisimilitude and Belief Change for Conjunctive Theories”, *Erkenntnis*, **75.2**, 183-222.

Kuipers, T., (2000), *From Instrumentalism to Constructive Realism*, Springer, Dordrecht.

---

*Thermodynamics vs. Statistical Mechanics: A Matter of Logic*

THOMAS MÜLLER

University of Konstanz

thomas.mueller@uni-konstanz.de

---

**1. Introduction**

It may seem plausible that thermodynamics (TD) should reduce to statistical mechanics (SM). After all, SM refers to a system's fundamental molecular microstate, while TD refers to macroscopic variables only, and there is nothing more to a TD system than its microscopic constituents. However, reduction fails. The main problem, well known since the inception of SM at the end of the 19th century, lies with the Second Law of TD. That law states that the entropy of a closed system (e.g., a gas in an insulated container) never decreases over time. There have been numerous attempts at deriving the resulting temporal asymmetry (the "thermodynamic arrow of time") from the laws of SM, but these laws are temporally symmetric, and all extant derivations of the Second Law import some kind of temporal asymmetry over and above the dynamical laws of the microlevel. TD thus appears to have some content over and above SM. But then again, a gas is a gas; the gas as a TD system doesn't have any extra parts or secret powers that the gas as an SM system is lacking. What is going on?

In this paper we will try to take seriously the idea that a TD system is indeed, at any moment of its existence, nothing over and above an SM system, in the sense that the TD system is identical to the SM system. But we still want to allow for different laws applying to TD systems vs. SM systems. The difficulty in this, we argue, is mostly a logical one, and we will show that there is a formally rigorous way to overcome the difficulty. On the account we are proposing, based on case-intensional logic, sortal predication plays a crucial role. That logical resource allows one to say that the gas as a TD system is a thing of a different sort than the gas as an SM system, and so has different persistence conditions resulting in different laws. The Second Law belongs to the persistence conditions of TD systems but not of SM systems. We thereby gain a new perspective on the famous reversibility objection.

**2. Logical background**

There have been many attempts at tackling intertheory relations such as reduction with formal-logical tools. A leading idea is that laws of the less general theory should be derivable from the laws of the more general theory, perhaps with the help of bridge principles, and perhaps allowing for approximations. We will not look at these issues here, but rather on a fundamental but mostly neglected aspect of reduction, viz., sameness

of subject matter of reducing and reduced theory. Since laws are general statements, they involve quantification, and the quantifiers in both theories thus have to range over the same entities.

Which account of quantification can ensure that TD and SM turn out to be suitably related? What do the quantifiers range over? The subject matter of both theories, we may agree, is physical systems (e.g., gases), and we can also agree that these systems are such that at any moment, a system is fully characterized by its (dynamical) microstate (specifying both position and momentum for each of the system's particles).<sup>1</sup> Now, what are the values of the variables occurring (often implicitly) in the laws of TD or SM?

This is a question about quantified intensional (temporal, modal, or temporal/modal) logic. There are many such logics out there, each with its own merits and drawbacks. For the task at hand, most logics are inadequate, either because they treat variables as rigid designators and thus cannot represent identity in a flexible way (e.g., Kripke, 1963), or because they let identity depend too much on a conversational background (e.g. Lewis, 1968). Rather than critically going over a list of logics, we will proceed by sketching how case-intensional first order logic (CIFOL; Bressan 1972; Belnap and Müller 2014a) offers a transparent and useful account of the sameness of subject matter for TD and SM while allowing for different laws.

In CIFOL, one posits a set of *cases*, which in our application would be a set of times,  $T$ .<sup>2</sup> Then there is an *extensional domain*,  $D$ , which in our application would be the set of microstates. Here is how CIFOL differs from almost all other intensional logics: In CIFOL, any term (proper name, variable, etc.) has both an *intension*, which is a function from cases to extensions, and (accordingly) an *extension* in each case. Identity is extensional; it may be true that  $a = b$  in one case, but not in another. Predication, on the other hand, is intensional: Whether a predicate  $P$  applies to a term  $a$  in some case, may depend on  $a$ 's whole intension, not just on  $a$ 's extension in a case at hand. Quantification employs intensional variables; necessity (with " $\Box\varphi$ " symbolizing "necessarily  $\varphi$ ") is truth in all cases. Accordingly, " $\Diamond\varphi$ " for "possibly  $\varphi$ " means truth in some case.

### 3. TD vs. SM: Different sorts, different laws

Given CIFOL with cases as times and extensions as microstates, here are the main benefits for a discussion of the interrelation of TD and SM:

- The value of a variable, an intension, is a function from times  $T$  to microstates  $D$ . Thereby, a variable specifies a logically possible time course of development of a microsystem. Such time developments can thus be quantified over.

---

<sup>1</sup>Albert (2000, 17) argues that a system's momentary state should specify only particle positions, not momenta. The framework described here can be adapted to accommodate that view.

<sup>2</sup>It would be more adequate (but also more complex) to take cases to combine a temporal and a modal dimension. Formal details for such a combination are provided by Belnap and Müller (2014b).

- Intensional predication makes available sortal predicates  $\Sigma$  (“SM system”) and  $\Theta$  (“TD system”) that can classify intensions as belonging to the respective class of systems. Such classification reflects the persistence conditions of systems in the respective class.<sup>3</sup>

Given the sortal predicates  $\Sigma$  and  $\Theta$ , we can express the basic idea of the fundamentality of SM as follows:

$$(1) \forall x [\Theta x \rightarrow \Sigma x] \wedge \exists y [\Sigma y \wedge \neg \Theta y],$$

any TD system is also an SM system, but not the other way round.

- And we can make sense of contingent identity claims:

$$(2) \exists x \exists y [\Theta x \wedge \Sigma y \wedge \blacklozenge x = y \wedge \neg D x = y],$$

there can be a TD system and an SM system that are identical in some case but not in all cases.

These features make it possible to develop a useful answer to the much-discussed reversibility argument, whose physical background is the following: For any possible microstate  $(\bar{x}, \bar{p})$  there is an equally possible microstate  $(\bar{x}, -\bar{p})$  that describes the motion-reverse of the original state. A system starting in that state would thus show the temporal reverse of the development of the original system. So, a system developing in accord with the Second Law of TD might as well develop in such a way that the Second Law is violated: just reverse all the particles’ motion.

In our logical framework, we can show what this argument proves and what it doesn’t. Let the term  $\alpha$  name a TD system (so that  $\Theta\alpha$ , and thus, by (1),  $\Sigma\alpha$  is true), and let  $z$  be  $\alpha$ ’s intension (a map from times to microstates). Now let  $t \in T$  be the time of the proposed motion reversal. We can compute, from  $z$ , an intension  $zI$  that describes a system  $\alpha I$  as evolving like  $\alpha$  until  $t$  (making true case-relative identity claims,  $\blacklozenge\alpha = \alpha'$ , as in (2)) and then on in the motion-reverse way. As a TD system,  $\alpha$ ’s entropy is non-decreasing; this means that from  $t$  on, the entropy of  $\alpha I$  will not increase. Assuming that the entropy is not constant, we thus have a system  $\alpha'$  that violates the Second Law after  $t$ .

Now the status of  $\Theta$  and  $\Sigma$  as sortal predicates allows us to express, in a formally rigorous way, the idea that while both  $\alpha$  and  $\alpha I$  are SM systems, only  $\alpha$  is also a TD system, whereas  $\alpha'$  is not. What happens at  $t$ —motion reversal—amounts to a possible change for an SM system, but it violates the persistence conditions of TD systems. Thus, motion reversal amounts to the *destruction* of the TD system, it is not a possible change for such a system. In this way, we can make sense of the idea that TD systems are nothing but SM systems, while yet obeying different laws. And we can also avoid mixing the natural-historical question of *why so many SM systems around us are TD systems*, which may point to the relevance of special initial conditions of the

<sup>3</sup>Due to lack of space we do not dwell on the main formal characteristics of sortal predicates in CIFOL, modal constancy and modal separation. Briefly, modal constancy means that a sortal predicate is assumed to stick to a system throughout its existence, and modal separation means that a sortal predicate allows one to separate different individual systems in any case. See Belnap and Müller (2014a) for details.

early universe (the “Past Hypothesis”), with the logical issue of the interrelation of TD and SM. Logically, TD cannot be reduced to SM, but this does not imply any physical or metaphysical extravagancies.

### References

- Albert, D. Z. (2000). *Time and chance*. Harvard University Press, Cambridge, MA.
- Belnap, N. and Müller, T. (2014a). CIFOL: Case-intensional first order logic. (I) Toward a theory of sorts. *Journal of philosophical logic*, 43:393–437. DOI = 10.1007/s10992-012-9267-x.
- Belnap, N. and Müller, T. (2014b). BH-CIFOL: Case-intensional first order logic. (II) Branching histories. *Journal of philosophical logic*, 43:835–866. DOI = 10.1007/s10992-013-9292-4.
- Bressan, A. (1972). *A general interpreted modal calculus*. Yale University Press, New Haven, CT.
- Kripke, S. (1963). Semantical considerations on modal logic. *Acta philosophica Fennica*, 16:83–94.
- Lewis, D. K. (1968). Counterpart theory and quantified modal logic. *Journal of philosophy*, 65(5):113–126.

---

*On a Rationale for Cognitive Values*

GERTRUDE HIRSCH HADORN

ETH Zurich

hirsch@env.ethz.ch

---

In the recent debate on the role of values in the assessment of scientific theories, models or hypotheses, cognitive values are distinguished from epistemic values. Typically, epistemic values are taken as the core criteria to judge whether it is justified to assume that a theory, model or hypothesis is true (or approximately true, or empirically adequate, or reliable).

This traditional conception of epistemic values includes criteria such as empirical accuracy and logical consistency. Cognitive values include criteria such as simplicity, broad scope, and explanatory power. There is less agreement on the functions of cognitive values as well as on their relation to the empirical and logical criteria. Recent proposals in line with the traditional conception of epistemic values by Steel (2010), Douglas (2013) and Elliot and McKaughan (2014) come up with different suggestions. Firstly, I show that these proposals highlight important aspects, but do not provide an adequate basis to systematize theory appraisal in science. Secondly, I propose to conceive cognitive values as standards or criteria for assessing the relevance of theories, models or hypotheses for a given problem or purpose. Assessing relevance is seen as complementary to assessing evidence. Taking climate science as an example, I discuss how to specify and weight standards for evidence and for relevance in the case of understanding the climate system on the one hand, and of predicting regional climate on the other hand.

Steel (2010) takes epistemic values as criteria that promote the attainment of truth. He distinguishes between intrinsic and extrinsic epistemic values in terms of how they play this role: intrinsic epistemic values are constitutive (or necessary) for true statements, extrinsic ones are only helpful in determining the truth value. Steel takes cognitive values as extrinsic epistemic values, arguing that, e.g., broadening the scope of a theory may help minimizing error. This proposal does not explain (i) why scientists sometimes rightly trade performance on empirical accuracy for performance on cognitive values and (ii) why not any claim that is warranted according to its performance regarding empirical accuracy and logical consistency counts as a theoretical claim in science. I conclude, that evidence alone is inappropriate as a systematic basis to conceive of values in the sense of criteria or standards for theory assessment in science.

Douglas (2013) proposes a justification of values in theory appraisal by distinguishing necessary values, which she calls “minimal epistemic values”, from optional values, which she calls “ideal desiderata”.

Furthermore, she distinguishes between applying minimal epistemic values and ideal desiderata to theories per se or to theory- evidence relationships. Minimal epistemic values like empirical accuracy and internal consistency must be instantiated to a sufficient degree. Ideal desiderata include scope, simplicity and explanatory power. Ideal desiderata applied to theories per se, Douglas terms “cognitive values” since she takes them as an optional cognitive aid in the search for potential flaws or potential evidence for theory, just like Steel’s extrinsic epistemic values. However, if applied to theory-evidence relationships, she terms these values “epistemic assuring values”, since she assumes that in this case ideal desiderata provide genuine epistemic import. She argues that we have the more reason to assume that a theory is true or reliable, the broader its scope is or the more explanatory power the theory has, given the supporting evidence. However, this conception builds on the questionable assumption that these properties have a systematic relation to the truth-value of a theory (Laudan 2004).

Furthermore, the conception of epistemic assuring values does not set the priorities right. Instead of taking criteria like simplicity to qualify the inference from evidence to epistemic status, I suggest to conceive them as properties, which we want theories qua scientific theories to instantiate to some degree, while empirical evidence is required to warrant theoretical claims having such properties. So, from my perspective, also Douglas’ account suffers from taking evidence as the only systematic basis for conceiving criteria for theory assessment in science.

Elliot and McKaughan (2014) conceive of theories, models and hypotheses as scientific representations. These representations can be evaluated along two dimensions, which they phrase “fit with the world” and “fit with the needs of their users” (Elliot and McKaughan 2014, 1). While they stick with the traditional conception of epistemic values, they go a step further than Douglas by conceiving of the role of cognitive values not as a subsidiary one in relation to epistemic values, but as a dimension of evaluation on its own. This is an important improvement. However, by restricting considerations on cognitive values to practical use, such as in policy-related research, they are silent about the role of criteria such as simplicity or explanatory power in basic research. This lack is due to taking “pragmatic” in the sense of practical use. To provide an appropriate systematic basis for conceiving standards in theory appraisal, I suggest to build on Suárez’ (2004) inferential conception of scientific representation. In accordance with his distinction between two aspects of (mis-)representation, I distinguish between standards for relevance and standards for evidence to assess goodness of theory in representing an intended target for a given problem or purpose. Taking climate science as an example, I discuss differences in specifying and weighting standards for evidence and for relevance in the case of understanding the climate system on the one hand, and predicting regional climate on the other hand (Knutti 2008, Parker 2014).

## References

- Douglas, Heather E. 2013. “The Value of Cognitive Value.” *Philosophy of Science* 80:796–806.  
Elliott, Kevin C., and Daniel J. McKaughan. 2014. “Non-epistemic Values and the Multiple Goals of Science.”



Philosophy of Science 81:1-21.

Knutti, Reto. 2008. "Should We Believe Model Predictions of Future Climate Change?" *Philosophical Transactions of the Royal Society A* 366:4647–64.

Laudan, Larry. 2004. "The Epistemic, the Cognitive, and the Social." In Machamer Peter, and Gereon Wolters, eds. *Science, Values, and Objectivity*. Pittsburgh/Konstanz: University of Pittsburgh Press/Universitätsverlag Konstanz: 14–23.

Parker, Wendy S. 2014. "Simulation and Understanding in the Study of Weather and Climate." *Perspectives on Science* 22:336–356.

Steel, Daniel. 2010. "Epistemic Values and the Argument from Inductive Risk." *Philosophy of Science* 77:14–34.

Suárez, Mauricio. 2004. "An Inferential Conception of Representation." *Philosophy of Science* 71: 767-779.

---

*The Structure of Science: From Diachronic and Synchronic Accounts*

HANNE ANDERSEN

Aarhus University

hanne.andersen@ivs.au.dk

---

The historical development of disciplines or specialties, as well as their subdivision into fields or domains, was a major topic of interest within the historically inclined philosophy of science that flourished from the 1960s and some decades onwards. Philosophers of science such as Kuhn, Lakatos, Laudan, Toulmin, and Shapere described the development of science by focusing on the development of individual areas of science, and the development of these areas were then described in terms of, for example, paradigm-induced normal science and paradigm changing revolutions (Kuhn, 1970), progressing and degenerating research programs (Lakatos, 1971), successive research traditions (Laudan, 1977), or domains connected through history by chains-of-reasoning (Shapere, 1977). On these traditional accounts of how individual areas within science developed over time, a scientific discipline (or specialty, field or domain) could be understood at the same time as an epistemic unit consisting of a set of closely related cognitive resources such as, for example, concepts, models and theories, and as a social unit consisting of highly similar experts who were employing and at the same time developing their shared cognitive resources. On these models, each of the practitioners within a given specialty could be seen as epistemically autonomous agents who were each able in similar ways to recognize the same, potential new research puzzles that could be solved in ways similar to previously recognized puzzle. Hence, a specialty or discipline could be characterized by this close, bipartite relation between the scientific community and the cognitive resources that members of this community employed, while the individual scientists and the research activities that they engaged in were tokens of the types of similar community members working on similar problems.

However, over the last century, science has grown increasingly collaborative, and most scientific knowledge today is produced by groups in which multiple scientists collaborate in order to combine their knowledge, manpower, materials and other resources (e.g. Beaver & Rosen, 1978; 1979; 1979; Beaver, 2001; Godin & Gingras, 2000; Wuchty, Jones, & Uzzi, 2007). Further, much scientific research today cut across disciplinary boundaries in addressing complex problems that reach beyond what can be solved within the individual disciplines (Braun & Schubert, 2003; Porter & Rafols, 2009). As a consequence, the relation between social community and cognitive resources is much more complicated than assumed by the philosophers of science working on the diachronic development of science half a century ago.

Hence, instead of focusing on the historical development of cognitive resources and the social community in a close bipartite relation, while reducing the individual scientists and the activities that they engage in

to tokens of disciplinary types, the analysis presented here turns the picture upside down and focuses instead on the individual research activity as it is spanned synchronically by the dimensions of cognitive resources and epistemic relations between the scientists employing them. On this background, the analysis presents a continuous, two-dimensional spectrum of research activities from which four ideal types can be described. The account takes this synchronic structure of research activities as its central focus. Based on analyses of the cognitive resources employed in individual research activities it is examined how they relate to domains in a historical process, and how their distribution among the researchers involved gives rise to relations of epistemic dependence. For the sake of analytical clarity I shall first examine how the cognitive resources employed in a research activity relate to domains and how to understand the individual's expertise on such a picture. Next, I shall examine the epistemic dependence between scientists, and combining the analyses of cognitive resources and epistemic dependence I shall provide a renewed view of how to understand disciplines and specialties in terms of different ideal types of research activities in a two-dimensional spectrum.

Finally, I shall show how current understandings of accountability and quality control in science are rooted in a disciplinary ideal that is untenable for collaborative and interdisciplinary research. Analyzing values that drive science in various directions I shall further sketch the main mechanisms underlying the perceived tension between disciplines and interdisciplinarity and argue for a redefinition of accountability and quality control for interdisciplinary and collaborative science.

---

*Measuring the Unmeasurable. Engineering, Mathematics, and the Computer: A New Mixture*

HANS HASSE

University of Kaiserslautern

hans.hasse@mv.uni-kl.de

JOHANNES LENHARD

Bielefeld University

johannes.lenhard@uni-bielefeld.de

---

Measurement is surely a central part of scientific activity. It quantifies aspects of the world and hence prepares them for scientific, in particular mathematical, treatment. At the same time, measurement somehow makes contact to the world, thereby ensuring a degree of reality. There exist very different and controversial opinions in the philosophy of science about the role of measurement. Think of Bridgeman's operationalism that intended to guarantee meaningfulness of scientific terms by defining properties via measurement practices (cf. Chang 2009 for a balanced picture). Or consider the much-debated "theory-ladenness" that would see measurement practices as theoretical and therefore indirect.

There is an important new twist in the philosophy of science that now considers the changing nature of measurement practices themselves. This new movement highlights the role of simulation modeling in measurement. An illustrative sample includes Morrison (2009), who diagnoses a "changing face of experimentation" when simulation modeling is intertwined with measurement. Parker (2015) discusses how measuring practices involve models and calculations when data sets that count as observational are also simulation-based. Carusi et al. (2012) maintain the crucial role for empirical contact, but show that this contact might be based on iterative relations where models are developed. Finally, Tal (2013) suggests a model-based epistemology of measurement.

The present paper aims at contributing to this discussion. We want to focus on engineering, in particular on chemical process engineering. There, the quantitative values of parameters are directly relevant for the construction of machinery. Hence the investigation of measurement practices happens against the background of actual usage in technology.

Engineering sometimes has to work with quantities that are neither empirically measurable nor can they be determined by theory. We claim that such quantities are determined by measurement practices that mix and intertwine mathematics, simulation modeling, and experiments. We will illustrate this claim by discussing an example from chemical process engineering, namely the design of a distillation column for producing certain materials.

While our claim is in good agreement with the standpoints mentioned above, it does not merely give a further instance, but also offers something in addition. Our emphasis is on how mathematics is used as a tool. Since at least Descartes, mathematics counts as the link that connects quantifying measurement with

systematic scientific treatment. In this line of thinking, however, mathematics deals with fundamental structures. We want to investigate cases that differ from this line in a number of aspects.

We will present a scheme of simulation modeling that highlights a feedback loop of model adaptation. This loop links model parameters to measurements and the most interesting case is when such measurements are not possible. This can occur for a number of reasons, especially when the quantities in question are not accessible with existing technology. However, the simulation models then have to be modified, which gives parameterization and tuning a particular significance. We will study a comparatively simple example from process engineering: the evaporation of an ethanol-water mixture. Modeling this involves both physical laws and models which come in the form of physical laws but are mere correlation tools. Such models often include parameters without physical meaning. Interestingly, they nevertheless often have names which sound physical – maybe to conceal their empirical character. They are tuned by fitting to results of empirical measurement. Such approaches, we will argue, have to find a non-trivial balance between theory and curve fitting – the latter alone would likely restrict the model to a very narrow range of application.

In a second step, we will discuss how sub-models of vaporization and other aspects are coupled together for building a (global) model of the process that happens in a distillation column. The point is that the global model of the process consists of sub-models that might be inconsistent, but show an adequate fit in relevant quantities.

The theoretical foundation of such global model arguably is poor, if not defective. Therefore, the explanatory potential is restricted. However, this is not a devastating finding! As long as the predictive virtues are satisfying, the global model remains satisfying and is used in design. This is not plain trial-and-error, but validation remains important. Validation or confirmation of the model do not happen by a fundamental analysis, rather via sensitivity studies, in which a model has to show sufficient agreement with known data. Overall, parameterization and tuning play an essential role in making accessible quantities that cannot be empirically measured. Mathematics then does not grant consistency, but helps working with inconsistent sub-models. In such circumstances, predictive performance plays a pivotal role at the cost of explanatory capacity.

## References

- Carusi, A., Burrage, K., Rodríguez, B. (2012). "Bridging Experiments, Models and Simulations: An integrative approach to validation in computational cardiac electrophysiology." *American Journal of Physiology: Heart and Circulatory Physiology* 303(2), H144-55.
- Chang, H. (2009). "Operationalism", *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/fall2009/entries/operationalism/>>.
- Morrison, M. (2009). "Models, Measurement, and computer simulation. The Changing Face of Experimentation." *Philosophical Studies* 143, 33–57.
- Parker, W. S. (2015). "Computer Simulation, Measurement, and Data Assimilation." *British Journal for the Philosophy of Science*, forthcoming.

Tal, Eran (2013). "Old and New Problems in Philosophy of Measurement." *Philosophy Compass* 8/12, 1159–1173.

---

*Measuring Unification*

IOANNIS VOTSIS

University of Duesseldorf

votsis@phil-fak.uni-duesseldorf.de

---

Scientists tend to opt for simpler and more unified hypotheses. Such considerations, though aesthetically pleasing, are often viewed as at best pragmatic in matters of choosing between rival hypotheses. In this talk, I put forth a novel conception and an associated measure of unification, both of which are demonstrably more than just pragmatic considerations. The discussion commences with a brief survey of some failed attempts to conceptualise unification. It then proceeds to an analysis of the notions of confirmational connectedness and disconnectedness, which are essential ingredients in the proposed conception of unification and its associated measure. Roughly speaking, the notions attempt to capture the way support flows or fails to flow between the content parts of a hypothesis. The more the content of a hypothesis is confirmationally connected, the more that content is unified. Since the confirmational connectedness of two content parts is determined by purely objective matters of fact, the proposed notion and measure of unification are themselves objective, i.e. not merely pragmatic considerations.

Attempts to devise a satisfactory conception of unification abound. One of the earliest is Friedman (1974) where it is argued that understanding is generated when we trim down the number of independently acceptable law-like assumptions that feature as *explanantia* in the derivation of an *explanandum*. The lower that number the more unified an explanation. Friedman's account was in great part motivated by a desire to avoid trivial explanations. It had already been observed that deriving an *explanandum* from a set of premises is not sufficient to turn those premises into a genuine explanation. Friedman sought to avoid this problem by limiting the derivations that yield genuine explanations to those that unify phenomena. Though highly influential, his account soon faced a number of insurmountable difficulties. As Kitcher (1976) and others pointed out, Friedman's account rules out trivial explanations only at the expense of also ruling out some genuine ones. Several other attempts at conceptualising unification have been made with similar problems. They include Forster (1988), Kitcher (1989), Schurz and Lambert (1994) and Thagard (1993).

While it ultimately fails, Friedman's account does at least get one fundamental thing right. By emphasising the role of the acceptability of law-like assumptions his account places a premium on the link between unification and confirmation. The proposal in this talk is in tune with this appraisal and indeed elevates the link with confirmation to the single most important ingredient in our quest to understand unification.

According to this account, unification is to be understood as a measure of confirmational connectedness. But what is confirmational connectedness and its opposite confirmational disconnectedness?

Roughly speaking, the notions attempt to capture the way support flows or fails to flow between the content parts of a hypothesis. The more the content of a hypothesis is confirmationally connected, i.e. support flows between its content parts, the more that content is unified. Let us use 'x F<sub>r</sub> y' to denote that y is a *relevant deductive consequence* of x. In formal terms, confirmational connectedness can be articulated thus:

Any two content parts of a non-self-contradictory proposition Γ expressed as propositions A, B are *confirmationally connected* if, and only if, for some pair of internally and externally non-superfluous propositions α, β where A F<sub>r</sub> α and B F<sub>r</sub> β: either (1) where 0 < P(α), P(β) < 1, P(α/β) ≠ P(α) or (2) there is at least one true or partly true atomic proposition δ such that α ∧ β F<sub>r</sub> δ, α ≠<sub>r</sub> δ and β ≠<sub>r</sub> δ.

An explication of the notions in the *analysandum* cannot be pursued here due to obvious limitations of space. Suffice it to say that the probabilities are meant to be objective. That is, probability statements indicate true relative frequencies and/or true propensities of things happening like events, states---of---affairs or property instantiations. An objective interpretation of the probabilities captures the intuition that the confirmational (dis---/)connectedness of the content of a hypothesis is determined by facts about the world, i.e. it is not a subjective matter.<sup>1</sup>

We are now ready to express the unification u of a proposition Δ with the following function:

$$u(\Delta) = 1 - \frac{\sum_{i=1}^n d_i^{\alpha, \beta}}{\sum_{i=1}^n t_i^{\alpha, \beta}}$$

where d<sub>i</sub><sup>α,β</sup> denotes the number of disconnected pairs α, β in a given content distribution i, t<sub>i</sub><sup>α,β</sup> denotes the total number of connected plus disconnected pairs α, β in a given distribution i and n denotes the total number of content distributions. To determine the number of disconnected pairs in a given content distribution we count how many times a *different* pair of relevant deductive consequences α, β fails to satisfy either clause (1) or (2). Any pair that is not disconnected is counted as connected. The higher the value of u(Δ) the more unified the content of Δ.

The objectivity of the unification measure arises from the fact that its output is entirely dependent on the notions of confirmational (dis-/)connectedness. And whether these notions are satisfied for any given pair of propositions is a matter wholly determined by the way the world is like, i.e. the world decides what is physically connected to what. In short, unification, under the proposed measure, is not in the eye of the beholder.

**References**

---

<sup>1</sup>I ask the reader to momentarily put aside worries about our epistemic access to objective probabilities, though these worries will be addressed in the talk.



- Forster, M. (1988) 'Forster Unification, Explanation and the Composition of Causes in Newtonian Mechanics', *Studies in History and Philosophy of Science*, vol. 19: 55-101.
- Friedman, M. (1974) Explanation and Scientific Understanding. *Journal of Philosophy* 71(1): 5-19.
- Kitcher, P. (1976) Explanation, Conjunction, and Unification. *Journal of Philosophy* 73(8): 207-212.
- Kitcher, P. (1989) 'Explanatory Unification and the Causal Structure of the World' in *Scientific Explanation: Minnesota Studies in the Philosophy of Science*, P. Kitcher and W. Salmon (eds.), Minneapolis: University of Minnesota Press, pp. 410–505.
- Schurz, G. and K. Lambert (1994) 'Outline of a Theory of Scientific Understanding', *Synthese*, vol. 101: 65-120.
- Thagard, P. (1993) *Computational Philosophy of Science*, Cambridge, MA: MIT Press.

**Historical, Social and Cultural Studies in****Contributed Papers****Philosophy of Science III**

Chair: Francesca Biagioli

Room 5E, Friday 15:30 – 17:30

---

*The Argument from the Good Lot: Unconceived Alternatives and 19th Century Bacteriology*

RAPHAEL SCHOLL

Center for Philosophy of Science, University of Pittsburgh

raphael.scholl@gmail.com

---

Laudan's pessimistic induction rests on a brute fact about the history of science: even highly successful scientific theories have regularly turned out to be mistaken. Recently,

P. Kyle Stanford has argued that the "problem of unconceived alternatives" (PUA) can explain this brute fact and thus show us *why* the pessimistic induction holds. On Stanford's view, many scientific inferences are eliminative: scientists begin by corraling a number of candidate hypotheses that may explain a phenomenon of interest, proceed to rank these hypotheses according to some standard such as explanatory power, and then accept the most successful candidate. The problem is that even if ranking is reliable, eliminative inference only leads to truth if the true hypothesis is among the candidates to begin with. In van Fraassen's words, inferences to the best explanation may well choose the best of a "bad lot". In a series of detailed case studies, Stanford gives historical support to the problem. He shows that past scientists regularly failed even to conceive of powerful candidate hypotheses which we have since come to accept. What is more, past scientists tended to overestimate their ability to exhaust the space of likely or plausible hypotheses.

However, Stanford's PUA does not apply to an important and widespread type of scientific inference: experiments to determine causal relevance. Causal inferences commonly proceed from an exhaustive hypothesis space defined by the contradictories "C is a cause of E" and "C is not a cause of E". Such inferences from "good lots" leave no room for unconceived alternatives in Stanford's sense — aspirin either does or does not cause the relief of headaches. Yet the stability of some causal claims does not imply a static view of science. Causal inferences involve well known methodological issues (such as confounding) that may give rise to uncertainty and even controversy. Moreover, individual causal claims are limited and require the investigation of causal co-factors (why is aspirin ineffective in some patients?), intermediate steps (how does aspirin relieve headaches?), and alternative causes (what else relieves headaches?).

A contrast in 19th century bacteriology will serve to illustrate and test the claim that biological practice is shaped by inferences from good lots. Robert Koch (1843–1910) investigated a series of infectious diseases including anthrax and cholera. In the case of anthrax, he was able to establish the life cycle of *B. anthracis* within a few years: a series of experiments — causal inferences proceeding from good lots — demonstrated

the causal factors required for the bacterium to grow, to form spores, to reactivate, and to infect livestock and humans. Even though Koch became involved in a controversy regarding anthrax with Pasteur, the causal claims established experimentally were not, in general, at issue, and they remain accepted to this day. By contrast, Koch's investigation of the cause of cholera was much less successful. Even though his expedition to Egypt and India in the early 1880s identified *V. cholerae*, Koch was unable to demonstrate that the bacterium is the cause of cholera, mainly for lack of an animal model. Koch offered explanatory considerations favoring his causal hypothesis, but — just as Stanford's PUA suggests — opponents such as Pasteur and von Pettenkofer did not find this to be compelling evidence. In the absence of evidence from an experimental causal inference proceeding from a good lot, Koch's claim remained controversial for over a decade. A careful study of the historical sources demonstrates that the historical actors were sensitive to just the methodological nuances sketched above.

In summary, the historical record suggests that a principled distinction is possible between (1) inferences that are vulnerable to the PUA (eliminative inferences potentially proceeding from a bad lot) and (2) inferences that are not vulnerable to the PUA (causal inferences proceeding from a good lot). This opens the possibility of selective realism about biological claims, and especially for *prospective* realism: Scientists such as Koch and his opponents were able to identify in advance which of their claims were reliable.

---

*Debating Causation in the Life Sciences: A Systems Perspective on Causes and Effects of Cancer*

SARA GREEN

Center for Philosophy of Science, University of Pittsburgh

saraehrenreichgreen@gmail.com

---

This paper reexamines philosophical accounts of causal explanation in light of ongoing debates in the life sciences regarding the role of genomics for understanding complex diseases. Many life scientists see genomics as a powerful tool for biomedical research, opening for the possibility to develop patient-specific models for disease prevention. Others argue that making sense of the effects of gene mutations requires a better understanding of - and contextualization within - the higher-level dynamics and organization of tissues or even whole organisms. I highlight a set of philosophical implications underlying this debate by drawing on a contemporary controversy on the cause of cancer. The case illustrates important ramifications of a systems perspective for our understanding of causation and causal explanation in the life sciences. In particular I discuss the implications of non-sequential and global causal effects for our conceptualization and discovery of causes through experimental interventions and mechanistic schemes.

A hotly contested issue is whether cancer is a cell-based disease or results from failure of tissue organization. Conflicting views on whether the best approach is a pathway-centered or systems approach have intriguing theoretical and practical implications. Some researchers have suggested that if cancer is a tissue-based disease, mutations may be a result of cancer, rather than the cause. These accounts differ with respect to the delineation and nature of the phenomenon to be explained, and of the relevant scale to approach the causal analysis. At the most fundamental level, it is debated whether specific 'cancer cells' exist, given a set of difficulties in distinguishing between malignant and benign tumors (the latter being thought of as a negative control). That is, it is a controversial issue whether causes of cancer can be fruitfully studied at the molecular scale, or whether we need a theory of tissue organization. Among the practical implications are issues pertaining to the design of relevant experiments, models and diagrams, and about what researchers take to be the most promising strategies for cancer treatment.

Settling this debate is challenging, given that there is some experimental evidence in support of both of these positions, and given that interventions on target genes or molecules may influence hundreds or thousands of other interactions in the networks that these are embedded in. A further complication is that these networks may change over time such that interventions may restructure the dynamics for gene expression and cell signaling. Systems biology research has recently gained insight to the multivariate nature of regulatory and signaling networks. In the context of cancer research, the effect of some mutations on tumor development has been shown to depend on the state of the network as a whole, resulting in

apparently conflicting experimental results. While there are strong correlations between mutations and tumor development, some mutations are at the same time associated with gene products that advance or suppress carcinogenesis, depending on the context of the network or tissue. Such results question the assumption of a specific causal identity of system components associated with the static interpretation of mutations where properties translate directly into cause-specific effects. These non-linear effects are, however, not taken to be arbitrary but mediated by dynamic interfaces between biological processes at different scales. Systems-oriented researchers hope that an identification of the constraints on stable dynamic states of biological networks as a whole can help understand and predict the nonlinear effects of lower-scale perturbations on phenotypic states such as cell types. In this paper I examine a framework proposed as a solution to this dilemma (see Creixell et al. 2012; Huang 2009) and relate this perspective to existing interventionist accounts of causation and mechanistic approaches to causal explanation in the life sciences (Woodward 2003, 2013).

Insights to complexity and strategies for dealing with the challenges should, in my view, reflect back on the concept of causation in philosophy of science. A systems-oriented view argues that some diseases and disorders are better characterized (and treated) as altered perturbed network states in state space. From this perspective, cancer is to be understood as an attractor within the possible states of the dynamics of regulatory or signaling networks. This framework has recently been argued to unify the conflicting conceptions of cancer through the notion of mutual causation (Huang 2014). The idea is that mutually dependent factors co-evolve in a spiral-like fashion over time. In this framework, mutations can introduce changes to the organization of the network as a whole, making other network states accessible. Reshaping of the landscape topology may, in turn, lead to genetic instability and further mutations. This perspective shifts the focus from properties of specific molecular causal activities to higher-level dynamics of the network as a whole. It is therefore relevant to examine the compatibility of this framework with the traditional understanding of complex diseases as diseased or broken mechanisms that can be uncovered through ideal molecular interventions. The case of cancer research illustrates a broader and pervasive problem in the life sciences of coming to grips with two-way degeneracy and plasticity of biological systems. How can the same set of genetic perturbations give rise to a variety of discrete cell phenotypes, and how can different genetic instructions give rise to the same phenotypic states? Why are certain cell states, including cancer states, highly robust to perturbations? To address such questions, I argue that we are forced to go beyond the linear pathway view of causation and beyond the common understanding of complex disease in terms of diseased or broken mechanisms.

## References

- Creixell, P., Schoof, E. M., Erler, J. T., & Linding, R. (2012). Navigating cancer network attractors for tumor-specific therapy. *Nature Biotechnology*, 30(9), 842-848.
- Huang, S., Ernberg, I., & Kauffman, S. (2009). Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in Cell & Developmental Biology*, 20(7), 869-

876.

Huang, S. (2014). When correlation and causation coincide. *BioEssays*, 36(1), 1-2.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2013). II—Mechanistic explanation: Its scope and limits. *Aristotelian Society Supplementary Volume*, 87(1) 39-65.

---

*Metaphysics Naturalized? The Case of Classification in the Sciences*

THOMAS REYDON

Leibniz Universität Hannover

reydon@ww.uni-hannover.de

---

While the topic of kinds and classification has long been a focus of work in the philosophy of science, as well as other areas of philosophy, a generally accepted account is still lacking. Moreover, there is no general agreement about the kind of account that is being searched for, or the criteria that such an account should meet. In part, this is due to the persistent problems confronting the search for a good theory of natural kinds, as many of the kinds found in the sciences are usually thought of as paradigmatic natural kinds. In response to these problems in the philosophy of science (though not in other areas of philosophy) there is a trend to move away from the metaphysics of kinds and classification and turn to epistemological issues, as well as questions regarding practical uses of kinds and classifications in various contexts. This trend fits well with the renewed interest in doing naturalistic philosophy of science and the increasing calls for bringing philosophy of science closer to scientific practice.

However, abandoning the search for a metaphysics of kinds and classifications is too quick. A metaphysical account of kinds is a crucial element of the explanation of why some kinds and classifications are used in the sciences with more success than others, and why some ways of grouping things turn out not to be useful at all. After all, barring cases of epistemic luck the reason for the epistemic and practical success of kinds and classifications must be that they adequately connect to some or other feature of the world.

While an account of kinds and classifications thus needs to encompass metaphysical elements, it is not clear whether naturalistic and practice-oriented philosophy of science is at all able to provide such elements. In this respect, naturalistic philosophers of science face two problems. First, metaphysics cannot be read off from either epistemology or practice: simply examining scientific kinds and classifications and the ways in which investigators in the various areas of science employ them will not reveal their metaphysical underpinnings. Second, once the metaphysics is elucidated for individual cases, these different metaphysical pictures need to be unified into an overarching account of kinds and classifications – an issue which some authors hold to be insurmountable (cf., Dupré, 1993).

Thus it seems that at least *some* a priori considerations should be allowed to enter into the picture to guide the metaphysical analyses of individual cases. But as a priori metaphysics is suspect from a naturalistic viewpoint, the challenge for a naturalistic and practice-oriented metaphysics of kinds and classification is to bring a priori considerations into play without rendering the account unacceptably non-naturalistic. In this talk I address this challenge and explore what a thoroughly naturalistic metaphysics of scientific kinds

and classifications could look like. Two core notions in the debate on kinds and classification are important in this context, namely the notions of naturalness and of normativity.

The notion of naturalness obviously occupies center stage in the search for a theory of kinds and classification in the sciences. While in both scientific and everyday contexts various kinds of kinds are used as the basis of epistemic as well as non-epistemic practices (including inference, explanation, mechanistic/functional analysis and decomposition, information storage and retrieval, heuristics, etc.), the kinds used in the sciences are often thought to occupy a special position among these kinds. But the question what makes a particular kind special – usually translated into the question what exactly the naturalness of a kind consists in – remains unclear. Proposals range from having essences, featuring in laws of nature, or being causally supported to merely being kinds of naturally occurring entities, but no proposal has been able to win the day. In part this is due to the normative requirements on philosophical theories of scientific kinds and classifications – or so I will argue.

Theories of scientific kinds and classifications should not only explain the various epistemic and non-epistemic uses of natural kinds, such as what makes a kind suitable as a basis for inferences. In addition, they should provide normative criteria to distinguish “good” (or: natural) kinds from other kinds of kinds. However, a problem for extant proposals regarding what it is that makes a kind natural is that to the extent to which they encompass normative criteria for identifying natural kinds, they selectively pick out a subset of epistemically successful kinds but fail to cover all such kinds (e.g., [blinded for review]). Conversely, accounts that cover all epistemically successful kinds tend to apply to epistemically less successful and unsuccessful kinds too. The solution to this dilemma that I want to propose draws inspiration from Nelson Goodman’s work on the problem of induction. According to Goodman (1983), the philosophical account of induction (i.e., the rule(s) of induction) and actual inductive practices mutually support and constrain one another. In one direction, the account of induction provides criteria with which good inductions can be distinguished from unacceptable ones. In the other direction, inductive practices support the philosophical account of induction in the sense that the justification of the account consists in its adequately describing widely accepted inductive practices. Goodman considers this mutual dependence of philosophical theory and actual practice to constitute a virtuous circle – a sort of mutual fine-tuning of the theory and the practices to which it is supposed to apply until a good fit is achieved.

Arguing on the basis of an example from the life sciences, I contend that Goodman’s scheme provides a general way of developing philosophical accounts in close connection with actual practice that can be applied to the development of a metaphysics of scientific kinds and classification. In particular, any a priori elements of such a metaphysics are rendered unproblematic from a naturalistic perspective, as they remain open for revision if the examination of particular cases demands this.



**References**

Dupré, J. (1993): *The Disorder of Things: Metaphysical Foundations for the Disunity of Science*, Cambridge (MA): Harvard University Press.

Goodman, N. (1983): *Fact, Fiction, and Forecast (Fourth Edition)*, Cambridge (MA): Harvard University Press.

---

*Philosophy in Unified Science: The Bipartite Metatheory Conception* (cancelled)

THOMAS UEBEL

University of Manchester

thomas.uebel@manchester.ac.uk

---

This talk is dedicated to detailing an aspect of the work of the Vienna Circle that has received surprising little attention: its philosophical metatheory (or philosophy of philosophy). The aim is to articulate and defend a conception of the nature of philosophy of science that, I claim, can be found among the members of the so-called left wing of the Vienna Circle (Carnap, Frank, Neurath — Hahn died too early to see much of its development). This conception does not only contradict common ideas of how the Vienna Circle thought about philosophy but also offers a useful intervention in current debates about, e.g., the standing of formal epistemology vis-à-vis various forms of naturalism.

Needless to say, in the Circle this issue was not discussed under the heading I give it, but under that of the relative place of philosophy and unified science. What is well-known is that the Vienna Circle abjured metaphysics and sought to develop an understanding of scientific reason free of questionable assumptions arrogated as a priori insights, though it has long been debated whether in fact it succeeded or could possibly succeed in that. More recently it has become recognized that leading members of the Circle put forward competing conceptions of their enterprise. Schlick championed a Wittgenstein-derived approach to philosophy as meaning-determination, Carnap turned philosophy into a purely formal “logic of science”, while Neurath rejected the idea of philosophy altogether seemingly submerging all in the projected unified science. The question arises how the apparent incompatibilities can be contained that threaten the viability of the Circle’s unity by calling into question its cohesion from the inside.

The position I will articulate and develop does recognize sharp differences but also offers reconciliation. According to it philosophy loses its autonomy vis-a-vis the sciences but is neither abolished in favour of pure science nor rendered into a purely formal-logical inquiry. Rather, philosophy of science is regarded as a scientific inquiry of the second-order, a reflective inquiry into scientific reason that must do without epistemic resources peculiar to itself but is able to help itself to all the tools, techniques and conceptualizations that the empirical and formal sciences make available. Philosophy is integrated into unified science as its metatheory without being reduced to any one or the collection of its object-level disciplines. That this conception rejects the idea of a “first” philosophy in its traditional foundationalist garb or even in its more recent Wittgensteinian guise as pursued by Schlick is obvious, as is its promise to harmonize the approaches of Carnap and Neurath (and Frank). What is not obvious is whether the promise can be cashed in the form of a programme that (i) can reasonably be ascribed to the historical actors, (ii)

consistently realizes their ambitions, and (iii) makes a contribution to contemporary meta- philosophy. I will seek to provide a positive answer to all three questions.

The talk will begin with a basic outline of what I call the “bipartite metatheory conception”. According to it, what is nowadays called philosophy of science comes in two parts, Carnap’s formalist logic of science and what Neurath called the “behavioristics of scholars” or Frank the “pragmatics of science”. The logic of science investigates scientific theories in typically axiomatized form and considers their internal structure and their relation to their evidential base in formal (syntactic or semantic) and logical terms (deductive and inductive). The pragmatics of science investigates scientific practice by means of the empirical sciences of science, the psychology and sociology as well as the history of science. So while, e.g., the logic of science investigates abstract relations of evidential support, the pragmatics of science investigates concrete theory choice and change. Following this outline, the talk will address several objections to the interpretive thesis presented.

Besides charting the emergence of the bipartite metatheory conception in the writings of Carnap, Frank and Neurath in the early 1930s, attention must be paid to the origin of the bipartite metatheory conception in the need to coordinate the different emphases they placed in their own early theories of science on descriptively adequate or idealizing rational reconstructions, reflecting the historical influences of the naturalism of Mach and the formalism of Hilbert. Against that background, the skeptical objection can be countered that Carnap not only did not work in but also had no theoretical interest in empirical studies of science. I will argue that attending closely to what he says on a variety of occasions provides good reasons to resist this skepticism and detect intentional division of labour instead.

Another issue concerns the question, also frequently raised in a sceptical mode, of the actual compatibility of the theorizing of Carnap and Neurath, especially with regard, first, to issues in the reconstruction of scientific theories, in particular with regard to basic observation statements, and, second, what Neurath’s rejection of the semantic conception of truth says about his understanding of the common programme which the bipartite metatheory thesis makes him a part of. Again I will argue that skeptical objections arising from these questions can be overcome. While a close reading of Carnap reveals a flexibility on his part that allows Neurathian variants of observation sentences to be considered physicalistically legitimate, a close reading of Neurath’s objections in as yet unpublished exchanges with Carnap reveals that his admittedly unsatisfactory arguments do not force him into a position incompatible with the bipartite metatheory programme.

Finally, yet another issue concerns the fact that the conception was rarely discussed by the theorists I designate as its proponents and hardly ever was advertized as such. This delicate matter is not simply settled by contextualizations like the one indicated above but also requires discussion of the use of explication in historiographical method. Here Carnap’s concept of explication (so named only in 1950 but operative by then for much longer) will be argued to provide the required basis of convergence.

## Symposia & Contributed Papers VIII

**The Tension between a Naturalistic and  
a Normative Approach to Explanation and Understanding**

**Symposium**

Organizer: Jan Faye

Chair: Jose Diez

Room 5D, Saturday 09:30 – 11:30

---

*An Evolutionary and Cognitive Approach to Understanding*

JAN FAYE

University of Copenhagen

faye@hum.ku.dk

---

*On Scientific Understanding without Explanation*

ANTIGONE M. NOUNOU

University of Athen

amnounou@gmail.com

---

*From Explanation to Understanding: Normativity Lost?*

HENK W. DE REGT

VU University Amsterdam

h.w.de.reg@vu.nl

---

*Normativity and the Inferential Account of Understanding*

PETRI YLIKOSKI

University of Helsinki, Linköping University

petri.ylikoski@helsinki.fi

---

**General Description**

Among philosophers working on scientific explanation, there seems to be a growing consensus that explanation is somehow connected to some form of understanding. This stands in sharp contrast to Hempel's view, according to which understanding is a psychological notion with no constructive bearings on an

acceptable notion of scientific explanation. It is well-known that Hempel argued that the aim of explanation was not to gain understanding but to deduce what was to be explained from a set of premises containing at least one law statement. In his view, the kind of understanding we experience in connection with explanation refers to a psychological state we are brought into whenever we realize that the occurrence of the explained phenomenon is to be expected on the basis of our knowledge of the laws in question and the particular circumstances. According to Hempel the expectation itself, although being a psychological state, should be considered to be a completely rational state of mind as it is grounded in the knowledge of the logic of deduction and the concomitant understanding is caused by a successful deduction. Others, like Michael Friedman, have suggested that it is not expectation but unification which gives us understanding, i.e. the insight that a high-level law unites different low-level laws formerly considered mutually independent. In general, philosophers have associated explanatory understanding with different psychological features such as confidence, expectation, feeling of certainty, or intellectual satisfaction.

Thus, the picture that emerges from the literature of the old days is that explanatory understanding is a concomitant phenomenon, a purely psychological feeling that was caused by the mental operation like successful deduction, unification, fitting the explanation-seeking phenomenon into a general world-picture, or getting to know the inner mechanism of things in the world. And this is a consequence of the fact that the aim of explanation is something different from understanding.

In recent years new proposals for understanding “understanding” and its relation to explanation have appeared. See for instance, de Regt, Leonelli & Eigner (2009) and Faye (2014). To begin with most of today’s theories of explanation consider understanding to be the aim of explanation, since it is argued that the aim of explanation should contain a response to the question why we seek explanation in the first place. The answer seems to be that we need explanations for cognitive or epistemological reasons because they provide us with understanding. Hence, ‘understanding’ is no longer regarded as a merely psychological by-product of explanation but has become an important player in epistemology and cognitive science. The result of this development seems to be a reversal of roles, whereby what was previously regarded as the aim of explanation is now considered to be secondary to the kind of understanding one strives for.

Note that by placing understanding in the front of the explanatory enterprise, the pragmatic aspects of explanation also move to the foreground. For both the purpose of the explanation and the context in which the explanation takes place play an important role in such accounts of explanation. A significant number of philosophers have already acknowledged that we cannot get on with scientific explanation unless we incorporate these pragmatic features of explanation into the notion of explanation itself. In spite of all their differences we find such an attitude in Charles S. Peirce, John Dewey, Michael Scriven, Sylvain Bromberger, Nicolas Rescher, Bas van Fraassen, Peter Achinstein and many more.

However, even though “understanding” has moved to the forefront of the debate it is not at all clear how this concept should be spelled out in details. The extant literature covers different approaches: some with a

naturalistic flavour, others with a much more normative quality. On the one hand, naturalistically oriented approaches may, for instance, hold that understanding in science is not merely connected to explanation. They may point to other forms of understanding embedded in the scientific practice and to the kind of understanding that follows from interpretation of data or classification of new phenomena. Hence, in science manifestations of various forms of understanding are conveyed as cognitive abilities, as acquired "tacit" skills of experimentation, and reflective understanding that enable scientists to produce interpretation and explanation and to be informed by others' interpretation and explanation. On the other hand, more normative approaches may focus on the standards for having scientific understanding in contrast to everyday understanding both in general and in relation to a particular scientific practice. The general standards could be coherence, consistency, truth, relevance, soundness, and unification supplemented with particular scientific theories.

If understanding in science should not again degenerate into a purely psychological notion, it seems to be a requirement of any account of understanding that the epistemic state in which one is placed by a cognitive act of understanding must fulfill certain standards of rationality. Both the naturalist and the normativist would probably agree that any understanding based on deliberation requires certain normative commitments. But the naturalist would emphasize that it is the empirical study of scientific practice by, say, cognitive science that discloses the standards of understanding that scientists are committed to. These empirically discovered standards may very well vary from one science to another and from science to everyday life. In contrast, the normativist may argue that the standards of understanding are closely connected to some a priori claims about epistemic commitments with respect to justification, personal responsibility, and adequate evidence. For example, a normative theory might tell us that a person understands in an epistemically responsible way, if and only if he or she feels obliged to uphold or reject an explanation based on certain criteria of good science. Therefore the normativist would be sceptical about empirical studies that they can't possibly address the normative goal of scientific understanding. Because, according to the normativist, the norms of intelligibility scientists actually follow are not necessarily those they ought to follow. It is not every alleged kind of understanding we may encounter in science that can reasonably count as a form of scientific understanding. A scientist must be ready to show that she is entitled to attribute to herself or others scientific understanding because her explanation or interpretation meets some well-established norms of scientific intelligibility. However, the naturalist would be much more tempted to argue that the criteria of intelligibility whereby scientists *do* arrive at their understanding are *the same as* the criteria by which scientists *ought* to arrive at their understanding.

The symposium attempts to cast some light on the compatibility of and tensions between these two lines of thinking including questions like: What is the epistemic difference between explanation and understanding? Can understanding and intelligibility be separated and, if they are regarded as distinct, how should they then be defined. Are there other forms of understanding in science than explanatory and

interpretive understanding? Is it possible, for instance, to be a naturalist about the standards of intelligibility without committing oneself to the naturalistic fallacy? Answers to some of these questions are not only important for theories concerning scientific explanation but have a significant impact on how to conduct philosophy of science.

### References

- Faye, J (2014), *The Nature of Scientific Thinking. On Interpretation, Explanation, and Understanding*. London: Palgrave Macmillan.
- Khalifa, Kareem (2013). The Role of Explanation in Understanding, *British Journal for Philosophy of Science*. 64, 161–187.
- Newman, Mark (2013), EMU and Inference: what the explanatory model of scientific understanding ignores, *European Journal for Philosophy of Science*. DOI 10.1007/s13194-013-0075-0
- H.W. de Regt (2013), Understanding and explanation: Living apart together? *Studies in History and Philosophy of Science* 44, pp. 505-510.
- H.W. de Regt, S. Leonelli & K. Eigner, eds (2009), *Scientific Understanding: Philosophical Perspectives*. Pittsburgh: University of Pittsburgh Press.

**Abstracts****1. Jan Faye: *An evolutionary and cognitive approach to understanding***

Based on my recent book *The Nature of Scientific Thinking* (2014) I argue that understanding can be grasped in evolutionary terms as a cognitive organization of beliefs. This has a number of implications. First, animals arguably possess non-verbal understanding. Second, human comprehension covers different forms of understanding such as innate, embodied, and reflective understanding. Third, 'truth' and 'falsity' do not characterize understanding but only the beliefs organized by our mind. Fourth, it is not only various forms of explanation but also various forms of interpretation that provide us with understanding. Explanations as well as interpretations help us to organize and unify our body of beliefs. Fifth, it is the cognitive organization of our beliefs that causes our sense of understanding. Taking these consequences into consideration, I shall argue that non-verbal forms of understanding (cognitive schemas) have purely innate standards of success in virtue of their contribution to survival. These standards may be described as mechanisms that secure consistency, coherence, relevancy, informative reliability, inference-aptness in relation to belief-formation. As humans eventually gained the capacity of reflective understanding the need of conscious interpretations and verbal request for explanations became immanent. However, it is also clear that interpretations and explanations are guided by personal and common interests. Hence, interpretation and explanation are always conducted in relation to a context of purposes that is a result of supported theories, background beliefs, and empirical knowledge. The context influences the kind of criteria of reflective understanding one upholds. The question is not whether science as part of a social enterprise follows the innate mechanism of successful understanding inherited from our ancestors but whether science establishes more reflection-based norms for epistemic understanding. I argue that science does establish such norms. Nevertheless, I argue that these epistemic norms are all context-dependent.

**2. Antigone M. Nounou: *On Scientific Understanding without Explanation***

Considering that understanding is a mental state, one could hardly dispute the idea that a good scientific explanation may confer scientific understanding of the phenomenon explained. Thus, one might assert that the causal explanation of a phenomenon gives rise to understanding of the underlying causes, a deductive-nomological explanation allows for the comprehension of the role laws play in its occurrence, a unificationist explanation enables one to fathom how that particular phenomenon fits in the bigger picture of things, etc. Still, even when scientific theories and models are known to misrepresent phenomena, and hence they do not provide acceptable explanations, they afford additional, distinct kinds of scientific understanding; or so I argue. In particular, I examine models of the nuclear structure—such as the liquid drop model—and statistical models of critical behaviour and phase transitions, which save the phenomena of the systems they describe despite the fact that they misrepresent them. I argue that what we gain from such models is a form of scientific understanding that comes prior to—and even irrespective of—scientific explanations of



phenomena. This form of understanding I contrast with the other distinct form of scientific understanding that does not stem from scientific explanations and which is typically associated with interpretations of theories. Unsurprisingly, the typical case study is quantum mechanics and a conclusion to take home is that the kind of scientific understanding that is intertwined with a theory's interpretation is necessary for the explanations of phenomena that will be based it and its models. In closing, I would also like to suggest that the taxonomy of the kinds of understanding that we may arrive at by this type of philosophical analysis is at least a useful, if not also a necessary, preparatory step for further analysis whether our proclivities are normative or naturalistic.

### **3. Henk W. de Regt: *From explanation to understanding: normativity lost?***

In recent years the notion of understanding has occupied central stage in philosophical debates about the nature of scientific explanation, and a number of philosophical theories of understanding have been presented in the literature. Since understanding is typically associated with the pragmatic and psychological dimensions of explanation, shifting the focus from explanation to understanding may seem to imply a shift from accounts that embody normative ideals to accounts that provide accurate descriptions of scientific practice. Not surprisingly, many 'friends of understanding' sympathize with a naturalistic approach to the philosophy of science. However, this raises the question of whether the proposed theories of understanding can still have normative power. I will argue that this question can be answered in the affirmative, at least for my own contextual theory of scientific understanding. Although my theory has first and foremost a descriptive and explanatory aim, namely to describe the criteria for understanding employed in scientific practice and to explain their function and historical variation, it can (at least to some extent) serve as a basis for normative assessment of scientists' actions. I will develop two arguments in order to substantiate this conclusion. The first argument hinges on the distinction between prescriptive and evaluative normativity. Although the criteria for understanding and intelligibility that my theory specifies cannot be used as prescriptive normative rules for guiding scientists' actions, I will argue that they can be used as evaluative norms for assessing whether theories are intelligible and whether understanding has been achieved. The second argument is directed at the objection that explanation is part of the (normative) context of justification while understanding belongs to the (descriptive) context of discovery. I will argue against such a sharp context distinction and in favor of norms for understanding that derive from study of scientific practice.

### **4. Petri Ylikoski: *Normativity and the inferential account of understanding***

A central challenge for any theory of explanation is to make sense of how scientists evaluate explanations. While traditional philosophy of science has not paid much attention to these evaluative practices, the inferential account of understanding is based on the idea that focusing criteria used in these evaluations is both important and fruitful. Thus the account is based on the distinction between understanding and the

sense of understanding (Ylikoski 2009). While the former refers to an ability to make correct what-if - inferences about the phenomenon of interest, the latter consists of a metacognitive feeling that tells us when we have understood or grasped something. Empirical studies show that the sense of understanding is a highly fallible indicator of understanding, thus motivating the idea that there is a difference between understanding and thinking that one understands. Another key idea of the approach is the idea that instead of there being one dimension of explanatory “power”, the explanatory goodness has at least five different dimensions that can be articulated with the help of contrastive-counterfactual account of explanation underlying the inferential approach (Ylikoski & Kuorikoski 2010).

These ideas provide the foundation for the inferentialist solution to the tension between naturalistic and normative approaches that is the topic of this symposium. The inferential account is fully naturalistic: it is both based on and inspired by recent cognitive science work on explanatory cognition (Ylikoski 2009), extended cognition (Kuorikoski & Ylikoski 2015), and distributed cognition. However, rather than identifying understanding with any specific mental state, the approach makes it possible to critically evaluate not only individual scientist’s understanding, but also explanatory practices within scientific fields. Nonetheless, this critical stance is not based on a priori philosophical ideas about understanding or explanation, but in the ability of the approach to connect the field-specific epistemic goals with actual explanatory practices. This potential will be discussed with examples from mechanism-based explanations in the social sciences.

### References

- 2015 ‘External Representations and Scientific Understanding’, with Jaakko Kuorikoski, *Synthese*, in press.
- 2014 ‘Understanding with Theoretical Models’, with Emrah Aydinonat, *Journal of Economic Methodology* 21: 19-36.
- 2014 ‘Agent-Based Simulation and Sociological Understanding’, *Perspectives on Science* 22: 318-335.
- 2014 ‘Analytical Sociology and Rational Choice Theory’, with Peter Hedström, *Analytical Sociology: Norms, Actions and Networks* (edited by Gianluca Manzo), Wiley: 57-70.
- 2013 ‘Causal and constitutive explanation compared’, *Erkenntnis* 78, 277-297.

**Probabilities, Chances and Statistics**

**Symposium**

Organizer: Mauricio Suárez

Chair: Nina Retzlaff

Room 5E, Saturday 09:30 – 11:30

---

*On Individual Risk*

ALEXANDER P. DAWID

Cambridge University

apd@statslab.cam.ac.uk

---

*Unsharp Best System Chances*

LUKE FENTON-GLYNN

University College London

l.glynn@ucl.ac.uk

---

*Against Ontic Chances (cancelled)*

JENANN ISMAEL

University of Arizona

jtismael@email.arizona.edu

---

*Counterfactual Probabilities, Chances and Robust Explanations*

AIDAN LYON

University of Maryland

alyon@umd.edu

---

*Propensities, Chances, and Experimental Statistics*

MAURICIO SUÁREZ

London University, Complutense University of Madrid

msuarez@filos.ucm.es

**General Description**

Probabilistic chance is of course a topic with a long and distinguished history and pedigree in the philosophy of science. The history of philosophical attempts to grapple with objective probability, and what appears to be its ubiquitous role in modern statistical science, arguably begins already in the 19th century, with Charles Peirce's writings on the topic, if not before. Kolmogorov's axiomatization of the probability calculus in 1933 made the concept formally precise, but this did not settle issues of interpretation. On the contrary two main schools emerged as regards objective probability, having their origin in Richard Von Mises' and Karl Popper's frequency and propensity interpretations respectively. Most philosophers in either 'frequency' and 'propensity' school would agree that objective probabilities seem in some sense presuppositions in much scientific inquiry – particularly in areas where statistical methods are widely employed; yet, there are nonetheless profound disagreements as to how to understand the nature of such probabilities. Roughly, thinkers schooled in the Humean or Bayesian traditions have tended to see probabilities as reducible to further notions or concepts that are empirically or epistemically more accessible, such as experimental statistics, frequencies, or credences. By contrast, defenders of propensities have tended to understand probabilities as reducible to metaphysical chances, or primitive dispositions in nature.

In other words, the 'frequency' and 'propensity' interpretations of probability have this much in common: They both aim to reduce the concept of probability to something else, by interpreting probabilities in terms of other notions they regard as less problematic, either by way of being empirically accessible or more explanatory. The frequency interpretation tradition has attempted to reduce probabilities to finite frequency ratios in actual sequences ('finite frequentism') or limiting frequency ratios in hypothetical sequences ('hypothetical frequentism'). The propensity tradition has attempted to reduce probabilities to metaphysical chances, dispositional properties of chance set ups, or causal relations between dispositional properties and their manifestation conditions. Nowadays, a third school may be emerging that refuses to reduce probabilities to either metaphysical chances (understood as dispositional properties) or frequencies (whether finite ratios in actual data, or limiting ratios in hypothetical data). The contributions to this symposium are either fully in defence of this third school, or they approach different issues raised in relation to it.

More particularly the symposium focuses on an array of topics and issues where this dispute is not an idle one in metaphysics, but becomes relevant to methodological and experimental practice. Thus we shall be particularly concerned with what, on each of these schools regarding chance and probability, follows for the practice of probabilistic modelling, experimental statistical testing, and the methodology of causal inference. More specifically, we ask questions regarding how, on either of these views, we may go about understanding i) the role and function of probabilistic or statistical modelling; ii) the nature of the relation between chances, probabilities, and experimental statistics, and iii) the empirical status of probability.

**Abstracts****1. Philip Dawid: *On Individual Risk***

Writing recently about her decision to have a preventive double mastectomy, the actress Angelina Jolie said: "My doctors estimated that I had an 87 percent risk of breast cancer and a 50 percent risk of ovarian cancer, although the risk is different in the case of each woman." Where do such figures come from, and what if anything do they mean? Is it even possible to assign risk values to individual events? We survey a variety of possible explications of individual risk, these in turn being based on a variety of interpretations of probability, themselves classified as either "groupist" or "individualist". Attempts to interpret individual risk from a "group to individual" perspective eventually founder on the insoluble problem of the reference class. We present an alternative "individual to group" interpretation, which, dispensing with reference classes, relates instead to the level and detail of the information used, that being allowed to differ from one individual to another. This approach leads to asymptotically unique individual risk values - which however are typically not computable. However this concept too has its limitations. In the end, a fully satisfactory understanding of individual risk remains elusive.

**2. Luke Fenton-Glynn: *Unsharp Best System Chances***

Much recent philosophical attention has been devoted to the Best System Analysis (BSA) of laws and chance. In particular, philosophers have been interested in the prospects of the BSA (and variants upon it) for yielding *high-level* laws and chances. In this vein, influential arguments have recently been advanced for thinking that the best system for our world is one that entails the fundamental dynamical laws together with the probabilistic principles of Statistical Mechanics (SM), and also special science laws and chances. But a foundational worry about the BSA lurks: there do not appear to be uniquely appropriate measures of the degree to which a system exhibits theoretical virtues, such as simplicity and strength. Nor does there appear to be a uniquely correct exchange rate at which the theoretical virtues of simplicity, strength, and likelihood (or *fit*) trade off against one another in the determination of a best system. Moreover, it may be that there is no *robustly* best system: no system that comes out best under *any* reasonable measures of the theoretical virtues and exchange rate between them. This worry has been noted by several philosophers, with some arguing that there is indeed plausibly a set of tied-for-best systems for our world (specifically, a set of very good systems, but no robustly *best* system). Some have argued that this entails that there are no Best System laws or chances in our world. I argue that, particularly when we consider systems that entail high-level laws and chances, it *is* plausible that there is a set of tied-for-best systems for our world, but that it doesn't follow from this that there are no Best System chances. (I argue that the situation with regard to laws is more complex.) Rather, it follows that (some of) the Best System chances for our world are *unsharp*.

**3. Jenann Ismael: *Against Ontic Chances*** (cancelled)

There is a great divide in views about the metaphysical status of chance. According to ontic views, chances are beables and beliefs about chance are beliefs about first-order matters of fact. According to epistemic views, chances are not beables. Beliefs about chance are either credences, or beliefs about what credences one ought to have about categorical matters. Epistemic views are often thought to be attractive because they explain the connections between chance and credence (if chances are beables in their own right, why should beliefs about them guide credence in categorical matters?), and because they leave room for a complex and informative story about why should believers like us adopt the chances as their credences (why chance rather than any number of other functions that can be defined and which ones might play the same role). Ontic views are attractive because chances seem to play a fundamental role in physics. I look at reasons for preferring an epistemic view that are based on their departure from familiar puzzles about the way beliefs about chance interact with ignorance about categorical facts. I conclude with some remarks about how these considerations bear on the interpretation of the quantum state.

**4. Aidan Lyon: *Counterfactual Probabilities, Chances and Robust Explanations***

Some of our best scientific explanations make references to probabilities, and this has led some philosophers to conclude that those probabilities are objective chances, because they cannot be subjective probabilities. For example, classical statistical mechanics explains why ice cubes melt in terms of probabilities of them being in micro-states that lead to them being in melted macro-states in the future. And since ice cubes seem to melt independently of what we know about them, the probabilities in question should be understood as objective physical chances (e.g., Popper 1982, Loewer 2001).

However, this line of reasoning quickly runs into a problem: classical statistical mechanics assumes that the world is deterministic, and it would seem that deterministic worlds cannot be chancy. So, some philosophers have concluded that the probabilities in question must be subjective (a.k.a. “epistemic”) probabilities after all (Schaffer 2007). This has an important ramification: the probabilities in question are epistemic in nature, and so the explanations in question are about what we should expect to happen.

I argue that both lines of reasoning are mistaken, and their mistakes come from not first clearly identifying the different conceptual roles that probabilities play in the sciences. I’ll show that by delineating at least three concepts of probability — counterfactual probabilities, chances, and credences — many conceptual confusions can be avoided, especially ones involving determinism and chance. I’ll argue that many of the probabilities that appear in high-level sciences should be understood as counterfactual probabilities, given the work they play in scientific explanations. This makes counterfactual probability a kind of objective probability that is distinct from chance.

**5. Mauricio Suárez: *Propensities, Chances, and Experimental Statistics***

Probabilistic modelling may be most generally described as the attempt to characterise (finite) experimental data in terms of models formally involving probabilities. I argue that a coherent understanding of much of the practice of probabilistic modelling calls for a distinction between three notions that are often conflated in the philosophy of probability literature. A probability model is often implicitly or explicitly embedded in a theoretical framework that provides explanatory – not merely descriptive – strategies and heuristics. Such frameworks often appeal to genuine properties (elements of reality or ‘beables’) of objects, systems or configurations, with putatively some explanatory function. Thus, I claim, it becomes useful to distinguish probabilistic dispositions (or single-case propensities), chance distributions (or probabilities), and experimental statistics (or frequencies). I illustrate the distinction with some elementary examples of games of chance, and go on to claim that it is readily applicable to more complex probabilistic phenomena, notably quantum phenomena.

I then argue that it is possible to understand the role of these three notions in probabilistic modelling in terms of Bogen and Woodward’s (1988) three-tier or tripartite distinction between theory, phenomena and data. Thus I suggest that in the context of probabilistic modelling, propensities are best understood as explanatory posits of *theory*, which both ground and explain chance or probability distributions. These distributions in turn are often best understood as models of *phenomena* in the sense described by Woodward and Bogen. Finally, relative frequencies of particular experimental outcomes in a given sequence constitute experimental *data*. It follows from the application of the tripartite distinction that propensities are typically not to explain particular outcomes or experimental data but rather the phenomena in the form of chance or probability distributions. The statistical data in turn may be used to directly confirm (and therefore also to test) probabilities, but not propensity ascriptions. The ascription of particular propensities – as Charles Peirce noted long ago – is rather to be justified (or criticized) by abductive means in terms of their explanatory qualities.

**References**

- Bacciagaluppi, G. (2013). A Critic Looks at QBism, [http://philsci-archive.pitt.edu/9803/1/comments\\_on\\_QBism\\_-\\_final.pdf](http://philsci-archive.pitt.edu/9803/1/comments_on_QBism_-_final.pdf)
- Ballentine, L. E. (1970). The statistical interpretation of quantum mechanics. *Reviews of Modern Physics* 42, 358–381.
- Ballentine, L. E. (1990). *Quantum Mechanics*. Englewood Cliffs: Prentice Hall.
- Bassi, A. and G. Ghirardi (2003). Dynamical reduction models. *Physics Reports* 379, 257.
- Bogen, J. and J. Woodward (1988), Saving the Phenomena, *The Philosophical Review* XCVII (3), 303-352.
- Bub, J. (1997). *Interpreting the Quantum World*. Cambridge: Cambridge University Press.
- Dawid, A. P. (2007), Counterfactuals, hypotheticals and potential responses: A philosophical explanation of statistical causality, in F. Russo and J. Williamson (eds.), *Causality and Probability in the Sciences*, London College Texts, pp. 503-532.
- Deutsch, D. (1999). Quantum theory of probability and decisions. *Proceedings of the Royal Society of London*

A455, 3129–3137.

- Durr, D., S. Goldstein, R. Tumulka, and N. Zanghi (2004). Bohmian mechanics and quantum field theory. *Physical Review Letters* 93, 090402.
- Fuchs, C. (2010). QBism, the Perimeter of Quantum Bayesianism. arXiv:1003.5209 .
- Greaves, H. (2007a). On the Everettian epistemic problem. *Studies in the History and Philosophy of Modern Physics* 38, 120–152.
- Greaves, H. (2007b). Probability in the Everett interpretation. *Philosophy Compass* 38, 120–152.
- Harrigan, N. and R. W. Spekkens (2010). Einstein, incompleteness, and the epistemic view of states. *Foundations of Physics* 40, 125.
- Hofer, C. (2007). The third way on objective probability: A skeptic's guide to objective chance. *Mind* 116, 549-96.
- Ismael, J. (1996). What chances could not be. *British Journal for the Philosophy of Science*, March 1996
- Lewis, D. (1980). A subjectivist's guide to objective chance. In R. C. Jeffrey (Ed.), *Studies in Inductive Logic and Probability*, volume II. Berkeley: University of California Press. Reprinted, with postscripts, in David Lewis, *Philosophical Papers, Volume II* (Oxford University Press, Oxford, 1986); page numbers refer to this version.
- Loewer, B. (2001). Determinism and chance, *Studies in History and Philosophy of Modern Physics*, 32: 609–20.
- Lyon, A. (2012). From Kolmogorov, to Popper, to Renyi: There's no escaping Humphreys' paradox (when generalised). *Chance and Temporal Asymmetry*, A. Wilson (ed.), Oxford University Press.
- Mises, v. R. (1957), *Probability, Statistics and Truth*, New York: Dover.
- Mumford, S. and R. L. Anjum (2011), *Getting Causes out of Powers*, Oxford University Press.
- Peirce, C. S. (1878 / 1955) On the doctrine of chances" in *Philosophical Writings of Peirce*, J. Buchler (ed.), New York: Dover, pp. 157-172.
- Popper, K. (1957) The propensity interpretation of the calculus of probability, and the quantum theory, in S. Körner (ed.), *Observation and Interpretation in the Philosophy of Physics*, New York: Dover, pp. 65-70.
- Popper, K. (1982). *Quantum Theory and the Schism in Physics*, New Jersey: Rowman and Littlefield.
- Reichenbach, H. (1956), *The Direction of Time*, University of California Press.
- Schaffer, J. (2007). Deterministic chance?, *The British Journal for the Philosophy of Science*, 58 (2): 113–40.
- Suárez, M. (2013) Propensities and pragmatism. *The Journal of Philosophy*, vol. CX (2), 61-92.
- Timpson, C. G. (2008). Quantum Bayesianism: A study. *Studies in History and Philosophy of Modern Physics* 39: 579-609
- Wallace, David (2013) Probability in physics: stochastic, statistical, quantum, <http://philsci-archive.pitt.edu/9815/>).
- Woodward, J. (2005). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.



**General Philosophy of Science VII****Contributed Papers**

Chair: Jesus Zamora Bonilla

Room 5F, Saturday 09:30 – 11:30

---

*Conceptualizing Uncertainty: An Assessment of the Latest Uncertainty Framework of the Intergovernmental Panel on Climate Change*

NICOLAS WÜTHRICH

London School of Economics and Political Science

n.wuethrich@lse.ac.uk

---

Climate change has the potential to generate tremendous ecological, economic, and social impact. Although it is commonly agreed that this change is also driven by anthropogenic factors, it is far from clear what the optimal responses to these changes in our climate system are. One reason for this is that we are facing severe uncertainty regarding the physical facts about the phenomenon of climate change. The Intergovernmental Panel on Climate Change (IPCC) was established to synthesize the latest scientific knowledge on climate change and to communicate it to policy makers. To this end, the IPCC developed as a key element an uncertainty framework.

In this paper, I assess the latest version of the uncertainty framework. The paper is divided into three parts. First, I look at the meta-documents which introduce and explain the uncertainty framework. I argue that there are substantial conceptual issues which need attention. Secondly, I focus on the full report of Working Group I, which focuses on the physical science basis of climate change, to explore how the uncertainty framework is put into practice. I show that the conceptual problems of the framework manifest themselves in concrete practical problems for the authors of the assessment report. Based on these observations, I suggest, thirdly, improvements to make this tool more fruitful in the highly relevant context of climate policy-making and, potentially, in other areas of policy-making as well.

The uncertainty framework equips the scientist with a confidence and a likelihood metric to qualify her statements about the causes and effects of climate change (see IPCC 2010 and Mastrandrea et al. 2011). An example for the use of this two-dimensional uncertainty framework is the following:

“In the Northern Hemisphere, 1983-2012 was likely the warmest 30-year period of the last 1400 years (medium confidence)” (IPCC 2013, 3)

*Confidence* describes the validity of a finding. The confidence judgement is the result of the aggregation of two sub-metrics ‘evidence’ and ‘agreement’. The *evidence judgment* is the result of an aggregation across different dimensions: the type of evidence, its amount, its quality, and its consistency. The *agreement judgment* is capturing the consensus across the scientific community on a given finding. The *likelihood metric*

is a quantified measure of uncertainty and expresses a probabilistic estimate of the occurrence of events or outcomes. The verdict 'likely', for example, is associated with a probability range of 66-100%.

The exposition of the uncertainty framework reveals fundamental conceptual problems. The problems can be grouped into three categories: problems associated with (a) the lack of definitions of key terms, (b) the lack of specifications of relations between key terms, and (c) epistemological assumptions. Subsequently, I highlight a selection. *In relation to category (a)*, the agreement sub-metric of the confidence scale is defined both as the degree of consensus between scientific publications on a finding as well as the number of competing causal explanations for a finding. Furthermore, the dimension 'quality of evidence' is not specified at all. *In relation to category (b)*, the relationship between the sub-metric 'agreement' and the dimension 'consistency of evidence' is not clear because consistency is introduced as the degree to which evidence supports single or multiple explanations or projections. In addition, the meta-documents support two, mutually inconsistent interpretations of the relationship between the confidence and the likelihood metric. A first interpretation sees confidence statements as meta-judgment about the validity of the finding whereas likelihood statements are, uncorrelated, intra-finding judgments about the probability of an event or outcome described in the finding. A second interpretation understands confidence statements and likelihood statements as conveying the same information, whereas confidence judgments are used for qualitative and likelihood judgments are used for quantitative evidence. *In relation to category (c)*, the fact that robust evidence (i.e. multiple lines of high quality, independent, and consistent pieces of evidence) could, according to the uncertainty framework, appear in combination with low, medium, or high agreement in the scientific community is puzzling from an epistemological point of view.

These conceptual problems give rise to concrete practical issues for the authors of the assessment report. For example, the authors fill the lacuna which is generated by the absence of the specification of the relationship between the agreement sub-metric and the dimension 'consistency of evidence' by equating these two terms. This yields an over-emphasis on agreement in comparison to evidence if confidence in a finding is evaluated. Moreover, the absence of a specification of the dimension 'quality of evidence' and the evidence categories 'limited' and 'medium' give rise to a large ambiguity in the application of the uncertainty framework. The authors have to make decisions and there is no indication present in the report that these decisions are taken in a consistent and non-arbitrary way.

Taking these findings into account, I identify the *construction of the confidence metric* as the key issue which needs attention for improving the framework. I motivate *three changes*. First, the agreement sub-metric has to be interpreted uniquely as consistency of evidence indicating whether the set of evidence supports qualitatively agreeing or disagreeing estimates of parameters. Secondly, the evidence sub-metric should be reduced to an aggregation of the dimensions 'quality' and 'amount' where quality of evidence should be made dependent on the different categories of evidence (e.g. mechanistic understanding, observational data, or model results). Thirdly, all categories of the two sub-metrics (e.g. robust evidence or

medium agreement) should be introduced in the meta-documents together with paradigmatic examples. These three changes do not only provide the IPCC with a conceptually more coherent uncertainty framework but also allow exploring whether this confidence metric can be used in other policy contexts. One example is macroeconomics where we face complex systems with expectation-driven feedback mechanisms and considerable disagreement about the relevant causal structures.

### References

- IPCC. 2010. "Guidance Note for Lead Authors of the IPPCC Fifth Assessment Report on Consistent Treatment of Uncertainties" IPCC Cross-Working Group Meeting on Consistent Treatment of Uncertainties (July 6-7, 2010): 1-6.
- IPCC. 2013. "Climate Change 2013 – The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of Intergovernmental Panel on Climate Change" Eds. Stocker, T. F., Qin, D., Plattner, G. L., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. M., 1-1552. Cambridge University Press: Cambridge.
- Mastrandrea, Michael D., Katharine J. Mach, Gian-Kasper Plattner, Ottmar Edenhofer, Thomas, F. Stocker, Christopher B. Field, Kristie L. Ebi, and Patrick R. Matschoss. 2011. "The IPCC AR5 guidance note on consistent treatment of uncertainties: a common approach across the working groups" *Climate Change* 108: 675-691.

---

*Causality and Natural Kinds*

OLIVIER LEMEIRE

KU Leuven

olivierlemeire@gmail.com

---

Most realist theories of natural kinds consider causality a crucial feature of the metaphysical nature of kind. According to (neo-)essentialists, kinds are characterized by (a set of) necessary and sufficient properties that cause many others properties of the kind. According to one often-discussed example, the essence of water is to be composed of H<sub>2</sub>O, which causes many other properties of water like its boiling and freezing point. Boyd's very popular Homeostatic Property Cluster Theory (HPC) of natural kinds relaxes the need for kinds to be characterized by essences, but maintains the requirement that kinds are grounded in the causal structure of reality. According to Boyd, natural kinds without essences are characterized by clusters of properties that regularly co-occur because of the workings of a homeostatic mechanism. Common to both the essentialist view and Boyd's HPC Theory is the idea that the causal ground of kinds explains their epistemic fertility: they support multiple inductive projections and explanations.

Recently, Matthew Slater has presented a novel argument against the view that natural kinds need to be grounded either in causal properties or in causal mechanisms. Like Boyd, Slater argues that kinds are associated with *clusters* of properties none of which need to be essential. Against Boyd, Slater argues that homeostatic *mechanisms* are neither sufficient nor necessary for (biological) kinds to be natural. Furthermore, Slater maintains that causality in general is not a necessary criterion of natural kinds. In response to causal theories of natural kinds, Slater presents the Stable Property Cluster Theory (SPC) according to which kinds are associated with a cluster of properties whose co-occurrence is counterfactually stable. Ultimately, it is the stability of this clustering that grounds the epistemic fertility and naturalness of kinds.

I will present two arguments against Slater's SPC Theory of natural kinds and in favor of a causal theory of natural kinds. The first argument is aimed specifically at Slater's non-causal theory, and aims to show the necessity of a causal ground for natural kinds. The second argument is aimed at any theory that takes the regular co-occurrence (or clustering) of properties as being a necessary criterion of natural kinds. This includes Slater's SPC Theory, but also essentialist theories and Boyd's HPC Theory. This second argument aims to show that although members of natural kinds must minimally share one or some causal properties, the regular or counterfactually stable *co-occurrence (clustering)* of many properties is not a necessary feature of natural kinds.

For the first argument, it is important to note that Slater specifies that the type of *clustering* of properties required for natural kinds is such that whenever we find some properties (any sub-cluster) associated with a kind, it is likely that we will find the other properties associated with it as well. For example, if we find a liquid that has the boiling point and freezing point of water, we are likely to also find the other properties of water as well. The fact that we can do this for many sub-clusters associated with the kind, is what, according to Slater, makes natural kinds so epistemically fruitful. But this is too strong. To see this, consider that the likelihood is not very high that one would find many other properties of water given that something is a colorless and odorless liquid, as there are several other liquids with these properties. Nevertheless, this fact does not make water any less of a natural kind. Similar examples can be multiplied. The likelihood of finding all the properties associated with the kind *water* given a particular sub-cluster of its properties is not what grounds the epistemic fertility of water. Rather, what does make water epistemically useful is that wherever we find *H<sub>2</sub>O*, we are likely to find all the other properties typical of water. Natural kind categories are often scientifically useful not because they refer to clusters of properties but because they refer to causal properties that result in a cluster of properties.

However, and this is the second argument, not all natural kinds are associated with clusters of properties. To be sure, many natural kinds are defined so as to support multiple inductive projections and hence are associated with a cluster of co-occurring properties. But this does not apply to all natural kinds. Natural kinds crucially ground generalizations, but these do not necessarily apply to all or even many of its members. Take, for example, the geological kind *igneous rock*. These are rocks that are formed as the result of the cooling of magma or lava and they come in two general types, depending on whether the cooling has occurred above the Earth's surface (extrusive/volcanic rock) or beneath the surface (intrusive/plutonic rock). Although there are many generalizations about igneous rocks that apply to both types, geologists also make generalizations about *igneous rock* that only apply to one of these subtypes. However, if natural kinds would be those kinds that support the most projections to unobserved instances, it is clear that the subtypes would be the only natural kinds and there would be no need for *igneous rock* as a kind. These subtypes are characterized by more properties that co-occur regularly or counterfactually stable. Nevertheless, *igneous rock* is also a natural kind because it allows for systematic explanations. It is the cooling of magma that causes, and hence allows us to explain, many of the common properties of plutonic and volcanic rocks, *but also many of the properties in which they differ*. Similar examples can be multiplied, as there are, for example many polymorphisms of biological species or gender specific effects of human diseases. Actual scientific classification shows that kinds are not only epistemically fruitful because they causally ground clusters of co-occurring properties, but also because they are associated with causal properties that account for many other properties, although these do not need to co-occur in all or most kind members. Based on these arguments I develop a view which I call the Causal Unification Theory (CUT) of natural kinds.

---

*Pan-Perspectival Realism*

PAUL TELLER

University of California at Davis

prteller@ucdavis.edu

---

In a brief sketch my argument is as follows:

- 1) Because the world is so complex, all human representation is to some extent imprecise and/or inaccurate.
- 2) Perception is, or constitutively involves, representation. Consequently,
- 3) Not just scientific, but also perceptual, knowledge is qualitatively affected with the limitations as claimed in 1).
- 4) So scientific and perceptual knowledge are of a piece, and both an ever refinable but never exact view of the way things really are.

What are the limitations of 3)? Standard scientific realism fails for a simple semantic reason: Because of 1) we can't attach words to anything specific. (*Whose* atoms? Dalton's? Perrin's? Today's chemists? Indeed, many field theorists insist that there are no particles.) Standard scientific antirealism, on the other hand, assumes that when we let go of the theoretical, we could fall back on the things and properties we know by perception. But, because of (3) that fails as well. (Consider the complexities of the still idealized current science of color.) So the stakes are high: If we can't find some more nuanced way to be a realist, (3) will have to be interpreted as some kind of representational idealism.

For the same reasons that the sense data theory failed, a catalogue of representations, like a catalogue of pictures, can't get us even an "accurate enough" grip on the world. We have to understand the world as *like* one filled with our ordinary physical objects. Thinking in terms of ordinary physical objects, though still an idealization, tells us that our world is one very like one occupied by these physical objects, despite the fact that if we examine too closely these turn out to involve idealization.

Finally, once we see how this works for the objects of perception, the same goes, and for the same reasons, for the objects described by a successful science. This substantiates the long-standing conclusion that, because of the absence of any observational/theoretical distinction, objects of perception and those of science stand or fall together. This is as good as realism gets.

To fill in this argument sketch I will introduce the notion of semantic alter-egos. Any statement made sufficiently precise will be, strictly speaking, false. But if it is sufficiently accurate, the statement can be turned into a truth by making it less precise. Consider the false "John is six feet tall precisely", the idealized semantic alter-ego, vs. "John is six feet tall close enough", the former's truth apt semantic alter-ego. The latter counts

as, literally, true when the former, though false, is accurate enough to function as a truth. Alter-egos because they do the same semantic work.

The application: How should we interpret “There is a chair in front of me”? If by “the chair” one insists on a physical object with completely determinate identity conditions the statement is, if not false, at least not true. Because of indefinite spatial and temporal boundaries, no exact identity conditions apply. Talk of “the chair” is an idealization. But if we understand “the chair” itself as an imprecise term, the imprecise statement counts as, literally, true because its idealized semantic alter-ego functions as a truth. One can de-idealize not only by increasing accuracy but by decreasing precision.

Now consider an application to “atom”. From the pen of Dalton, Perrin, or today’s chemist, if we take “atom” as an imprecise term, we can take all to have written truths. Alternatively, we can regard the concept of an atom as that of an idealized particle, with different details of idealization in different theoretical contexts. Quine admonished: No entity without identity. Quantum theories reject any such, but idealized atoms function brilliantly in various theoretical contexts.

The conceptual duality afforded by semantic alter-egos makes sense of our use of the tools of reference in face of the fact that no such representational reference gets things exactly right. Still, representational tools tell us a great deal about how the world works in ways that admit of right and wrong, better and worse; so we should count them as informing us of how things really are.

Some will persist in asking: But do physical objects/atoms exist – really?! We must reject this question. A “yes” answer conveys that a) our referential tools have been attached to referents, b) that these referents have exact identity conditions, and c) all this is not an idealization. We know we can’t have any of the three. It’s not that we positively know that there are no identity-bearing concreta. There may or may not be, but either way our referential tools are not so attached. A “no” answer conveys that our referential language is completely wide of the mark. That at best is disastrously misleading. We do much better than some kind of referential idealism. The scientific realism debate has been irresolvable because limited to two equally unviable alternatives.

---

*The No Miracles Argument without Base Rate Fallacy*

RICHARD DAWID

LMU Munich

richard.dawid@univie.ac.at

---

The talk is based on joint work with Stephan Hartmann.

The No Miracles Argument (NMA) is one of the most influential arguments for scientific realism. It argues that the predictive success of science would be a miracle if mature scientific theories didn't tend to be approximately true.

In his book (Howson 2000), Colin Howson argues that the NMA commits the base rate fallacy and therefore is logically flawed. The argument was reemphasised more recently in (Callender and Magnus 2003). Howson and the latter authors offer a Bayesian formalization of the NMA. In this formalization, the hypothesis that a theory  $H$  is approximately true is updated based on the evidence that  $H$  was predictively successful. The inference to a high probability of truth relies on the assumption that the probability of predictive success is high if the theory is approximately true but low if the theory is false. Bayesian analysis then immediately shows that the inference to a high probability of approximate truth of  $H$  would be invalidated by a very low prior of its approximate truth. Specifying the prior would amount to begging the question, however, since a reasonably high prior probability of approximate truth of scientific theories is exactly what the anti-realist denies. Thus, the NMA implicitly relies on specifying the prior probability of the approximate truth of  $H$  but has not basis for doing so. In other words, it commits the base rate fallacy.

Our analysis of Howson's argument starts with the observation that the NMA can be and has been understood in two different ways. First, the fact to be explained by the realist conjecture may be taken to be the predictive success of an individual theory. We call an argument on that basis the individual theory-based NMA. Second, the fact to be explained may also be taken to be the tendency of predictive success in mature science or in some specific segment of mature science. We call an argument on that basis the frequency-based NMA. We then go on to show that Howson's formalization of the NMA is a reconstruction of the individual theory-based NMA. Howson successfully demonstrates that the individual theory-based NMA is logically flawed. The frequency-based NMA, however, has a significantly different structure. It is based on observing a frequency of predictive success already before assessing the truth of a given individual theory  $H$ . Therefore, once one looks at a new theory  $H$  and updates under its predictive success, the prior probability of its truth cannot be freely chosen any more. Based on the law of total probability, a lower bound on the prior probability of the truth of  $H$  is enforced by the predictive success rate in the field in conjunction with the assumptions on the correlations between approximate truth/falsity and the probability of predictive



success. A formalization of the frequency-based NMA shows that it does not fall prey to the base rate fallacy. To the contrary, even fairly low tendencies of predictive success can imply a high posterior probability for a theory's approximate truth. It must be emphasised that avoiding the base rate fallacy does not amount to having a valid argument for scientific realism. The scientific realist is still faced with the difficult task to justify the assumptions regarding the correlations between truth and predictive success.

The second part of the talk addresses the question as to who did endorse the first, logically flawed version of the NMA, and who endorsed the second, logically sound version. Based on looking at the original formulations, it will be argued that the initial understanding as presented by Hilary Putnam and Richard Boyd clearly is of the frequency-based type. Later discussions often show a lack of awareness of the distinction between the two interpretations of the NMA. Some discussions clearly do adopt the individual theory based version of the NMA.

To conclude, the base rate fallacy constrains the way in which a NMA can be formulated in a meaningful way. The form of the NMA that avoids the base rate fallacy is the form in which the argument was initially presented.

### References

- Boyd, R. (1985): "Lex Orandi and Lex Credendi", in: P. Churchland and C. Hooker (eds.): *Images of Science: Essays on Realism and Empiricism with a Reply from Bas C. van Fraassen*. Chicago: University of Chicago Press, pp. 3--34.
- Howson, C. (2000): *Hume's Problem: Induction and the Justification of Belief*. Oxford: The Clarendon Press.
- Magnus, P. D. and C. Callender (2003): "Realist Ennui and the Base Rate Fallacy", *Philosophy of Science* 71: 320--338.
- Putnam, H. (1975): "What is Mathematical Truth?" in H. Putnam: *Mathematics, Matter and Method*, Collected Papers Vol. 2. Cambridge: Cambridge University Press.

---

*No Alternatives for What? Non-Empirical Evidence in the Case of String Theory*

RADIN DARDASHTI

Munich Center for Mathematical Philosophy/LMU Munich

radin@gmx.de

---

In a recent paper Dawid, Hartmann and Sprenger have shown within a Bayesian framework that the observation that there is no alternative theory to one's theory, at a given time and despite considerable effort, confirms the theory. This so-called No Alternatives Argument (NAA) is crucial in cases where empirical evidence is missing, as in String Theory. Unlike common theory confirmation the confirming evidence in this case is called non-empirical, since it is not a deductive or inductive consequence of the theory that there are no alternatives. The main focus of this paper is how one can obtain such non-empirical evidence in an objective, unbiased way. The conclusion is that in cases where the NAA is most needed (i.e. in theories where empirical evidence is missing), it is usually not yet applicable, while in cases where one does have enough non-empirical support, empirical evidence can be given too (as in the case of the Higgs mechanism), and so the NAA is not needed. The paper is divided into three parts, which I will discuss now.

**1. What is the precise definition of non-empirical evidence in the NAA?**

In the first part we critically analyse the definition of non-empirical evidence in the NAA and argue that its formulation in Dawid et al. (2015) is inadequate for the purposes of theories of quantum gravity, i.e. for those cases where it is most needed. We offer an extension and a problem-relative reformulation of the definition of non-empirical evidence, which allows for an application of the NAA to the relevant theories. Any NAA is then always relative to the specific set of problems  $P$  the theory is meant to solve. There remain two open questions: First, how do we individuate theories? And second, what is the specific problem set?

**2. How to individuate theories?**

The first problem arises from the need to individuate theories. It is obviously crucial for the No Alternatives Argument that it be possible to claim that there are no alternatives to one's theory. This, however, implies the possibility to individuate theories, since only if I can count theories, can I claim that the number of theories solving a problem is one. After arguing why the answer offered by Dawid et al. (2015) is not satisfactory I propose an alternative criterion of theory individuation, which for the purposes of the NAA offers a pragmatic solution to the problem. This criterion offers a problem-relative individuation, since

different problem sets can lead to different individuations. I will consider several examples to illustrate the applicability of this criterion.

### **3. *What is the right problem set?***

The second problem is due to the problem-relative statement of the non-empirical evidence in NAA. If one says one has no alternative, one always needs to specify with respect to what problem set there is no alternative. E.g. in the case of String Theory the statement is that it “is the only viable option for constructing a unified theory of elementary particle interactions and gravity”. But who determines what the relevant problem in need of a solution is? The determination of this set of problems is a priori highly non-trivial, especially in the cases where the NAA is most crucially needed, where the determination of the problem set can be dependent on the research program within which the scientist works. While we offer a pragmatic solution to the theory individuation problem, the problem-determination problem remains and leaves us with two possible interpretations of the NAA result in this light:

The first possible interpretation follows from the fact that if there is no way to justify the problem set independently, any scientist may regard her own favorite set of problems. This seems especially adequate in the context of theories of quantum gravity, where each research community has their own set of problems and favorite methods by which they aim at solving them. This has the following more general and rather undesirable consequence: within each research project one can find a unique problem set such that (according to the criterion of theory individuation) there will be no alternatives to that theory. What is the meaning of the confirmatory result of Dawid et al. in this light? The argument does not trivialise completely, since scientists work on the specific theories they are working on because they consider the theory they use as most appropriate considering the set of problems they wish to address. If there were many alternatives able to address the same problem set, their trust in their specific approach may decrease. The confirmatory result that follows from the NAA should then be understood as a justification for the scientists to work on the theory they use given their specific problem set. The confirmatory result should then not be understood as a result confirming the theory per se but as a result which accounts for the scientific practice.

The more interesting conclusion would be that there is a preferred problem set. In this case the NAA by itself may provide theory confirmation. Consider, however, the unification of all fundamental forces. Whether or not this should be considered as a problem in need of an explanation is non-trivial. So these claims go beyond the empirically justified problems. So if they are not empirically justified, one can only evaluate them by considering the appropriateness of the assumptions within the bigger research program. For instance, if unification has been the right guide in the development of theories in the past then they may be in the future as well. However, I argue that this kind of meta-inductive support is not available for theories of quantum gravity. This may lead to the unfortunate consequence that the NAA in cases it is most needed,

it usually will not yet be applicable, while in cases where one does have enough non-empirical support, empirical evidence can be given too, and so the NAA is not needed.

---

*The Borel-Kolmogorov Paradox and Conditional Expectations*

MIKLOS REDEI

London School of Economics  
m.redei@lse.ac.uk

ZALAN GYENIS

Hungarian Academy of  
Sciences  
gyz@renyu.hu

GÁBOR HOFER-SZABÓ

Hungarian Academy of  
Sciences  
szabo.gabor@btk.mta.hu

---

Suppose we choose a point randomly with respect to the distribution given by the uniform measure on the surface of the unit sphere in three dimension. What is the conditional probability that a randomly chosen point is on an arc of a great circle on the sphere on condition that the point lies on that great circle? Since a great circle has measure zero in the surface measure on the sphere, the Bayes formula cannot be used to calculate the conditional probability in question. On the other hand one has the *intuition* that the conditional probability of the randomly chosen point lying on an arc is well defined and is proportional to the length of the arc. This tension between the “ratio analysis” (Bayes formula) of conditional probability and our intuition is known as the Borel-Kolmogorov Paradox. The tension seems to be aggravated by the fact that different attempts to replace the Bayes formula by other, apparently reasonable, methods to calculate the conditional probability in question lead to different values.

The Borel-Kolmogorov Paradox, published first in 1909 [2], has been discussed both in mathematical works on probability theory proper [6][p. 50-51], [1][p. 441], [12], [13], [18], [14][p. 65], and in the literature on philosophy of probability [2][p. 100-104], [3], [17], [5][p. 470], [16], [8]. One can discern two main attitudes towards the Paradox: a radical and a conservative.

According to radical views, the Borel-Kolmogorov Paradox poses a serious threat for the standard measure theoretic formalism of probability theory, in which conditional probability is a defined concept, and this is regarded as justification for attempts at axiomatizations of probability theory in which the conditional probability is taken as the primitive rather than a defined notion [4], [19], [3]. Such axiomatizations have been given by Popper [9], [10], [11], and Renyi [15] (see [7] for a recent analysis of Renyi’s and Popper’s approach).

According to “conservative” papers the Borel-Kolmogorov Paradox just makes explicit a pitfall in naive conditioning that can be avoided within the measure theoretic framework by formulating the problem of conditioning properly and carefully. Once this is done, the paradox is resolved. Kolmogorov himself took this latter position [6] [p. 50-51]. Billingsley [1] [p. 441], Rao [13] [p. 441] and Proschan and Presnell [12] [p. 249] write about the Borel-Kolmogorov Paradox in the same spirit (Proschan and Presnell call the Borel-Kolmogorov Paradox the “equivalent event fallacy”).

The present paper falls into the conservative group: We claim that the Borel-Kolmogorov Paradox is in perfect harmony with measure theoretic probability theory. Specifically: conditional probabilities with respect to probability zero events can be defined and treated in a consistent manner if one uses the theory of conditional expectations as the conditioning device. One can obtain the intuitively correct uniform distribution on any great circle by choosing a suitable conditional expectation: the conditional expectation determined by the Boolean algebra  $O$  generated by the circles parallel to a given great circle. Calculating the conditional probabilities on a great circle  $C$  using the conditional expectation determined by the Boolean algebra  $M$  generated by all the meridian circles that intersect  $C$  at the same two points (North and South Poles), one obtains a conditional probability on  $C$  that is non-uniform however. It will be shown that the difference between the  $M$ -conditional and  $O$ -conditional probabilities on great circles does not indicate a paradoxical dependence of conditional probabilities of the *same event* with respect to the *same conditioning* conditions in *different co-ordinatization* but a sensitive dependence of conditional probabilities of the *same event* on *different conditioning* Boolean subalgebras with respect to which conditional probabilities are defined in terms of conditional expectations. This latter dependence is however not only not paradoxical but entirely natural and expected once the concept of conditional probability is defined properly in terms of conditional expectations.

It also will be argued that both the  $M$ -conditional and  $O$ -conditional probabilities are intuitively correct, if one has the proper concept of conditional probability defined by conditional expectations, and it does not make sense to ask whether the  $M$ -conditional or the  $O$ -conditional expectations/probabilities are the correct ones: The algebras  $M$  and  $O$  represent different conditioning circumstances and the conditional probabilities they lead to are answers to different questions – not different answers to the same question. In certain applications  $M$ , in certain other applications  $O$  represent some circumstances that are described correctly by the corresponding conditional probabilities. This is an advantage, showing the flexibility of probability theory in modeling random phenomena. There is no “absolute” notion of conditional probability – conditional probabilities are truly *conditional*: they depend on a full set of conditions, i.e. on a Boolean subalgebra.

Thus, under close and careful scrutiny, the “paradox” in the Borel-Kolmogorov Paradox evaporates: There is no clash between the *correct* intuition about what the conditional probabilities with respect to probability zero events are and the technically proper concept of conditionalization via conditional expectation.

## References

- [1] P. Billingsley. *Probability and Measure*. John Wiley & Sons, New York, Chichester, Brisbane, Toronto, Singapore, Third edition, 1995.
- [2] E. Borel. *Éléments de la Théorie des Probabilités*. Librairie Scientifique A. Herman & Fils, Paris, 1909. English translation: by J. Freund, “Elements of the Theory of Probability”, Englewood Cliffs, 1956, Prentice-Hall.
- [3] A. Hájek. What conditional probability could not be. *Synthese*, 137:273–333, 2003.
- [4] W. Harper. Rational belief change, Popper functions, and counterfactuals. *Synthese*, 30:221–262, 1975.

- [5] E.T. Jaynes. *Probability Theory. The Logic of Science*. Cambridge University Press, Cambridge, 2003. edited by G. Larry Bretthorst.
- [6] A.N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin, 1933. English translation: *Foundations of the Theory of Probability*, (Chelsea, New York, 1956).
- [7] D. Makinson. Conditional probability in the light of qualitative belief change. *Journal of Philosophical Logic*, 40:121–153, 2011.
- [8] W. Myrvold. You can't always get what you want: Some considerations regarding conditional probabilities. *Erkenntnis*, 2014. Forthcoming, online August 5, 2014, doi:10.1007/s10670-014-9656-3.
- [9] K. Popper. A set of independent axioms for probability. *Mind*, 47:275–277, 1938.
- [10] K. Popper. Two autonomous axiom systems for the calculus of probabilities. *The British Journal for the Philosophy of Science*, 6:51–57, 1955.
- [11] K. Popper. *The Logic of Scientific Discovery*. Routledge, London and New York, 1995. First published in English in 1959 by Hutchinson Education.
- [12] M.A. Proschan and B. Presnell. Expect the unexpected from conditional expectation. *The American Statistician*, 52(3):248–252, 1998.
- [13] M.M. Rao. Paradoxes in conditional probability. *Journal of Multivariate Analysis*, 27:434–446, 1988.
- [14] M.M. Rao. *Conditional Measures and Applications*. Chapman & Hall/CRC, Boca Raton, London, New York, Singapore, 2nd, revised and expanded edition, 2005.
- [15] A. Rényi. On a new axiomatic theory of probability. *Acta Mathematica Academiae Scientiae Hungaricae*, 6:268–335, 1955.
- [16] M. Rescorla. Some epistemological ramifications of the Borel-Kolmogorov Paradox. *Synthese*, 2014. Forthcoming. Published online November 20, 2014. DOI 10.1007/s11229-014-0586-z.
- [17] T. Seidenfeld. Remarks on the theory of conditional probability: Some issues of finite versus countable additivity. In V. F. Hendricks, editor, *Probability theory*, pages 167–178. Kluwer Academic Publishers, 2001.
- [18] T. Seidenfeld, M.J. Schervish, and J.B. Kadane. Improper regular conditional distributions. *The Annals of Probability*, 29:1612–1624, 2001.
- [19] B.C. van Fraassen. Representation of conditional probabilities. *Journal of Philosophical Logic*, 5:417–430, 1976.

---

*Symmetries and the Identity of Physical States*

SIMON FRIEDERICH

University of Groningen

email@simonfriederich.eu

---

The importance of symmetries in physics has been a recurring topic in the philosophy of science in recent years. According to a (preliminary and imprecise) characterization, symmetries are (or induce) mappings of a theory's state space onto itself which connect states that are in some sense "physically equivalent". Philosophical debates about symmetries often start from recognizing that "physical equivalence" can have (at least) two different meanings in this context. Distinguishing between these two meanings by deciding which one applies in which case is perhaps the main challenge for the philosophical analysis of symmetries.

According to the first meaning of "physical equivalence", symmetries are descriptive redundancies in that any two states related by a symmetry transformation represent one and the same physical state of affairs in mathematically distinct ways. According to the second, symmetries operate between physically distinct states of affairs, but in such a way that there is no empirically detectable difference between states connected by symmetries for observers who can only make observations inside the region where the symmetry transformations operate. How to classify symmetries in actual physical theories in terms of this distinction is a nontrivial task with respect to which there are controversial views. The debate is often formulated in terms of the question of which symmetries have "direct empirical significance" and which do not. Roughly speaking, those symmetries which connect physically identical states of affairs are (or correspond to) those which do not have any direct empirical significance, whereas those which operate between physically distinct states of affairs are (or correspond to) those which have some.

Clearly distinguishing between symmetries that have direct empirical significance and those that do not is the same as specifying which mathematical states correspond to the same physical state. Therefore, the debate about symmetries and their direct empirical significance concerns the identity of physical states and is thus intimately related to the very concept of physical state.

The present contribution builds on a recently proposed framework by Hilary Greaves and David Wallace (Greaves and Wallace 2014) and derives a result according to which, contrary to the claims made by Greaves and Wallace, only so-called *global* but not *local* symmetries can have direct empirical significance. Given a small number of intuitively plausible and/or uncontroversial assumptions it is shown that mathematical states that are related by local symmetries correspond to the same physical state.

The framework suggested by Greaves and Wallace is based on a distinction between subsystem and environment (mathematical) states, where state spaces  $S$  and  $E$  are postulated for the subsystem and the



environment, respectively. Elements  $u \in U$  of the universe state space  $U$  are assumed to be uniquely decomposable in terms of subsystem states  $s \in S$  and environment states  $e \in E$ . The operation of combining a subsystem with an environment state is denoted by “\*” (that is,  $u = s * e$  is the universe state that arises from combining  $s$  and  $e$ ).

The question of which symmetries  $\sigma$  have direct empirical significance is asked with respect to symmetries operating on subsystem states  $s$ . The essential idea of the present analysis is that whether  $\sigma$  has direct empirical significance depends on whether  $s*e$  and  $\sigma(s)*e$  correspond to the same physical state for all combinations of  $s \in S$  and  $e \in E$ . If they do, the symmetry  $\sigma$  has the character of descriptive redundancy and does not have any direct empirical significance. If there are  $s \in S$  and  $e \in E$  for which  $s*e$  and  $\sigma(s)*e$  correspond to distinct physical states, the symmetry  $\sigma$  does have direct empirical significance.

The central result of this paper rests on the following four assumptions:

Assumption (DES) :

A subsystem symmetry  $\sigma$  has direct empirical significance iff  $s$  and  $\sigma(s)$  correspond to different physical states for some  $s \in S$ .

Assumption (SUL) For all  $s, s' \in S$ :

$s$  and  $s'$  correspond to the same physical state iff  $s * e$  and  $s' * e$  correspond to the same physical state for all  $e \in E$  for which  $s * e$  and  $s' * e$  are defined.

Assumption (MAH)

For all  $s_1, s_1' \in S_1$  and  $s_2, s_2' \in S_2$  :

if, for all  $e \in E$  for which  $s_1 * s_2 * e$  and  $s_1' * s_2' * e$  are defined,  $s_1 * s_2 * e$  corresponds to the same physical state as  $s_1' * s_2' * e$ , then  $s_1$  corresponds to the same physical state as  $s_1'$  and  $s_2$  corresponds to the same physical state as  $s_2'$ .

Assumption (Ext)

Any local symmetry  $\sigma$  defined on the subsystem state space  $S$  can be extended to a symmetry without direct empirical significance defined on the state space  $V$  of a larger subsystem  $V \supset S$ .

Assumption (DES) captures the intuitive idea of what it means for a symmetry to have direct empirical significance. Assumptions (SUL) and (MAH) are intended to make aspects of our pre-theoretic notion of physical state precise (and are motivated along these lines). Assumption (Ext) is meant to express an element of non-controversial common wisdom among practicing physicists vis-à-vis local (e.g. gauge) symmetries.

Using (DES), (SUL), (MAH), and (Ext) it is easy to derive the result that, contrary to the claims made by Greaves and Wallace, local symmetries do not have any direct empirical significance. In other words, if the assumptions hold, mathematical states that are connected by some local symmetry transformation invariably correspond to the same physical state.

The presentation closes by considering a possible criticism of the assumption (MAH) that is formulated in a forthcoming paper in *Philosophy of Science* and rejects it as based on a misconstrual of (MAH) as an empirical assumption.

**References**

Greaves, H. and Wallace, D. (2014), Empirical consequences of symmetries, *British Journal for the Philosophy of Science*, 65:69-89.

---

*Functional Emergence of Spacetime in Quantum Gravity*

VINCENT LAM

University of Lausanne

vincent.lam@unil.ch

CHRISTIAN WÜTHRICH

University of California

wuthrich@ucsd.edu

---

The main research programs in quantum gravity tend to show that standard relativistic spacetime is not fundamental. The precise and different ways in which it is not fundamental depend on the particular quantum theory of gravity, but they all seem to suggest a radical picture according to which spacetime itself is not part of the fundamental physical ontology. This perspective raises gnawing worries about the very characterization of this non-spatio-temporal physical ontology, about the emergence of the usual spatio-temporal quantities that constitute our everyday macroscopic experience, and about the very possibility of empirical evidence, including the experimental confirmation of these theories themselves. This latter point is especially problematic: if space and time are necessary preconditions of theory confirmation in empirical science, then a theory denying the fundamental existence of spacetime undermines the very possibility of its own empirical justification. Consequently, such a theory would seem empirically incoherent. This threat of empirical incoherence has also been voiced in the context of the interpretation of quantum mechanics, in particular as an argument in favor of Bell's notion of 'local beables', which are the fundamental elements of the physical ontology that are localized in a bounded region of spacetime. According to this argument, no contact between theory and empirical evidence is possible without local beables. The worry, then, is that no contact with empirical evidence is possible without fundamental spacetime quantities.

In most of the physics literature on quantum gravity, this challenge of empirical incoherence amounts to the usual constraint of consistency with the superseded theories: in particular, any theory of quantum gravity should recover in some appropriate regime the smooth relativistic spacetime picture of the theory of general relativity. This consistency constraint is a central concern in all quantum gravity programs and may typically involve approximation and limiting procedures. In this context, the issue is a technical one. However, from a conceptual point of view, the worry is that the consistency constraint is a necessary but not sufficient condition for the challenge of empirical incoherence to be met. To many, it remains unclear in what precise sense spacetime quantities, including local beables, can emerge from a fundamental non-spatio-temporal ontology.

This contribution aims to show how the tools of functionalism can help to avoid the threat of empirical incoherence. Our central claim is that spacetime need not be fully recovered in some strong ontological sense in order to provide a ground for empirical evidence and everyday experience, but only its functionally relevant features. Just as mental states can be functionally defined by their roles executed by the underlying

ontology of neural states, spacetime can be functionally understood in terms of its roles in physical theories and these functions may be executed not by relativistic spacetime, but rather by an underlying ontology of non-spatio-temporal structures described by quantum gravity.

The first step of this strategy is the functional characterization of the relevant spacetime features, such as the metrical and inertial structure. This latter should in particular allow the functional characterization of the crucial notion of spatio-temporal localization, which is at the heart of the argument for local beables to avoid empirical incoherence. The second step involves showing that the non-spatio-temporal structures under consideration can play the right sort of functional role. The details of the functional instantiation of relevant spacetime features (in particular spatio-temporal localization) by non-spatio-temporal entities need to be worked out in concrete cases. We will focus here on two important research programs in quantum gravity: loop quantum gravity and causal set theory. We will show in these two cases how the required approximation and limiting procedures can be functionally understood such that the right sort of functional roles are instantiated. Beside the fact that the general relativistic limit involves many unsolved technical issues in both loop quantum gravity and causal set theory, we argue that the functional perspective developed here averts the conceptual issues related to the emergence of spacetime from fundamental non-spatio-temporal entities, here causal sets or spin networks (or spin foams in the covariant approach to loop quantum gravity). To the extent that these latter can be understood in the appropriate regime as being functionally related as standard spacetime quantities such as dimensionality, topology, timelike and spacelike distances, spacetime volumes, or the like, they just are (functionally) spacetime quantities in this limit. There is no further question about the emergence of these spacetime quantities and therefore no threat of empirical incoherence on this basis.

We will focus in particular on how causal sets and spin networks can functionally reproduce spacetime localization, which grounds the notion of local beables and the very contact between theory and empirical evidence. In loop quantum gravity, we will discuss the functional (and approximate) implementation of a standard spacetime lattice picture familiar to quantum field theory, which allows for localization, and from there how smooth relevant features can be functionally reproduced in the appropriate limit, such as the connection and the associated parallel transport. In causal set theory, quantities can be recovered from the fundamental causal set that approximate the dimension, topology, and distances of the approximating spacetime.

What makes the non-spatio-temporal entities described by quantum gravity concrete physical entities, rather than merely abstract mathematical ones? The standard criterion for distinguishing the concrete from the abstract relies on spacetime itself: concrete entities are in spacetime, abstract ones are not. Clearly, such a spacetime criterion is just not available for characterizing a physical ontology of non-spatio-temporal entities. An alternative characterization of concrete entities involves some notion of causal efficacy: concrete physical entities as opposed to abstract mathematical ones can be considered as causally efficacious in

some sense. Whereas it does not seem obvious how to make explicit a precise notion of non-spatio-temporal causation, we argue that some weaker functional counterpart of causal efficacy could do the job here. The non-spatio-temporal structures are concrete physical structures in virtue of the (approximate) spacetime functions they instantiate. If the physically salient emergence of spacetime has been established, the status of the non-spatio-temporal entities as concreta is secured.

---

*What Even is Explanatory Pluralism? - On the Multiple Manifestations of Explanatory Pluralism in Theory and Practice (cancelled)*

HARDY SCHILGEN

University of Cambridge

hschilgen@gmail.com

---

This paper raises the following question: How do philosophers' theories of explanatory pluralism fare with regard to their fit with current social scientific practice? After setting up a taxonomy of three major theories of explanatory pluralism to date, the paper intends to make two claims: first, that philosophers focus too extensively on instances of non-complementary pluralism; second, that non-complementary theories of pluralism fail to qualify as thoroughly pluralistic.

Pluralism is taken to be different ways in which explanatory kinds can stand in relation to each other. These explanatory kinds are assumed to come in the form of micro- and macro-explanations. Now, the relations that different explanatory kinds are said to stand in, differ depending on the theory of explanatory pluralism. On the one hand, *non-complementary* theories of pluralism hold that the different explanatory kinds (micro- and macro-explanations) stand in a non-complementary relationship in virtue of being separate and autonomous. Both micro- and macro-explanations as such are, in principle, taken to be capable of providing a complete explanation of one given phenomenon. Non-complementary theories of pluralism further split up into non-competitive (Weber & Van Bouwel, 2002) and competitive versions (Kitcher, 1991). On the other hand, *complementary* theories of pluralism claim that different explanatory kinds (e.g. micro and macro) cannot be treated as separate and autonomous, but instead have to be integrated (or: complemented) in order to provide a complete explanation of a given phenomenon. Micro- and macro-explanations taken separately, cannot fully explain any given phenomenon (Marchionni, 2008; Mitchell, 2002, 2009).

First, I intend to show that Philosophers of Science to date have focussed too extensively on instances of non-complementary pluralism (Mitchell, 2002: 56). This is problematic in that their theories of pluralism fail to accommodate instances of complementary ('integrative') pluralism that are, however, frequent in social scientific practice. I will make use of several examples from the social sciences in order to illustrate what integration of explanations means in practice. For now, a more general indication should suffice: while the micro-components of a mixed explanation (e.g., individual-level mechanisms) are contextualized by the macro-structure, the macro-components (e.g., social-structural terms) are complemented by the 'stories'

that micro-mechanisms tell. The explanatory levels are neither reducible to one another, nor can they be separated. The kind of integration one ends up with amounts to something that Kincaid referred to as ‘unity without reduction’. As a result, mixed explanations that are complementary in virtue of this very inter-level kind of integration keep two explanatory kinds within the explanation of one single phenomenon and thus qualify as explanatory pluralist. However, due to their overly narrow theories of pluralism most philosophers to date fail to recognize such instances of complementary pluralism coming in the form of mixed explanations. Some important questions are neglected because of this large-scale omission: How does the integration of explanations work? Can we stick to the traditional level-of-analysis conception, given that levels are often significantly entangled with one another? Beyond that, a renewed acknowledgment of mixed explanations may well help philosophers of science to finally recognize the empirical nature of the long-led methodological individualism-holism debate. For in mixed explanations, individualist and holist components are not mutually exclusive but complement each other within one explanatory account and are thus a matter of degree. Philosophers, however, have mainly (and unsuccessfully so) attempted to tackle the individualism-holism-debate by means of conceptual analysis aiming for a winner-takes-it-all solution in favour of either of the two (Kincaid, 2014). Instead, they should attempt non-general empirical answers.

In the second part of my paper I intend to go one step further and argue that both competitive and non-competitive versions of non-complementary pluralism do not qualify as thoroughly committed to explanatory pluralism.

Both conceptions of pluralism fail to imply genuinely pluralistic explanations. While they allow for various different explanatory kinds formulated on different organizational levels, the constituting explanations taken as such are not genuinely pluralistic explanations. It seems, however, that the representatives of non-complementary theories of pluralism can counter this criticism by simply denying that genuinely pluralistic explanations are a necessary condition for a pluralistic social science as a whole. One might well have, so the argument, a pluralist social science allowing for various different explanatory kinds explaining different social phenomena without necessarily having genuinely pluralistic explanations constituting this overall plurality.

However, I am not convinced by this response and argue that such an allegedly “pluralist overall social science” is not pluralistic, but merely non-reductionist. Simply allowing for a mutual co-existence of different explanatory kinds alone does not entail a genuinely pluralistic relation between them. More specifically I intend to defend the following two claims:

1. Non-reductionism is not a sufficient criterion for pluralism.
2. Non-reductionism is not (even) a necessary criterion for pluralism.

Thesis 1 holds, so will be argued, because there is more to pluralism than the mere non-reduction of explanatory levels. This goes against large parts of the current literature, as “philosophers who advocate pluralism (...) often see themselves as staking out a position that stands in direct opposition to reductionism” (Steel, 2004: 55). Slightly surprisingly, thesis 2 holds as well, for it is possible to show that a moderate

reductionism is compatible with a set of formative components of explanatory pluralism that most theories of pluralism could probably agree on (Steel, 2004).

### References

- Kincaid (2014) 'Dead Ends and Life Issues in the Individualism-Holism Debate', in J. Zahle and F. Collin (eds.) *Rethinking the Individualism-Holism Debate*, Switzerland: Springer International Publishing.
- Kitcher, P. (1991) 'The Division of Cognitive Labor', *Journal of Philosophy*, vol. 87, pp. 5-22.
- Marchionni, C. (2008) 'Explanatory Pluralism and Complementarity: From Autonomy to Integration', *Philosophy of the Social Sciences*, vol. 38, pp. 314-333.
- Mitchell, S. (2002) 'Integrative Pluralism', *Biology and Philosophy*, vol. 17, pp. 55- 70.
- Mitchell, S. (2009) *Unsimple Truths*, Chicago: The University of Chicago Press.
- Steel, D. (2004) 'Can a reductionist be a pluralist?', *Biology and Philosophy*, vol. 19, pp. 55-73.
- Weber, E.; Van Bouwel, J. (2002) 'Can we dispense with structural explanations of social facts?', *Economics and Philosophy*, vol. 18, pp. 259-275.



---

*Do Mechanism-Based Explanations make a Case for Methodological Individualism?*

JEROEN VAN BOUWEL

Ghent University

jeroen.vanbouwel@ugent.be

---

In the recent philosophy of social science literature, we notice an increasing support for mechanism-based social explanations. Earlier pleas for social mechanisms used to be closely linked to defenses of methodological individualism. However, more recent contributions seem to be loosening that link and developing a more sophisticated account – ascribing a less important role to micro-foundations (e.g. Ylikoski 2012, Little 2012). In this paper, I want to review the impact of the mechanism-approach on methodological individualism and draw more radical conclusions as regards the individualism/holism debate, severing the link between the social mechanisms-approach and individualism. Four steps will be taken:

(a) *We should consider more than two levels of social explanation.*

Several advocates of the mechanism-based approach to social explanations have been defending the relative explanatory autonomy of meso-level explanations (e.g. Little, 2012). This adds a welcome extra explanatory level in between the individualist micro-level and the macro-level. As such, it supersedes the dichotomous thinking in the individualism/holism debate in which there would always be an individual micro-level – which would always be the same (cf. point (b) below) – that is contrasted with a macro-level.

(b) *The levels of explanation are perspectival levels; neither absolute, nor unique.*

However, the advocacy of meso-level explanations still comes with a *microfoundations requirement*. Let us first zoom in on *microfoundations* (for the *requirement*, see step (c) below). In the philosophy of social science debate, the microfoundations are usually understood as individual-level microfoundations, see, for instance, most recent work on analytical sociology. It is presupposed that there is some comprehensive, unique, and privileged individual level, the level of individual actors (cf. Ylikoski 2012). However, microfoundations do not necessarily have to be understood in that way. They could also just be understood as looking for foundations on any lower-level, e.g., on a sub-individual level focussing on cognitive capacities and processes that might be important in explaining certain social phenomena. The latter understanding of microfoundations would be more in line with actual social scientific practice in which we notice that the specification and amount of levels of explanation is perspectival, depending on the phenomena, research approaches and explanatory interests involved. Thus, the *micro* in *microfoundations* should be understood as perspectival too, rather than absolute or unique.

(c) *Seeking for microfoundations and seeking for macrofoundations are good heuristics.*

Next, let us scrutinize the microfoundations *requirement*. This requirement stipulates “that all social facts, social structures, and social causal properties depend ultimately on facts about individuals within socially defined circumstances. Social ascriptions require microfoundations at the level of individuals in concrete social relationships.” (Little 2012, p.138) Advocates of the social mechanisms approach have often been defending that a macro-explanation would never be satisfactory, or, could only be satisfactory if a micro-level part of the social explanation was provided, e.g. Hedström and Swedberg (1998). Thus, they consider a reference to (individual actions on) the individual, micro-level as a condition *sine qua non* of a satisfactory explanation.

Daniel Little (2012) develops a different position. According to him, the microfoundations requirement should not be understood as a condition for satisfactory explanations, but rather as a form of confirmation or justification of a macro-explanation. Here as well Little takes into account the actual explanatory practice of social scientists and he avoids the ontological fallacies (i.e., mixing up ontological and explanatory issues) made by earlier advocates of microfoundations. However, Little’s requirement remains vague. It should be understood as constraining explanatory practice, but how would that exactly work? How is the microfoundations requirement operationalised (and how would it interfere with our explanatory practice)?

Petri Ylikoski (2012) ascribes a special role to microfoundations as *explanantia* of constitutive explanations. I will give counterexamples that question this alleged special role.

In short, the more recent accounts of the microfoundations requirement are more sophisticated than earlier accounts, but they still remain problematic as I will show. A fruitful role one could see for a microfoundations pursuit is as an engagement to compare one’s own explanatory practice and research approach with other practices and approaches. This might result in more interaction between different approaches through which approaches articulate themselves and their relations to others more explicitly and through which the strengths and weaknesses of the respective approaches are clarified. In this respect, we could not only encourage seeking for microfoundations as a heuristic, but, on the same basis, propose searching for macrofoundations as a fruitful heuristic. (Some use the term *macrofoundations*, but given that we think of it as something higher up, one could also use *macro-roof* or *macro-covering*.)

(d) *there are no general preference rules with respect to the level of social explanations.*

This brings me to the fourth step in which we have to draw more radical conclusions than Little and Ylikoski as concerns the impact of the social mechanisms-approach on methodological individualism and sever the link between the social mechanisms-approach and individualism. The social mechanisms approach has a lot to offer to help clarifying the explanatory reasoning going on in social science, but it cannot be used to make a case for methodological individualism. The special role ascribed to microfoundations by defenders of mechanism-based social explanations cannot be upheld. We have to conclude that there are neither general preference rules with respect to the level of social explanations, nor good reasons for a general microfoundations requirement.

I will defend that the debate on social mechanisms, microfoundations, explanatory autonomy, etc. should not be so much about developing the ultimate individualistic approach or holistic approach, but rather about understanding explanatory reasoning in social science and optimize the way in which different explanatory approaches interact, co-exist, can be integrated and/or develop some division of labour among each other, while making the best out of the strengths and limitations of the respective explanatory strategies of holists and individualists. Philosophers of social science might contribute in analyzing, visualizing and optimizing the interaction among these different approaches (as an example I will discuss Longino 2013 and some of my earlier work).

**References**

- Hedström, P. and R. Swedberg (eds.). 1998. *Social mechanisms*. Cambridge UP.
- Little, D. 2012. Explanatory Autonomy and Coleman's Boat. *Theoria* 74: 137-151
- Longino, H. 2013. *Studying Human Behavior*. University of Chicago Press.
- Ylikoski, P. 2012. Micro, Macro, and Mechanism. In *The Oxford Handbook of Philosophy of Social Science*. ed. H. Kincaid, 21-45. Oxford University Press.

---

*Modeling Inequality*

KARIM THEBAULT

SEAMUS BRADLEY

ALEXANDER REUTLINGER

MCMP (LMU)

MCMP (LMU)

MCMP (LMU)

karim.thebault@gmail.com

seamus.bradley@lrz.uni-  
muenchen.dealexander.reutlinger@lrz.uni-  
muenchen.de

---

Econophysics is a cross-disciplinary research field that applies models and modelling techniques from statistical physics to economic systems. Methodologically, econophysics is supposed to differ from conventional economic practice in that it uses the ‘paradigms and tools’ of statistical physics. Econophysicists hold that mainstream economics suffers from a number of defects that they would like to correct. Most importantly both the core principles and models of mainstream economic theory are argued to lack the evidential support of real economic data. Econophysicists suppose that models and modelling techniques drawn from an experimentally focused and mathematically sophisticated science such as statistical physics will give new, and more reliable insights. Not surprisingly, not all mainstream economists agree with such a dismal view of their science. Authors offer a sharp critique of econophysics on the grounds of some practitioners: i) redoing work which has been done within economics; ii) ignoring rigorous and robust statistical methodology; iii) assuming universal empirical regularities where there are none; and iv) using modelling techniques that are in certain senses inherently problematic or illegitimate, above all that econophysical models suffer from completely unjustifiable strong idealisations. Considering this final line of criticism in the context of econophysics models of income inequality is the main focus of this paper.

The hallmark of econophysics models is their success in capturing certain ‘stylised facts’ found within economic systems. Simple physics-inspired models can reproduce important distributional features of economic systems, such as the scale freedom of price fluctuations in financial markets, or the ‘power-law tail’ of the distribution of monetary income or wealth in populations. Perplexingly, recovery of the income (or wealth) ‘stylised fact’ can be achieved within extremely simple, and heavy idealised, econophysics models of monetary exchange. Drawing on analogies with statistical mechanics, these ‘kinetic exchange’ models of income or wealth distributions model economic agents as zero-intelligence particles who bump into each other and exchange money between them at random – in many respects just like the molecules in a gas. Despite painting an idealised picture of economic interactions that is quite far removed from reality, these simple statistical physics- inspired models are remarkably successful at capturing the broad features of the distribution of income within populations. The critique of kinetic exchange models found in the literature is primarily a methodological attack regarding the idealisations involved in the models: it is not the accuracy of these models in recovering real data that is in question. Rather, kinetic exchange type models for inequality

are argued to be inherently problematic or illegitimate on the grounds that their treatment of (for example) production, income and transactions is in conflict with 'economic reality'. In our paper we will assess the warrant of the criticisms drawing both upon the philosophical literature on modelling and idealisation, and upon the notion of a 'maximum entropy explanation'.

One important aspect of our paper, is a comparison between the various idealisations made in the econophysics model, and in the statistical physics model that inspired it. We will highlight how the idealisations can be justified in the two contexts of statistical mechanics and econophysics, and in so doing, rebut some of the criticisms aimed at this model in the literature. This serves as a platform to discuss some general lessons about the importance of background knowledge in justifying idealisations. More specifically, our analysis will involve a comparison between the respective idealisations relating to i) binary interactions; ii) conservation principles; and iii) the exchange dynamics. In each case the similarity in idealisation between the income and gas model will be contrasted with the difference in mode of justification. For instance, whereas in the gas case binary interactions are a legitimate approximation to certain density regime, for inequality, we argue that the binary interaction idealisation is only justified in terms of the model's retention of an explanatorily relevant factor (entropy maximisation) – for this reason we argue that the idealisation in question is a minimalist idealisation.

Although the idealisations involved in kinetic exchange models of inequality must be justified in different terms to those involved in kinetic exchange gas models, there can therefore still exist legitimate methods for their justification. To this end, we will consider in detail the foundations of explanations via maximisation of entropy, drawing upon recent work in the foundations of statistics and in philosophy. We will point to a number of conceptual problems relating to the employment of the notion of 'entropy' within economics, and critically examine the putative justification of the idealisations found in kinetic exchange income models via entropy maximisation. Our paper will thus have both specific implications for the debate regarding kinetic exchange models of inequality, and wider implications for analysis of models and idealisation in econophysics and beyond (including but not limited to a critical discussion of understanding econophysical explanations as structural mechanistic explanation). Our hope is to offer guidance with regard to both the practice of modelling inequality, and the inequality of modelling practice.

---

*Cooperative Game Theory, Philosophy and the Social Sciences*

---

STEFAN WINTEIN

Erasmus University Rotterdam

stefanwintein@gmail.com

CONRAD HEILMANN

Erasmus University Rotterdam

heilmann@fwb.eur.nl

---

**1. Introduction**

In comparison with other formal frameworks, cooperative game theory has been a largely neglected tool in philosophy, economics, and the social sciences alike. Investigations into cooperation, the evolution of norms and other interactive or interdependent problems and social phenomena often proceed without taking a cooperative game theory perspective. In this paper, we focus on one such case of neglect, namely theorising about fairness. We criticise recent work on fairness, such as Broome (1990) and Curtis (2014), and show how cooperative game theory both exposes and rectifies their shortcomings.

**2. Fairness and Bankruptcy**

Bankruptcy. A firm goes bankrupt and its liquidation value of 900 ('the estate') has to be divided amongst Ari, Betty, and Claire. Ari and Betty have a claim of 400 each and Claire has a claim of 200. Together, their claims add up to 1,000, so the three claims cannot be satisfied in full. How should we, in order to be fair, divide the estate?

Bankruptcy is a *fair division problem*: when in charge of allocating a limited amount of an available good, how do you ensure that you treat everyone fairly? Moreover, Bankruptcy is a fair division problem with a very specific structure, which is that of a claims problem.

**Definition** A *claims problem*  $C := (E, N, c)$  consists of an estate  $E > 0$ , a set of agents  $N = \{1, \dots, n\}$  and a claims vector  $c \in \mathbb{R}^N$  specifying a claim  $c_i$  for each agent  $i$  such that  $\sum_{i \in N} c_i > E$ .

Recently, Curtis (2014) has provided a new theory of fairness that spells out what he maintains Broome's (1990) well-known theory of fairness has left out: an actual method that describes 'how to be fair'. Curtis (2014:47) remarks that to the best of his knowledge 'neither Broome nor anyone else had laid down a theory of precisely what one must do in order to be fair'. However, Curtis' theory of fairness is only applicable to claims problems and fair division in such problems has been studied both *axiomatically* (see Herrero and Villar 2001 and Thomson 2003 for overviews) and from the perspective of *cooperative game theory* (see e.g. O'Neill 1982 or Curiel et al. 1988). The main point of this paper is to invoke this axiomatic and game-theoretic work on fair division in order to lay bare and rectify certain shortcomings in Broomean theories of fairness in general and in Curtis' theory in particular. In some more detail, we will argue as follows.

### 3. Fairness does not imply proportionality

Curtis argues that in a claims problem, the estate should be divided by division rule P, i.e. proportional to the claims. In fact, Curtis claims that P can be derived from a general principle of fairness (FC) that he adopts. In the axiomatic approach to claims problems, various division rules (including P) are characterized by means of sets of logically independent properties which, intuitively, can be interpreted as capturing some aspect of what it means to be fair. We invoke this literature to argue that FC does not imply P and, more importantly, that it is problematic to claim that P follows from more elementary properties that jointly characterize fairness.

### 4. Not all fair division problems are claims problems

Curtis' theory is restricted to claims problems, which echoes the Broomean thought that fairness is *only* concerned with claims. But now consider the following problem.

Gloves. A(nn) owns 1 left glove and 2 right gloves, B(ernie) own 3 left gloves and 1 right glove and C(harlie) owns 0 left gloves and 1 right glove. At the market, *pairs* of gloves can be sold for 1 Euro and each pair that is offered for sale is actually sold. All three agents cooperate and jointly earn 4 Euros. How to divide this amount fairly amongst A, B and C?

Gloves is a fair division problem but not a claims problem, or so we argue.

### 5. Fairness beyond claims problems: cooperative game theory

We want to divide fairly in Gloves but can seek no guidance from a Broomean theory of fairness. Cooperative game theory helps. Let us first list the earnings  $v$  for all different combinations of cooperating agents:

$$\begin{array}{cccc} v(\emptyset) = 0 & v(\{A\}) = 1 & v(\{B\}) = 1 & v(\{C\}) = 0 \\ v(\{A, B\}) = 3 & v(\{A, C\}) = 1 & v(\{B, C\}) = 2 & v(\{A, B, C\}) = 4 \end{array}$$

The abstract description of Gloves as given by above is the *cooperative game* induced by Gloves. More generally, a cooperative game consists of a characteristic function  $v$  which specifies the value that each group of cooperating agents (or coalition) can *guarantee* itself. In the literature, various *solution values*<sup>1</sup> have been proposed, all of which describe how the value of the grand coalition in a cooperative game should be divided (fairly) amongst its members. Hence, cooperative game theory can guide us when we want to divide fairly in situations that are outside the scope of Broomean accounts of fairness.

### 6. Claims problems in cooperative game theory

Interestingly, cooperative game theory has also been used to analyse claims problems. An essential part of such analysis is the cooperative game  $v^C$  that is induced by a claims problem C and that is defined as follows<sup>2</sup>:

<sup>1</sup> E.g. the Shapley value (cf. Shapley 1953) the  $\tau$ -value (cf. Tijs 1981) and the nucleolus (cf. Schmeidler 1969).

<sup>2</sup> Remember that, in a cooperative game,  $v(S)$  represents the amount that coalition S can guarantee itself.

$$v^C(S) = \max\{0, E - \sum_{i \notin S} c_i\} \quad \text{for each } S \subseteq N$$

Given a claims problem  $C$ , we may thus divide the estate  $E$  directly by applying a division rule (such as  $P$ ) to  $C$  or we may divide  $E = v^C(N)$  indirectly, by applying a solution value to the claims game  $v^C$ . A division rule  $r$  coincides with a solution value  $\phi$  when applying  $r$  to  $C$  coincides with applying  $\phi$  to  $v^C$ . Several division rules coincide with a solution value and are “doubly justified” in this sense<sup>3</sup>. However,  $P$  is not one of them. Even though the lack of double justification may not discredit  $P$  as a fair division rule, we think that this lack has to be addressed by anyone who advocates  $P$  as *the* fair division rule.

### References

- Aumann, R.J. and M. Maschler. 1985. Game Theoretic Analysis of a Bankruptcy Problem from the Talmud. *Journal of Economic Theory* 36: 195-213.
- Broome J. 1990. Fairness. *Proceedings of the Aristotelian Society* 91: 87-101.
- Curiel, I.J., M. Maschler and S.H. Tijs. 1988. Bankruptcy games. *Zeitschrift für Operations Research* 31: A143-A59.
- Curtis, B.L. 2014. To be Fair. *Analysis* 74:47-57.
- Herrero C. and A. Villar. 2001. The three Musketeers: four classical solutions to bankruptcy problems. *Mathematical Social Sciences* 42: 307-28.
- O’Neill, B. 1982. A problem of rights arbitration from the Talmud. *Mathematical Social Sciences* 2: 345-71.
- Schmeidler, D. 1969. The nucleolus of a characteristic function game. *SIAM Journal of Applied Mathematics* 17(6): 1163-70.
- Shapley, L.S. 1953 . A Value for n-person Games. In *Contributions to the Theory of Games, volume II*, eds. H.W. Kuhn and A.W. Tucker, 307-17. *Annals of Mathematical Studies* v. 28, Princeton University Press.
- Thomson, W. 2003 Axiomatic and game-theoretic analysis of bankruptcy and taxation problems: a survey. *Mathematical Social Sciences* 45: 249-297.
- Tijs, S.H. 1981. Bounds for the core and  $\tau$ -value. In *Game theory and mathematical economics*, eds. O. Moeschlin and D. Pallaschke, 123-32. North-Holland Publ Cie.

---

<sup>3</sup> For instance, the Talmud rule (T) was (implicitly) proposed as a fair division rule in the Talmud. Around 2000 years later, Schmeidler (1969) proposed the nucleolus as a fair solution value for cooperative games, and Aumann and Maschler (1985) showed that T coincides with the nucleolus.



## Symposia &amp; Contributed Papers IX

**Theory Choice meets Social Choice****Symposium**

Organizer: Alexandru Marcoci &amp; James Nguyen

Chair: Christian J. Feldbacher

Room 5F, Saturday 13:30 – 15:30

*Arrow's Theorem and the Rationality of Scientific Theory Choice*

SAMIR OKASHA

Bristol University

*samir.okasha@bristol.ac.uk**Can there be Neutral Choice Procedures in Science?*

MICHAEL MORREAU

The Arctic University of Norway

*michael.morreau@uit.no**On the Rationality of Theory Choice*

ALEXANDRU MARCOCI

LSE

*a.marcoci@lse.ac.uk*

JAMES NGUYEN

LSE

*j.nguyen1@lse.ac.uk**Evaluating Competing Theories via a Common Language of Qualitative Verdicts*

WULF GAERTNER

University of Osnabrück

*wulf.gaertner@uni-osnabrueck.de*

NICOLAS WÜTHRICH

LSE

*n.wuethrich@lse.ac.uk***General Description**

Scientists are often faced with multi-criterial choice situations. When evaluating theories, models or hypotheses they have to weigh up how well each competitor fares with respect to multiple scientific virtues. A trade off ensues. Perhaps they will choose a simpler theory (or model or hypothesis) despite the fact it doesn't fit the data as well as one of its competitors. Or perhaps

they will choose the theory with the broadest scope at the expense of the simplest one. And so on. Kuhn (1977) argued that there is no unique algorithm to do this. Even scientists who agree with one another about which virtues are important might still disagree about how well each member of their choice set fares 'all-things-considered'. And as such, theory choice isn't a fully rational enterprise. How to understand this, and the implications it has, are central questions within the philosophy of science. If theory choice fails to be rational, what sort of stance should we adopt to what our theories tell us about the world?

Okasha (2011) sparked a resurgent interest in the question of theory choice by placing the problem in the context of social choice theory. Scientific virtues are identified with voters who provide ordinal rankings of the theories under consideration: e.g.  $T_1$  is simpler than  $T_2$  is simpler than  $T_3$ , whereas  $T_1$  is more accurate than  $T_3$  which in turn is more accurate than  $T_2$ , and so on. The five Arrowian conditions are then taken to be desirable in the context of theory choice: a theory choice function should be defined over a universal domain; it should respect irrelevance of independent alternatives, and unanimity; no virtue should be dictatorial; and it should always deliver a transitive overall ordering. Arrow's Impossibility Theorem tells us that no such function exists. Okasha argues, pace Kuhn, that it is not that there is no unique theory choice function, but there is no function whatsoever.

In this symposium we investigate the prospects of rational theory choice in light of Okasha's argument, and explore in detail some implications of treating theory choice as a social choice problem. Okasha's and Morreau's contributions offer a debate about the applicability of universal domain in the context of theory choice. The condition is motivated in social choice by noticing that voters could hold different preference rankings, and a social choice function should be able to account for this. It would be undemocratic to rule out an individual's preference ranking *a priori*. But in the context of theory choice, at least some virtues appear 'rigid': if simplicity ranks Copernican astronomy above Ptolemaic astronomy, it could not have done otherwise. And as such certain profiles are ruled out from the domain of the theory choice function.

However, it is well known that impossibility results also hold even in more restricted domains. So the question is whether or not the rigidity of certain virtues suffices to avoid Okasha's threat. In light of this, Morreau considers strong neutrality as a further condition that we might want a theory choice function to respect. Roughly speaking, strong neutrality requires that different pairs of alternatives be treated in the same way. But if a theory choice function should respect this condition then another impossibility result looms, even in restricted domains.

Marcoci and Nguyen investigate the notions of rationality that Okasha's argument threatens. The non-existence of a theory choice function that satisfies the Arrowian conditions precludes certain ways of construing theory choice as a rational enterprise, but it does not preclude them all.

For example, we can restrict our focus to a specific individual profile, and ask whether or not a theory choice function is rational with respect to that profile. They explore two ways of developing this observation. Firstly by introducing a quantitative notion of rationality, where a theory choice function is rational to the extent that it is likely to deliver a transitive ranking over a given domain. Secondly, they introduce an, as yet unexplored, alternative notion of rationality, which they call neighbourhood rationality. The idea is that although theory choice functions might deliver intransitive results when applied to specific profiles in the universal domain, as long as there is an alternative profile 'nearby' which the function maps to a transitive ranking, scientific rationality is saved. They appeal to Kuhn's claims regarding the ambiguity of scientific virtues as motivating the shift from an intransitive to a nearby transitive profile. Such a map of what 'rationality' means in the context of theory choice clears up precisely what Okasha's *prima facie* threat threatens.

Gaertner and Wu<sup>thrich</sup> argue that Arrow's result can be avoided in the context of theory choice by introducing a scoring rule defined over a set of qualitative intervals for every epistemic value. Each virtue will rank each competing theory on a scale from 'insufficient' to 'very good', let's say. The introduction of the scoring rule provides a common language for each of the virtues to score competing theories, and as such it provides a common cardinal scale within which theory choice is to take place. They demonstrate that with this scoring rule in place, all five of Arrow's conditions are met (in a specific sense to be discussed), but the impossibility result no longer goes through.

The symposium will thus both further existing debates in the literature, and add new perspectives. Okasha's (2011) suggestion of construing theory choice as a social choice problem opened the door for a wealth of interactions between philosophers of science and choice theorists. The former have been interested in what makes theory choice rational, but have only recently begun to draw on the precise models of rationality developed by the latter. All of the contributions in this symposium import significant results from social choice theory into the context of theory choice, and all of them explore questions beyond Arrow's impossibility result. Furthermore, the discussions suggest a 'two-way' interaction between theory choice and social choice. Just as philosophers of science can use the tools from social choice theory to think about theory choice, the contributions to this symposium are of clear interest to choice theorists as well. For instance, Okasha and Morreau's debate over restricted domains concerns the limits of impossibility results; Morreau's contribution shows what the introduction of strong neutrality entails; Marcoci and Nguyen consider ways of saving rationality when faced with intransitivity; and Gaertner and Wu<sup>thrich</sup>'s introduction of a common language for qualitative verdicts leads to an interesting possibility result. The symposium will thus be appealing to both camps: theory choice meets social choice meets theory choice again.

**Abstracts****1. Samir Okasha: *Arrow's theorem and the rationality of scientific theory choice***

The problem of theory choice, in which scientists must choose between competing theories, is structurally quite similar to the problem of social choice, in which 'society' must choose between competing social alternatives. This structural parallel suggests that it could be fruitful to use formal results from social choice theory, such as Arrow's impossibility theorem, to shed light on theory choice in science. In a *Mind* 2011 paper, I examined the social choice / theory choice relationship in detail, arguing that Arrow's theorem poses a *prima facie* threat to the rationality of theory choice in science, but one that in certain contexts can be met. In a forthcoming reply to my paper, Michael Morreau (forthcoming) criticizes my argument on the grounds that Arrow's universal domain condition does not apply to standard cases of theory choice, so there is no *prima facie* threat. Here I reflect further on this debate between Morreau and I, and tentatively explore the idea of using single-profile, rather than multi-profile, social choice theory to shed light on scientific theory choice.

**2. Michael Morreau: *Can there be neutral choice procedures in science?***

Multi-criterial choice problems can be analyzed within Kenneth Arrow's (1951, 1963) framework for studying democratic social choice. The trick is to let criteria play the role of people. Multi-criterial choice differs substantively from social choice, though, and the requirements are different. For one thing, criteria unlike people do not have to be treated as equals. Therefore, multi-criterial choice need not be *anonymous*. Also, the possible criterial rankings might be known to be quite limited, and then there is no sense in demanding that choice procedures have *unrestricted domains*. These are among the requirements implicated in Arrow-style "impossibility" theorems. So the differences cast the possibilities for multi-criterial choice in an optimistic light. The focus in this talk will be on choice among models, hypotheses or theories by scientific criteria such as accuracy, simplicity and scope. *Strong neutrality* requires that different pairs of alternatives be treated the same way. In social choice this condition expresses an unethical "Welfarism", but it seems defensible in theory choice. It seems to say no more than that choice procedures are, in a sense, principled. This casts a gloomier light. Strong neutrality gives rise to impossibility results that are relevant even with highly restricted domains. It will emerge that these results do limit the possibilities for theory choice to some extent, certainly in certain "toy" examples, where domains can be shown to have the necessary *richness*.

**3. Alexandru Marcoci and James Nguyen: *On the rationality of theory choice***

Okasha's gambit of treating scientific virtues as voters allows him to import Arrow's Impossibility Theorem into the context of scientific theory choice. Assuming that a theory choice function (TCf) should meet the Arrovian conditions, Arrow's result immediately tells us that no such function exists. And this is said to threaten the rationality of theory choice.

In this paper we suggest differentiating between different versions of scientific rationality. If rationality requires a TCf that delivers a transitive ranking from *every profile* in *every domain* (or at least domains where two or more virtues are used to choose between two or more theories), then theory choice is irrational. If, for a given domain (same caveat), rationality requires a TCf that delivers a transitive ranking from *every profile* then theory choice is irrational with respect to that domain. But the rationality of a TCf can be analysed on a more fine grained level. We can ask, *for any given profile* whether or not a TCf is rational. This leads us to suggest a quantitative notion of rationality that turns on *how likely* it is for a TCf to deliver a transitive ranking. We suggest that in the domains relevant to theory choice (a relatively low number of virtues choosing between a relatively low number of theories), simple theory choice functions are highly rational.

We further investigate an, as yet unexplored, alternative notion of rationality that is particularly relevant in the context of theory choice. We call this *neighbourhood rationality*. A TCf is said to be neighbourhood rational if and only if any profile which it maps to an intransitive ranking contains a profile 'in its neighbourhood' that the function maps to a transitive ranking. We provide an exact mathematical framework in which to study the neighbourhood rationality of a TCf. What neighbourhood rationality corresponds to in the context of theory choice is intuitively compelling: ambiguity of scientific virtues. Our claim is that, as long as scientific virtues are sufficiently ambiguous, even qualitative all-or-nothing rationality is saved. Whenever we are faced with irrationality, we shift to an alternative profile in the neighbourhood of the original one, a shift justified by the ambiguity of a scientific virtue. We demonstrate that this can be done for the domains relevant to theory choice.

So whether or not Okasha's claim threatens the rationality of theory choice depends on the notion of rationality in question. There are plausible notions of scientific rationality that are not precluded by Arrow's result.

**4. Wulf Gaertner and Nicolas Wüthrich: *Evaluating competing theories via a common language of qualitative verdicts***

Scientists face situations in which they have to choose among competing theories. Kuhn (1977) claimed that several algorithms can be defended to select the best theory based on epistemic

values such as simplicity and accuracy. In a recent paper, Okasha (2011) reformulated Kuhn's discussion in an Arrowian social choice framework and stated that no theory choice algorithm exists which satisfies a set of intuitively compelling conditions. In this paper, we propose a solution to avoid the impossibility result. Based on Gaertner and Xu (2012), we suggest reconstructing theory choice with the help of a general scoring rule defined over a set of qualitative intervals for every epistemic value. The basic idea of the general scoring function we propose can be best captured by contrasting it with the Borda method. The Borda method is such that if there are  $m$  different alternatives and if all individuals are assumed to have strict orderings over the alternatives, then rank  $m_1$  is assigned to the top-ranked alternative,  $m_2$  to the second highest rank, and so forth. The suggestion that we wish to make is different from the Borda approach in at least two important aspects. First, we propose a common cardinal scale within which all individual evaluations have to take place. This scale, ranging from "very good" to "insufficient", let's say, establishes a common language of judgment to which all evaluating persons have to agree. Second, each and every agent is assumed to have their own scoring function. In Borda's scheme, all persons have the same scoring function, assigning strictly descending grades to the linearly ordered alternatives. In our model, some individuals may not announce such a descending sequence, thereby not filling all ranks from top to bottom. Coming back to Kuhn's problem, replace alternatives by competing theories and replace individual agents by criteria. The result of our procedure is attractive. Given a finite number of epistemic values and three or more alternative theories, the aggregation method we propose yields a complete and transitive ranking and the rule satisfies unrestricted domain, weak Pareto, non-dictatorship, and independence of irrelevant alternatives.

## References

- [1] Arrow, Kenneth. (1951). *Social Choice and Individual Values*, New York: John Wiley.
- [2] Arrow, Kenneth. (1961). *Uncertainty and the Welfare Economics of Medical Care*, *The American Economic Review*, 53:5, 941-973.
- [3] Gaertner, Wulf and Xu, Yongsheng (2012). *A General Scoring Rule*, *Mathematical Social Science*, 63:3, 193-196.
- [4] Kuhn, Thomas. (1977). *Objectivity, Value Judgement and Theory Choice*, in his *The Essential Tension*, University of Chicago Press, 320-329.
- [5] Morreau, Michael. forthcoming. *Theory Choice and Social Choice: Kuhn Vindicated*, *Mind*.
- [6] Okasha, Samir. (2011). *Theory Choice and Social Choice: Kuhn versus Arrow*, *Mind*, 120:477, 83-115.
- [7] Okasha, Samir. forthcoming. *On Arrow's Theorem and Scientific Rationality: Reply to Morreau and Stegenga*, *Mind*

**Local vs. Global Approaches to Realism****Symposium**

Organizer: Juha Saatsi

Chair: Ludwig Fahrbach

Room 5D, Saturday 13:30 – 15:30

---

*Forget Perrin (cancelled)*

PAUL DICKEN

University of New South Wales

email@pauldicken.com

---

*Should the Debate over Scientific Realism go Local?*

LEAH HENDERSON

Carnegie Mellon University

leahh@andrew.cmu.edu

---

*Kinds of Evidence for Realism: Revisiting the Case of Atomism*

STATHIS PSILLOS

University of Western Ontario, University of Athens

spsillos@uwo.ca

---

*A Case for Local Realism*

JUHA SAATSI

University of Leeds

j.t.saatsi@leeds.ac.uk

---

**General Description**

The issue of scientific realism remains very central in philosophy of science, and the debate over it has by now a long and complex history. It has become clear that tackling the realism question requires a multi-faceted approach with significant input from the history of science. However, important questions remain over exactly how the debate should proceed. In particular, there has been considerable discussion in recent years over whether the debate should primarily take a local or a more global approach. On the one hand, traditional arguments such as the No Miracles

Argument and Pessimistic Induction operate at a relatively global level. They are based on an attempt to gain a general sense of the reliability of the scientific method and use that information to modulate our epistemic confidence in theories produced by that method. On the other hand, some philosophers have argued that the realism debate should be pursued at a more local level by focusing on case-specific evidence and considerations, rather than on the global arguments. The aim of this symposium is to explore the question of the appropriate level of generality for conducting the realism debate. In part this issue is important because it will help to see exactly what kind of information needs to be extracted from the history of science, so that the energy put into historical investigations can be directed appropriately.

The symposium will focus on three key areas.

**i) Assessing new objections to global arguments for and against realism.**

In the history of the realism debate, there has been considerable discussion of possible defects of the global arguments for and against realism, such as the charge that the No Miracles argument is circular. Discussion of many of these issues is ongoing. However, in recent years new objections to these arguments have arisen which turn specifically on the idea that they are pitched at too high a level of generality.

Our symposium will explore these new challenges—summarized below—and the prospects of different realist responses.

**Base rates.** Arguably global realism must be framed probabilistically in order to capture the fact that no necessary connection exists between truth and the positive realist ‘indicators’ (predictive success; employing a particular method; whatever). That is, realism can at best establish a *high likelihood* of the (approximate) truth of a theory exhibiting the relevant indicators. But thus construed, realism arguably faces an insurmountable problem of inaccessible base-rates: in arguing that the indicators thus reliably indicate truth one must rely on assumptions concerning the proportion of true theories in the relevant population—an unknowable base-rate (Howson 2000, 2013; Magnus and Callender 2003).

Realists have responded to the base-rate problem in various ways, but no consensus has emerged (Psillos 2006, 2011; Worrall 2012). The significance and the interpretation of this problem, as well as the appropriate realist response, remain significant points of contention.

**Realist arguments and theory of induction.** Arguably global realist arguments are associated with universal theories on inductive reasoning in general. By contrast, local realism is naturally affiliated with a local theory of induction according to which the goodness of an ampliative inference is underwritten by case-specific material assumptions, instead of deriving from some universal template of reliable inductive inferences (Norton 2003, 2010). Problems with universal theories of induction have been taken to present a challenge also to global realism, in as far as it



aims to establish the reliability of a particular form of inductive reasoning (e.g. IBE) in the sciences (Saatsi 2009; see also Norton 2011).

***Realism as naturalistic philosophy of science.*** Arguably global realism can be in tension with its naturalistic aspiration to be properly informed by first-order scientific practices. This tension takes different forms. For one, the no-miracles argument is arguably in tension with first-order explanatory practices (Frost-Arnold 2010). For another, the fact that first-order scientific investigations are highly domain-specific in their methodologies is arguably in tension with the way in which global realism abstracts away from the particular contexts of theorizing (Dicken 2013, Dicken *forthcoming*).

**ii) Assessing the prospects of purely local approaches to realism.**

The challenges addressed in (i) motivate a move towards more local, case-specific realist and anti-realist arguments, concerned with individual scientific theories, in contrast with the sweeping arguments of the global debate. But an important task is to clarify what form such a local debate over realism would take. Can the realism debate be based on local, case-specific considerations without any support from more general arguments that abstract away from case-specific details? Would a completely local realism debate add anything of substance to science itself or would it amount to a mere repetition of scientific analysis of the first-order evidence?

The symposium will discuss the potential for realist and anti-realist arguments to be framed locally, and will address whether this constitutes a more productive direction for the realism debate to take. In part, this will be done by paying close attention to a key case study of Perrin's argument for the reality of molecules.

**iii) Clarifying the relationship between global realist arguments and first- order local scientific evidence.**

For those who still think that the global arguments should play a role in the scientific realism debate, engagement with the local challenge prompts a reconsideration and clarification of how the global arguments interact with first-order local scientific evidence.

Questions that the symposium will address include: How do the levels interact in particular cases? What is the relative importance of the interacting levels? Are the answers to these questions a context-sensitive matter or is a general account possible?

The local-global debate is closely related to another deep issue for the realism debate, which is to what extent it relies on distinctively philosophical considerations, as opposed to considerations already implicit in scientific evaluation of theories. Our discussion will present different points of view on this relationship.

The symposium presents the first focused examination of these interrelated issues by philosophers each of whom has recently contributed to the debate between local and global approaches to realism. The use of the Perrin case-study provides a significant unifying theme to the symposium.

## Abstracts

### 1. Paul Dicken: *Forget Perrin* (cancelled)

Traditional arguments in the scientific realism debate have concerned the overall reliability of our scientific methods. Recent literature in the debate has been particularly concerned with a statistical fallacy that appears to be common to such arguments—a failure to acknowledge the base-rate probability of an arbitrary scientific theory being true. A number of philosophers have responded to this worry by distinguishing between those global arguments that consider the reliability of our scientific methods in general, and those local arguments that target the reliability of specific scientific theories. Despite much interesting work on this issue there has yet to emerge a clear understanding of what is meant by a local argument for scientific realism, and how such reasoning can be differentiated from the first-order scientific practice with which it is supposedly concerned.

In this talk, I challenge one influential account of how first-order scientific practice can nevertheless constitute a philosophical argument for scientific realism—the case of Perrin’s argument for the reality of molecules. I argue that the philosophical conclusions that have been repeatedly drawn from Perrin’s work are in fact best understood as traditional, global arguments for scientific realism, and that the most plausible diagnosis for the interest shown in Perrin’s case is best explained in terms of the base-rates fallacy that it is supposed to resolve. Finally, I suggest that talk of a base-rate fallacy is misleading, since the worry is not that we manage to ignore a relevant piece of statistical information, but that such information is simply unknown. The problem is better seen as an instance of the more general problem of establishing our priors; and that within this framework some of the more recent appropriation of Perrin’s work is shown to be especially problematic.

### 2. Leah Henderson: *Should the debate over scientific realism go local?*

The No Miracles Argument (NMA) and the Pessimistic Induction (PI) are often regarded as pillars of the scientific realism debate. They are instances of what might be called ‘reliability-based’ arguments. According to these arguments, information from the history of science about the reliability of the scientific method plays an important role in determining our confidence in current well confirmed theories. Recently there has been an influential attempt to reorient the debate away from the reliability-based arguments and to ‘go local’. Localizers say that the reliability-based

arguments are pitched at the wrong level of generality – they are too ambitious, too general, too sweeping. Rather, localizers argue that a more productive way to conduct the realism debate is to focus on simply looking carefully at the empirical evidence that supports each scientific theory on a case-by-case basis.

This paper considers what the scientific realism debate would be like if it were conducted locally, by providing a close examination of the canonical case of Achinstein's local treatment of Perrin's experimental arguments for the existence of atoms. I argue that the local approach removes the key resources which give substance to the realism debate and also deprives the debate of its naturalistic status. I urge that if the realism debate eschews the reliability-based arguments, it will become rather barren, rather than more productive.

Nonetheless, the local challenge does give new urgency to the question of how exactly the reliability-based arguments figure into the overall assessment of our epistemic attitude towards the approximate truth of a theory. I will discuss how this challenge might be met.

### **3. Stathis Psillos: *Kinds of evidence for realism: revisiting the case of atomism***

I have argued in my recent work that a central philosophical issue in the scientific realism debate concerns the balancing of two kinds of evidence there is in favour of or against scientific theories. The first kind—what I have called first-order evidence—comes from the actual empirical and theoretical considerations (empirical data, predictions, explanatory successes, theoretical virtues etc.) that make scientists adopt, or object to, a particular scientific theory. The second kind of evidence—what I have called second-order evidence—comes from, by and large, philosophical considerations concerning science as a whole; in particular, the historical track-record of scientific theories and the existence or possibility of empirically equivalent rivals. In the present paper I will advance further this strategy by examining in detail how the two kinds of evidence played out in the case of the acceptance of atomism in the beginning of the twentieth century.

I will start by presenting the main philosophical (second-order) arguments against atomism that were proposed by Duhem, Ostwald, Stallo and Poincare and stress that these were not taken to be enough to debunk atomism unless a) first-order scientific evidence was summoned against it and b) an alternative theory (in this case, energetics) was shown to be superior (on first-order evidence) to atomism. I will then show how Perrin's work on the Brownian motion offered overwhelming first-order evidence in favour of the atomic hypothesis, which was taken to outweigh second-order philosophical objections to atomism. Finally, I will focus on Poincare's acceptance of the atomic hypothesis and argue that the best defence of realism in general has to be based on a combination of two kinds of argument, one based on first-order evidence in favour of particular scientific

theories and another based on more global, philosophical and second-order, considerations that are sensitive to the history of science and the track-record of scientific theories.

#### **4. Juha Saatsi: *A case for local realism***

In the first half of the paper I assess some recent challenges to the famous no-miracles argument for scientific realism. I will argue that the most serious challenge—largely unnoted in the literature—turns on the fact that the argument is *global* in a particular way. The relevant sense of global (as opposed to local) argument is usefully characterized in terms of the distinction between formal and material justifications of induction (Norton 2003, 2010). There are general problems with formal justifications of induction—exemplified in the case of the No Miracles argument—which can be used to motivate the idea that realism should be argued for by (more) local arguments that turn on (more) case-specific, material justifications of the relevant inductive inferences.

In the second half of the paper I explore the nature of such local realist arguments. I will clarify different senses in which local arguments can transcend first-order scientific evidence and reasoning so as to yield bona fide philosophical arguments, by virtue of (i) making transparent the relevant material justifications operative in science, and (ii) defending on a case-by-case basis those material justifications against an anti-realist skeptic. I will also explain why the local arguments are not threatened by the base-rate challenges that have been presented to the more global arguments (cf. Dicken 2013). I will illustrate the issues at stake in reference to Perrin's argument for the reality of molecules (as discussed in Achinstein 2002; Psillos 2011a).

#### **References**

- Achinstein, P. (2002). 'Is There a Valid Experimental Argument for Scientific Realism?', *Journal of Philosophy*, 99, 470–495.
- Dicken, P. (forthcoming). 'Three Degrees of Naturalism in the Scientific Realism Debate', in Daly, C. (ed.) *The Palgrave Handbook of Philosophical Method*, Palgrave Macmillan.
- Dicken, P. (2013). 'Normativity, the base-rate fallacy, and some problems for retail realism', *Studies In History and Philosophy Science*, 44, 563-57.
- Egg, M. (2012). 'Causal Warrant for Realism about Particle Physics', *Journal for General Philosophy of Science*, 43, 259–280.
- Fitzpatrick, S. (2013) 'Doing Away with the No Miracles Argument', *EPSA11 Perspectives and Foundational Problems in Philosophy of Science*. Springer International Publishing, pp. 141-151
- Frost-Arnold, G. (2010). 'The No-Miracles Argument for Realism: Inference to an Unacceptable Explanation', *Philosophy of Science*, 77, 35–58.

- Hacking, I. (1982). 'Experimentation and Scientific Realism', in Curd, M. & Cover, J. (eds.) *Philosophical Topics*,  
Philosophy of Science: The Central Issues, W.V. Norton and Company, 1998, pp. 71–87.
- Howson, C. (2000) *Hume's Problem: Induction and the Justification of Belief*, Oxford: Oxford University Press.
- Howson, C. (2013). 'Exhuming the No-Miracles Argument', *Analysis*.
- Kincaid, H. (2000). 'Global Arguments and Local Realism about the Social Sciences', *Philosophy of Science*, 67,  
667–678.
- Kitcher, P. (2001). 'Real realism: the Galilean Strategy', *The Philosophical Review*, 110, 151–197.
- Lipton, P. (2004). *Inference to the best explanation* (2nd ed.), London: Routledge. Magnus, P. & Callender,  
C. (2004). 'Realist ennui and the base rate fallacy', *Philosophy of Science*, 71, 320–338.
- Magnus, P. D. (2003). 'Success, Truth and the Galilean Strategy', *British Journal for the Philosophy of Science*,  
54, 465–474.
- Norton, J. (2003). 'A Material Theory of Induction', *Philosophy of Science*, 70, 647–670.
- Norton, J. (2010). 'There are no universal rules for induction', *Philosophy of Science*, 77, 765–777.
- Norton, J. (2011). 'History of science and the material theory of induction: Einstein's quanta, Mercury's  
perihelion', *European Journal for Philosophy of Science*, 1, 3–27.
- Psillos, S. (2011a). 'Moving Molecules Above the Scientific Horizon: On Perrin's Case for Realism',  
*Journal for  
General Philosophy of Science*, 42, 339–363.
- Psillos, S. (2011b). 'Scientific Realism with a Humean Face', in Saatsi, J. & French, S. (ed.), *The  
Continuum  
Companion to Philosophy of Science*, Continuum Press, pp. 75–95.
- Psillos, S. (2011c). 'The Scope and Limits of the No Miracles Argument', in Stadler, F. (ed.), *The  
Philosophy of  
Science in a European Perspective: Explanation, Prediction, and Confirmation*, Springer, vol. 2,  
pp. 23–35.
- Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*, London: Routledge.
- Saatsi, J. (2009). 'Form vs. Content-driven Arguments for Realism', in Magnus, P. & Busch, J. (eds.)  
*New Waves  
in Philosophy of Science*, Palgrave, pp. 8–28.
- Worrall, J. (2012). 'Miracles and Structural Realism', in Landry, E. & Rickles, D. (eds.) *Structural  
Realism:  
Structure, Object, Causality*, Springer, pp. 77–95.

---

*Events, Quantum Mechanics and the Passage of Time*

MAURO DORATO

University of Rome Three

dorato@uniroma3.it

---

Cosmology, astrophysics, geology, evolutionary biology and our perception of time essentially refer to events that happen in temporal succession. If this succession sufficed to provide a robust notion of becoming, and if the primitive ontology of our fundamental physical theories were about events, the latter could play a key role in grounding a mind-independent passage of time, and therefore in bridging the gap between the physical and the manifest image of time. In this paper I discuss three major difficulties raised by this project and conclude with a note of skepticism.

The first difficulty originates from unresolved metaphysical controversies about the nature of events: without at least a working definition, how can we decide whether the fundamental ontology of any physical theory is about events? The second problem is raised by the interpretation-dependence of the ontology of quantum mechanics (QM). The third concerns the definability of a robust notion of becoming in standard or modified Minkowski spacetime (STR), even granting that the difficulties raised by the second objection could somehow be overcome.

In order to concentrate on the last two problems, here I will simply take for granted that any reasonable metaphysical theory of events must presuppose that they be “sufficiently” localized in spacetime. In order to discuss the other two objections, this minimal assumption will suffice.

In fact, independently of its interpretation-dependence, the ontology of non-relativistic QM does not seem to involve just measurement *outcomes* (i.e., events), but also the *wave function* and, in any case, *quantum fields*. On the one hand, the hypothesis that the ontology of QM is fundamentally about events in temporal succession can add some additional precision and generality to the claim that the “primitive ontology” of QM is spatiotemporal (Allori et al 2008, Goldstein and Zanghì 2013). In fact, both the so-called GRW’s “flashes” and the worldlines traced by Bohmian particles can be subsumed, respectively, under an ontology of *discrete events* and of *continuous processes*. On the other hand, however, the “density of matter” version of GRW seems inimical to an ontology of fundamental events. The “contractions” of the density of matter in certain spacetime regions should be regarded as events, because they are changes *in* the preexisting matter field. However, this account seems to imply that events are not fundamental,

since they turn into local *properties* exemplified by a *substance* (the field), as in Kim's view (1976). Furthermore, since a field is typically considered to be a *global* assignment of quantities to spacetime regions, even if events were identified with such regions (Quine 1985) or properties thereof (Lewis 1986), they would be coextensive with them, no matter how big, and therefore possibly with the whole of spacetime.

Furthermore, these three theories presuppose the event-friendly view that the wave function is either non-fundamental or purely "nomological". On the former hypothesis, both configuration-space realism and many-worlds would make events derivative. On the latter hypothesis, the nomological character of the wave function can be cashed out in three ways: in terms of a Humean pattern of nonlocal matters of facts (Darby 2012, Esfeld 2013), in terms of an holistic power exemplified in all of spacetime (Esfeld et. al. 2013), or in terms of a primitive posit. On Humeanism, the wave function supervenes on local matters of fact at certain spacetime regions, which can be regarded as *events*. On dispositionalism, the quantum state associated to the wave function would be a global, holistic power, and therefore events would be just its manifestation. On primitivism, the quantum state would be co-extensional with spacetime, and therefore non-localized.

Finally, even if the prospects for an ontology of events were more promising in quantum field theory (QFT) – in particular within AQFT (Haag 2013, Pashby 2014) – in such a way as to overcome, at least partially, QM's interpretation dependence, two additional difficulties would remain. The first is that the standardly assumed fundamentality of fields over localized particles pushes towards extendedness in spacetime, and therefore against localizability as a necessary condition for eventhood. The second is that the *causal* criteria for the identity of Davidsonian events, invoked by (Bartels 1999) in order to give an account of what happens in laboratories – run the risk of being circular, as Davidson himself later recognized.

Coming now to third objection, suppose to identify physical events with spacetime regions or points: the "addition" of new regions to the growing block (Broad 1938, Tooley 1997), or of new "nodes" to the causets (Sorkin 2007, Earman 2008, Dowker 2014, Callender and Wütrich 2015) would be an addition of events. However, this addition (or the attrition of branches as in Pooley 2013) would not amount to a qualitative change *in* the already existing universe, but rather to a *sui generis*, *absolute* change in what tenselessly exists (Broad's "coming to pass"). However, while a qualitative notion of change (a field having different values in different spacetime regions) is uncontroversial, the kind of becoming defined as an "addition" is purely *relational* and not absolute. Analogously, absolute changes given by "attritions" are not described by current physics and therefore cannot be regarded as bona fide physical events, even though the "tree models" of reality purport to represent them.

A similar skeptical conclusion applies to attempts to ground becoming on the invariant partial order given by temporally successive events in *standard* Minkowski spacetime (Savitt 2002, Dieks 2006, Dorato 2006). The relationality in this case means that given any point or region  $R$ , events in the future light cone of  $R$  have not become (or occurred), while those in the past light have become (happened). However, labeling such an occurrence “becoming” amounts to solving the problem by *fiat*, since a more plausible reading of this “occurring” is that the former events *are* in the future light cone of  $R$ , the latter *are* in the past of  $R$ . Since physics can hardly identify these regions  $R$  as Privileged, it cannot describe a genuine passage of time, even if its fundamental ontology were just about events.



---

*Do We Need a Primitive Ontology to Make Quantum Mechanics Empirically Coherent?*

MATTHIAS EGG

University of Bern

matthias.egg@unil.ch

---

Empirical support for any scientific theory comes from observation of things and events in space and time. Hence, if a theory makes no room for such entities (which John Bell famously called *local beables*), it might undermine its own empirical basis and thereby face the threat of what Jeffrey Barrett has called *empirical incoherence*. Some authors have argued that this is the case for quantum mechanics, unless we supplement it with local beables at the fundamental level (also called a *primitive ontology*). Put more formally, this argument for introducing a primitive ontology runs as follows:

- (1) A fundamental theory without local beables is empirically incoherent.
- (2) Quantum mechanics without a primitive ontology has no local beables.
- (3) Quantum mechanics without a primitive ontology is empirically incoherent.

The most detailed version of this argument (though without mention of the terms “empirical incoherence” and “primitive ontology”) was given by Maudlin (2007). My paper will start by questioning Maudlin’s defence of premise (2). On this basis, I will then argue that Ney’s (2015) response to Maudlin is doubly unjustified, firstly in its sympathy for premise (2), secondly in its rejection of premise (1).

The crucial question underlying premise (2) is whether local beables can be *derived* within a version of quantum mechanics that does not postulate them at the fundamental level. Maudlin admits that this might be possible in principle, but he thinks that present attempts to do so lack a clear rationale to regard the derived structure “as *physically* salient (rather than merely mathematically definable)” (Maudlin 2007, 3161). In response, Huggett and Wüthrich (2013, 283-284) point out that physical salience can be assessed “from above”, that is, by examining which theoretical structures yield correct empirical predictions. I think this is correct as far as it goes, but it does not completely dispel Maudlin’s worry. The example he discusses in this context concerns the two different types of local beables that can be associated with the GRW formalism: a matter field (GRWm) or flash-like events (GRWf). Since GRWm and GRWf are empirically equivalent, choosing between them “from above” is impossible. This is a familiar problem of underdetermination, but it is here combined with a less familiar one: not only is the choice between GRWm and GRWf underdetermined by the empirical evidence, but it is also

underdetermined by the underlying fundamental theory (GRW without local beables, called GRW0).

However, there is no reason why this twofold underdetermination should be any more worrying than the usual one we face in quantum mechanics anyway. Whoever wants to be a realist about quantum mechanics must opt for one of its versions, based on their non-empirical virtues. This is true for the primitive ontologist (who thinks of GRWm and GRWf in terms of *fundamental* ontology) as well as for the wave function realist who tries to *derive* local beables from GRW0. Therefore, insofar as underdetermination does not prevent us from realism about fundamental ontology, it should not prevent us from realism about derivative ontology either.

Still, one might find premise (2) appealing, because the wave function of quantum mechanics does not seem to be the kind of entity from which local beables could emerge. I will discuss one specific instance of this reasoning, namely Ney's (2015, 15) claim that the wave function could not play the functional role of a three-dimensional object such as a macroscopic pointer. This is directed against the wave function realist's appeal to functionalism, most prominently worked out by David Albert. The curious thing is that Ney cites Albert's claim that the wave function's dynamics (encoded in the Hamiltonian of the system) "plays the causal role constitutive of there being multiple classical particles in a three-dimensional space" (Ney 2015, 11), without specifying what is wrong with that claim. But if nothing is wrong with it, then why shouldn't the potential term in the Hamiltonian be such that these particles form a bound state that constitutes a pointer capable of interacting with other objects in just the way ordinary pointers do? To be sure, it would be hopelessly complicated to actually write down such a Hamiltonian, but *this* is not a specific problem of wave function realism; it confronts the primitive ontologist in precisely the same way.

Despite her sympathy for premise (2), Ney seeks to defend wave function realism against the charge of empirical incoherence by rejecting premise (1). The claim that empirical coherence presupposes local beables is based on our pre-theoretical beliefs about evidence, and these, she argues, should be replaced by what our best scientific theories tell us about the nature of evidence. But aren't our best scientific theories those which are best supported by empirical evidence? If so, the project of *first* appraising our scientific theories *and then* having our beliefs about evidence informed by them is incoherent.

Even setting this problem aside, Ney's (2015, 18) proposed reconceptualization of "evidence" by directly linking the wave function to a state of the world "that is properly described (nonexhaustively) as 'Theorists have acquired evidence for theory T'" does not look promising. It is even more problematic than the idea (criticized by Maudlin 2007, 3158-3159) that physical theories should make predictions about our experience. In order to do that, a theory would have to solve the mind-body problem. This would not suffice in the case of Ney's proposal, since there

are not even any bodies on her view. What her theory would have to solve is the *mind-wavefunction problem*, that is, the challenge of connecting the quantum state of the universe directly to mental states, without passing through the intermediate step of first connecting it to some local beables (pointers, observer's brains etc.), which can then be connected to mental states. Neither Ney nor anyone else has given us any idea how this is supposed to work.

### References

- Huggett, Nick and Christian Wüthrich (2013). "Emergent spacetime and empirical (in)coherence". *Studies in History and Philosophy of Modern Physics* 44, 276--285.
- Maudlin, Tim (2007). "Completeness, supervenience and ontology". *Journal of Physics A* 40, 3151--3171.
- Ney, Alyssa (2015). "Fundamental physical ontologies and the constraint of empirical coherence: a defense of wave function realism". *Synthese*. DOI 10.1007/s11229---014---0633---9.

---

*There are No Mathematical Explanations*

JAAKKO KUORIKOSKI

University of Helsinki

jaakko.kuorikoski@helsinki.fi

---

According to a widely accepted view, explanation is about tracking objective dependency relations (Pincock forthcoming, Woodward 2003, Ylikoski and Kuorikoski 2010). Although this view has been developed furthest in the case of causal explanation, many have recently proposed that the general idea is also applicable to various non-causal forms of explanations (e.g., Huneman 2010; Kuorikoski 2012), including explanations that many take to be essentially mathematical (Pincock forthcoming; Saatsi and Pexton 2013). In this paper I argue that if we accept the general idea of objective dependence as the basis of explanation, there cannot be mathematical explanations. What appear to be mathematical explanations are either highly abstract mechanistic explanations or reconceptualizations of the *explanandum* phenomenon in which mathematics as such does not have an explanatory role.

Marc Lange (2013) takes the distinguishing feature of mathematical explanations to be their modal strength: truly mathematical explanations show how the *explanandum* could not have been otherwise due to mathematical necessity (which is stronger than causal or nomological necessity). Although it is easy to see how mathematical necessity can provide modal information (and even answers to what-if-things-had-been-different-questions), it is harder to square with the idea that explanations show what the *explanandum* depend upon; how can the *explanandum* depend on anything, if it could not have been otherwise by mathematical necessity? Not surprisingly, Lange opts for a modal conception of explanation instead of an ontic dependence conception. Christopher Pincock, however, argues that in such cases (what he calls abstract explanations), there is an explanatory dependence between the *explanandum* (an abstract feature of a phenomenon) and a more abstract entity/structure, of which the *explanandum* is a special case. Such relations of abstract dependence are arguably objective matters of fact that science (mathematics) investigates. In a similar fashion, Saatsi and Pexton argue that cases of what they call geometric explanation of regularities (e.g., between scaling exponents and dimensionality) trace objective relations of dependence between abstract properties.

Even if one accepts that there are objective explanatory relations of dependence other than causation, I argue that there are serious difficulties regarding the concept of mathematical or abstract dependence. The main problem lies in providing truth conditions for the relevant

counterfactuals. As Lange notes, relations between mathematical entities are modally much stronger than relations of dependence between entities in space-time. The counterfactual case in which the *explanans* is different (the abstract object in question had different properties that it in fact has) is either inconsistent or definitional of some other abstract object. This is problematic regardless whether one entertains a realist metaphysics of abstract or mathematical entities or not. On the one hand, realist metaphysics is more congenial to the idea of true ontic dependence between abstract entities, but the epistemology and metaphysics of such dependencies are even murkier than the corresponding problems related to the mere existence of abstract objects. On the other hand, if one is non-realist with respect to abstract objects, then one face the challenge of explaining how abstract dependencies meet the requirement of *ontic* dependence, i.e., the requirement that explanatory relationships relate things in the world in contrast to merely epistemic relations of dependence between our representations of things in the world. This contrast is essential for the ontic view and motivates many of the features of Woodward's theory that are often wrongly attributed as being definitional of causal explanation (for example, the requirement that the relevant counterfactuals have to be same-object-counterfactuals, cf. Pincock forthcoming).

I argue that most examples of non-causal mathematical explanations presented in the literature (Huneman's topological explanations, Saatsi and Pexton's geometrical explanation, Pincock's explanation of the Plateau's laws for soap films) are to be interpreted as constitutive mechanistic explanations, albeit very abstract ones. The relevant explanatory dependency is that of constitution, a suitably ontic relation between the whole and its parts, analyzable with only minimal alterations to Woodward's theory.

What comes to Lange's mathematical explanations demonstrating the mathematical necessity of the *explanandum*, I argue that these cannot be reconciled with the requirement of ontic dependency, and that they are therefore not explanations. In such cases, the *explanandum* is reconceptualized in such a way which shows that there was, in fact, nothing to be explained to begin with. What is perceived as an explanatory advance with respect to phenomena is an increase in what I call *formal understanding*, i.e., better understanding of our systems of reasoning and representation. Formal understanding can be given a similar, broadly inferentialist, analysis as explanatory understanding in terms of answers to what-if questions, but the crucial difference is that it is about our tools of reasoning, not the world itself. (I argue that this is also a better analysis of Pincock's example of the derivation of the Plateau's laws from minimal sets.) This difference in the object of understanding answers Lange's and Pincock's challenge of providing an independent rationale for ignoring expert intuitions about what is and what is not an explanation.

**References**

- Huneman, P. (2010). Topological Explanations and Robustness in Biological Sciences. *Synthese* 177 (2):213-245.
- Kuorikoski, J. (2012). Mechanisms, Modularity and Constitutive Explanation. *Erkenntnis* 77 (3):361-380.
- Lange, M. (2013). What Makes a Scientific Explanation Distinctively Mathematical? *British Journal for the Philosophy of Science* 64 (3):485-511.
- Pincock, C. (forthcoming). Abstract Explanations in Science, *British Journal for the Philosophy of Science*.
- Saatsi, J. & Pexton, M. (2013). Reassessing Woodward's Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations. *Philosophy of Science* 80 (5):613-624.
- Ylikoski, P. & Kuorikoski, J. (2010). Dissecting Explanatory Power. *Philosophical Studies* 148 (2):201–219.
- Woodward, J. (2003) *Making Things Happen*. OUP.

---

*Presentism meets Black Holes again*

GEURT SENGERS

Erasmus University Rotterdam

geurtsengers@gmail.com

---

Presentism may still be the metaphysics preferred by common sense, but in the philosophical literature it has long been forced into the defences. In a recent publication in the *European Journal for Philosophy of Science* [1], Romero and P´erez add to its troubles. As in some earlier attacks on presentism, general relativity theory is at the heart of the issues. But where at least some of the earlier troubles were at least a modal step away (think of Gödel universes), the authors promise to bring it close to home by pinpointing black hole geometries as the culprit.

Romero and P´erez argue the only viable choice for a present in the presence of a (Schwarzschild) black hole is its horizon. Since distant events at different times are co-present with different parts of this horizon, it would spell trouble for the presentist if he is indeed forced to think all events on the horizon necessarily co-present.

The arguments that drive the authors to this conclusion are derived from the behaviour of lightcones near the black hole horizon. In familiar diagrams of the behaviour of these lightcones in Schwarzschild coordinates, these cones are seen to close up (from the outside in) and flatten out (from the inside) in a way that at the horizon aligns them with it. This would seem the presentist who wants to postulate a hypersurface between these cones no wriggling room, and force his choice to be the horizon itself. As a further argument forcing this choice, it is maintained that this hypersurface is the one orthogonal to the local time direction. Furthermore, the authors argue this to be a coordinate independent feature.

I will argue these arguments to be flawed. To begin with, one should not draw conclusions on the basis of Schwarzschild coordinates where they are known to become degenerate. The alignment of the cones is among the well known coordinate artefacts of the Schwarzschild system. Secondly, the argument based on orthogonality to the local time direction is confused. (Briefly, this would only hold for the direction of the Schwarzschild time axis, and no local observer could identify his temporal axis with that. I will also discuss another possible reading, and argue it to be flawed as well.) I will continue by pointing out that the authors arguments for coordinate independence do not establish the desired conclusion. (Briefly, lightlikeness of the horizon is indeed an invariant, but the needed flattening of the lightcones is not.) Finally, descriptions of the geometry in coordinates that do not become degenerate at the horizon show that it is indeed

possible to choose a spacelike foliation such as the presentist needs for his cause, and I will offer an explicit example. Whatever else the troubles of presentism, the Schwarzschild horizon does not seem to be among them.

In closing I will highlight some further issues arising in black hole geometries that indeed might spell insurmountable problems for presentism, namely the possibility of closed timelike curves in metrics of the Kerr family. Thus I hope to offer a more balanced picture of the impact of black hole physics on presentism.

**References**

- [1] Gustavo E. Romero, Daniela P´erez, "Presentism meets black holes", *Euro Jnl Phil Sci* (2014) 4: 293-308,  
DOI 10.1007/s13194-014-0085-6.



**General Philosophy of Sciences VIII****Contributed Papers**

Chair: Richard Dawid

Room 5E, Saturday 13:30 – 15:30

---

*Kuhn's Revolutions*

VASSO KINDI

University of Athens

vkindi@phs.uoa.gr

---

In the present paper I criticize a typical understanding of Kuhnian revolutions which I think is wrong. I have two aims: the first is exegetical, namely to show what Kuhn's revolutions were all about and the second conceptual, i.e., to show what Kuhn has contributed to our concept of revolution. I argue, against several commentators (including Tom Nickles, Brad Wray, Dan Garber, Stephen Toulmin) that Kuhn's model of science does not offer a narrative of scientific development that can be tested for its truth or falsity. Rather Kuhn's *extended* concept of revolution functions as a philosopher's tool to highlight diversity in the practice and history of science in order to undermine a particular philosophical picture which stressed uniformity and linear cumulative progress. Revolutions before Kuhn also marked discontinuity, yet they were thought to lead to progress which is not the case with Kuhn who invoked the concept of incommensurability.

Thomas Kuhn's *The Structure of Scientific Revolutions* (2012) is commonly taken to be a philosophy book which proposes a certain model for scientific development that is based on historical evidence, but also as a book of historiography which attempts either to give a particular shape to the past of science in the manner of philosophical history or to read off from historical data a pattern of scientific activity in the course of time. So, the pattern '*paradigm-normal science-anomaly-crisis-extraordinary science-revolution-new paradigm*' is supposed to depict accurately the different phases science undergoes.

In line with this understanding of the book, and independently of whether *Structure* is understood as philosophical or historiographical, critics and advocates alike have tried, and are still trying, to assess and test this schema empirically, to determine what particular historical entity (theory, tradition, discipline or practice) does or does not qualify as paradigm, to settle which historical episodes deserve to be called revolutions, whether Kuhnian revolutions occur in certain scientific disciplines such as mathematics or the life sciences, whether crises always precede revolutions, whether there are communication breakdowns, as Kuhn maintained, between adherents of incommensurable paradigms, whether there are indeed radical differences between concepts or whether there is continuity or gradual discontinuity in the history of science. For

instance, Dan Garber in his recent article “Galileo, Newton and All That: If It Wasn’t a Revolution What Was It?” (*Circumscribere* 7, 9-18, 2009), concentrating on the minutiae of particular revolutions, namely, the 17th century scientific revolution, maintained that “the simple model of a political revolution, or the model of paradigm-crisis-new paradigm that is at the heart of Kuhn’s *Structure* doesn’t really fit the world.” Garber thinks that “the model of a revolution seems strangely ill-suited to capture what went on in the seventeenth century”. (ibid.)

All these attempts completely disregard and do not care to make sense of Kuhn’s statement in the 90s that his model does not really need history and can be derived completely a priori. They do not have a coherent account of what Kuhn does in *Structure* in relation to history, e.g., whether his model is poorly evidenced, “unevidenced” or a priori. Some think that his evidential basis is very limited, others claim that he draws from it the wrong conclusions. They do not notice or care to consider that not only in the Postscript to *Structure* but already in *Structure* Kuhn speaks of big and small revolutions that may affect just the members of a scientific subspecialty. His critics have no explanation for that and prefer to criticize Kuhn’s model for supposedly having limited application to only major shifts in scientific development. They equally have no explanation for the fact that Kuhn explicitly renounces the conflation of history and philosophy. They prefer to accuse him of inconsistency when they think that he combines the two or to fault his model as inadequate when they notice that he does not use the concepts and categories of *Structure* in his historical research. In their view, the absence of paradigms, puzzles, crises, incommensurability of concepts, of standards or of perceptions from Kuhn’s historical work, speaks against his model: it is inapplicable to history.

I will present a different understanding of Kuhn’s revolutions. I will argue that in *Structure* Kuhn aimed at challenging the dominant view of science and developed the concepts and categories that would serve this purpose. *Paradigm*, *normal science* or *revolution* are not labels which are supposed to name concrete empirical phenomena in the history of science but concepts which serve as tools in order to deliver the point that science is not the cumulative enterprise that was taken to be at the time Kuhn wrote *Structure*. These concepts can certainly be applied, and are supposed to apply, to empirical phenomena by historians or philosophers but they were not developed to provide the uniquely accurate description of the history of science. They are supposed to provide an interpretive schema in which and by which scientific practice and development can be made sense of. The concept of revolution in this schema -and its correlative concept of incommensurability-, is important not because it identifies correctly particular events as revolutionary (this is a task that belongs to the historian) but because it draws attention to discontinuity rather than continuous accumulation in the history of science.

In the course of developing my interpretation, I consider and criticize relevant recent work by, among others, Tom Nickles and Brad Wray.

---

*How are Mechanistic Explanations Understood?*

PHYLLIS ILLARI

University College London

phyllis.illari@ucl.ac.uk

---

There are two important areas of debate in philosophy of science which concern scientific understanding, broadly construed. First, various major authors in the mechanisms literature claim that good mechanistic explanations are ‘intelligible’, although they do not say a great deal in detail about what they take this to mean. Secondly, the ‘Contextual theory of scientific understanding’ of de Regt and Dieks (2005) claims that intelligibility is a value that scientists in a particular community at a particular time confer on theories they can use. In the first place, accounts of mechanisms have been developed using paradigm cases from the life sciences such as the mechanisms of protein synthesis; while in the second place, de Regt and Dieks draw paradigm cases from physics. This, and different framing of ideas, means that it is not altogether clear how the contextual theory might be applied to understanding gained from mechanistic explanations.

The aim of this paper in brief is to examine what it means to understand a phenomenon mechanistically, by applying the core insights of the contextual theory of scientific understanding to existing work on mechanistic explanation, supplemented by the work of Sabina Leonelli on embodied knowledge.

This account will be developed by discussing our understanding of the mechanisms of supernovae, as a way of bringing together an account of mechanisms developed using paradigmatic examples from the life sciences with the contextual theory of understanding which was developed with the domain of physics in mind. It will build on the work of Illari and Williamson (2012), which argues that the mechanisms of supernovae do fit the account of mechanisms and mechanistic explanation that is becoming consensus in the mechanisms literature:

A mechanism for a phenomenon consists of entities and activities organized in such a way that they are responsible for the phenomenon. (Illari and Williamson, 2012, p120.)

This account clearly draws heavily on work of the major mechanists, Bechtel, Craver, Darden, Glennan and Machamer, as Illari and Williamson explain. Therefore I will also use this consensus account alongside the case of supernovae to argue that the contextual theory of scientific understanding can be used to show what it means to render a phenomenon intelligible by giving a mechanistic explanation. Examining the case of supernovae means that we see the mechanism for a purely physical phenomenon being understood in the context of understanding other things,

including physical laws, and general models, and I will make some comments on how this context might apply also to the life sciences.

I will begin, in section 2, by bringing the claims of mechanists together with the contextual theory of scientific understanding to build a theoretical account of understanding phenomena using mechanistic explanations. In section 3, I will develop this account by applying it to understanding the mechanisms of supernovae. It will turn out that the mechanisms of supernovae are particularly interesting for this purpose, as they enable a creative synthesis between the original one-theory- one-phenomenon assumption of the contextual theory, and the extreme diversity and locality that is typical of the life sciences and generates the focus on locally described activities and entities. As a teaser, consider the following from a textbook on astrophysics: 'Astrophysics does not deal with a special, distinct class of effects and processes, as do the basic fields of physics. ... astrophysics deals with complex phenomena, which involve processes of many different kinds. It has to lean, therefore, on all the branches of physics, and this makes for its special beauty. The theory of the structure and evolution of stars presents a unique opportunity to bring separate, seemingly unconnected physical theories under one roof.' (Prialnik, 2010, p28-9.) This section will show that even in the important branch of physics that is stellar astrophysics, there are a variety of things in addition to theories that are developed and used for a variety of different purposes. While section 3 will concentrate on theoretical understanding, in section 4 I will ask whether this is sufficient for an account of understanding stars, relying on work by Sabina Leonelli arguing for the importance of not merely theoretical knowledge but also what she calls 'embodied' knowledge to scientific understanding. While I will say little about the history of science, I will comment here on the work of Peter Dear (2008) on the history of what has been taken to be intelligible in science, particularly on what has often been regarded as an opposition of theoretical and instrumental aims. Some very significant differences between the mechanical philosophy of Descartes and Boyle, and the new mechanists, will become clear throughout the paper. There are many more that I will not have space to examine, with the extent of the differences in the developments made by new mechanists particularly exemplified by Bechtel (2010).

I will finish by drawing some wider conclusions, in section 5. While one major aim of the paper is to develop an account of what it means to understand a phenomenon mechanistically, with reference to stars, a second aim is also to offer a useful extension to the contextual theory of scientific understanding, and argue that Leonelli's work can be extended to apply to physics. While I will focus on how scientists understand phenomena, in line with both the contextual theory of scientific understanding, and comments by major mechanists, I will also comment in the conclusion on how to avoid potentially dubious implications of the view defended in this paper for work on the public understanding of science.

---

*From Classical Mechanics, to Special Relativity Theory, and Quantum Mechanics—Or: Why Structural Realists would Profit from Studying Structural Continuity by Means of Conceptual Spaces*

GEORGE MASTERTON

Lund University

geroge.masterton@fil.lu.se

FRANK ZENKER

Lund University

mail@frankzenker.de

PETER GÄRDENFORS

Lund University

peter.gardenfors@lucs.lu.se

---

A viable structural realist position in philosophy of science should be clear on the questions (1) what theoretical ‘structure’ is, and (2) how continuity of structure can be judged. But the answers in the literature remain abstract and are rarely rich enough to describe scientific practice adequately. We argue that (1) conceptual spaces provide a rich framework for identifying the structures of scientific theories; (2) by studying the types of changes in the underlying conceptual space that occur when one theory is replaced by another, the continuity in structure becomes more apparent than the tools typically used by structural realists allow.

We work with three physical theories presented in their phase-space formulations: Classical Mechanics (CM), Special Relativity Theory (SRT), and Quantum Mechanics (QM). A theory’s phase-space does not exhaust its conceptual space, but the phase-space constitutes its most central part. By treating the transitions from CM to QM and from CM to SRT in their phase-space formulations, the similarities between the conceptual spaces become evident, and allow for an initial assessment of the degree of similarity between their theoretical structures.

As our comparison makes (painfully) clear, judgments of conceptual similarity and continuity depend on assumptions that remain largely unaffected by, and hence are prior to, applying conceptual spaces; they must therefore be established (or criticized) on independent grounds. This result can go some way towards explaining varying intuitions as to whether CM is more similar to SRT than SRT is to QM, for instance. Phrased more positively, whether a given historical transition constitutes a mild (“conservative”) or a more radical (“revolutionary”) change thus turns into a better defined disagreement.

***Conceptual spaces as structures***

Conceptual spaces are not parts of a symbolic system with a syntactic or logical structure, but rather geometric structures that can be analyzed into their constitutive dimensions and properties. An empirical theory always presupposes a specific conceptual framework that provides the magnitudes, or *dimensions*, on which the formulation of this theory depends. The topological and

metrical structures of such magnitudes are tightly connected to the variety of methods by which these magnitudes are measured, as well as to (strict or probabilistic) inferential relations to other concepts that are thought to obtain in the given theory.

Dimensions are said to be integral if, to fully describe a quality, a value must be assigned to each dimension; dimensions that are not integral are said to be separate. A theory domain can now be defined as a set of integral dimensions that are separate from all other dimensions in a theory. That dimensions are separable, however, does not imply that they are separate; rather, dimensions may be integral despite the fact that two or more measurements of their values do not interfere with each other. Indeed, dimensions may be integral simply because they individually fail to constitute a domain.

As the example of physical space suggests, it is part of the meaning of 'integral dimensions' that they share a metric, that is, all distances between any two (different) points  $x$  and  $y$  located on the plane spanned by the integral dimensions  $D_1$  and  $D_2$  are expressible as multiples of the distance between the point of the origin and either of the points  $x$  and  $y$ . We argue that the separability of dimensions is closely connected with the commutation relations specified by the Lie algebra of a theory. For instance, the Poisson bracket  $\{x, px\}$  equals zero in CM, so it does not matter whether one measures position first and then momentum, or the other way around. That is, measurements of position and of momentum along the same axis do not interfere with each other in CM, and so the position and momentum dimensions on each axis are separable in the theory. Indeed, position and momentum along the  $x$ -axis are also separate in CM. But such separability is absent in QM, and hence the very same dimensions are integral in that theory.

### ***Why structural realists might care***

The application of conceptual spaces allows tracing *continuities* in scientific conceptual frameworks, which provide evidence against the "revolutionary" interpretation of scientific change famously proposed by Thomas Kuhn, an account which unduly assumes the primacy of the symbolic level of representation and *volens nolens* remains wed to a realist semantics.

Something similar seemingly holds for many proponents of structural realism, who tend to compare the structure of empirical theories via their symbolic representations. The hope here is that a relevant *continuous* structure may be represented by, say, the symbol  $S$  in theory  $T$ , and by the symbol  $S^*$  in theory  $T^*$ . But the structure that is thus represented must somehow differ from the symbolic representation itself, for it is only the latter entity that can be said to have undergone change if *structural* continuity shall nevertheless obtain in the course of  $T$  having been being replaced by  $T^*$ , irrespective of how continuity and replacement are best defined. Thus, while

structural realists do seek to *get* their hands onto theory structures, all they seem to end up *having* in their hands are symbolic representations thereof, rather than the structures themselves.

Compared to Abelian groups and Lie algebras, which are the favorite examples of structure used by structural realists, our representational framework is far less abstract and closer connected to scientific practices. We can therefore claim that employing conceptual spaces may help structural realists get their hands onto the structures they are after. Moreover, our five-typed categorization of inter- and inter-theory changes offers a finer grain than Kuhn's distinction between normal and revolutionary change, and so provides a richer toolbox to study both theory changes as well as changes to the associated conceptual framework.

This as background, we work with the phase-space formulations for classical mechanics, quantum mechanics, and relativity theory. Although these major physical theories superficially differ from one another when they are viewed in their standard guises, a great deal of continuity between them readily appears when we view them in their phase-space formulations. Using the phase-space descriptions of CM, QM and SRT thus provides a way of representing these theories in terms of conceptual spaces: the phase-space of a physical theory being the conceptual space of its basic domains.

***Example: Transition from CM to QM***

The transition from CM to QM, for instance, involves no change in the set of basic dimensions nor is there change in the relative importance of the non-basic dimensions, and also the phase-space geometry remains Euclidean. In their phase-space formulations both theories are, at their core, geometrically 6D Euclidean cotangent bundles linearly ordered in time, with the same set of derived dimensions. But there are two related changes of the underlying structure. The first one is that in CM momentum and position, energy and time, etc. are respectively separable domains; while in QM they are not. The second change in the transition from CM to QM is the deformation of the Lie algebra of observables, the classical Poisson bracket being replaced by the Moyal Bracket in QM. Hence, when going from CM to QM, there is equivalence between the *novel* impossibility of precise co-determination of observables in measurement and the changes in the Lie algebra of observables.

This exemplifies how intra-theoretical changes in the dependence of observables on phase-space location can also lead to inter-theoretical changes in the separability of dimensions. Furthermore, in QM we have a perfect demonstration of how structural changes of the latter kind are intimately connected with principled limitations of our measurement procedures. Notice, too, that momentum and position being no longer separable in QM implies that phase-space is a domain in this theory: that is, we can no longer consider phase-space to be composed of two



separable position and momentum domains as we can in CM, SRT and GRT. Hence, there is a change in the dependence of dimensions on phase-space location, which in this case also results in a change the separability of certain dimensions. But in going from CM to QM no change occurred to underlying geometry or dimensionality of the basic dimensions, and the derived dimensions retain their original level of importance.

Such insights, we claim, can assist structural realists in making their case for structural continuity in theory change.

---

*Explaining Complex Dynamics by Structural Mechanisms*

MEINARD KUHLMANN

University of Mainz

mkuhlmann@uni-mainz.de

---

There are different options for fixing the identity criteria for causal mechanisms. One option is to cash out the notion of mechanisms in terms of physical processes. Quite obviously, it wouldn't be very attractive to individuate a mechanism by a particular physical process in space and time, because then each mechanism would occur only once, making scientific generalizations impossible. Thus we should say that a mechanism can be specified by certain kinds of physical processes. Nevertheless, even then this approach has the incurable defect that it "obscures similarities between kinds of interactions among higher-level entities" (Glennan 2002, 346). If the interactions involved in mechanisms are understood as material processes, then tokens of interactions cannot be recognized as tokens of one common type of higher-level interaction because different physical instances of one type of interaction may be vastly diverse on the lower levels.

Thus it can be essential not to characterize mechanisms in terms of (fundamental) physical processes even though interactions between parts of a mechanism supervene upon physics. In other words, although higher-level interactions and thereby higher-level mechanisms are ultimately ontologically determined by the underlying physics, higher-level mechanisms are explanatorily autonomous. Describing a higher-level mechanism purely in terms of physical processes can even destroy its explanatory power. In order to identify the mechanisms one needs to abstract from its material manifestations. From these considerations we learn that it can be important to specify and individuate a mechanism by decomposing a given system into parts that fulfil certain functions – i.e. we need a "functional decomposition" (Bechtel and Richardson

2010) – where it is irrelevant for the identification of a mechanism how the function of one of its parts is realized physically.

What I have said so far has nothing specifically to do with complex systems. However, when it comes to identifying an apparently common mechanism across radically diverse complex systems, such as ferromagnets and financial markets (e.g. see Mantegna and Stanley 2000), the identity criteria have to be yet more abstract. My thesis is that we need to focus on certain structural features. I want to claim that, for instance, phase transitions in ferromagnets and financial markets can be studied in a common framework because the same structural mechanisms can be invoked in both cases. I propose to distinguish two different classes of structures in structural mechanisms, namely (i) structural start and boundary conditions, and (ii) emerging dynamical structures. If one has identified a structural mechanism, then one knows that a certain set of structural start and boundary conditions (i) is essential for producing certain dynamical structures (ii). What one needs in order to be sure that one has really found a mechanism and not just an artefact is the fulfilment of a certain robustness condition.

Structural start and boundary conditions may concern connectivity, dimensionality, topology and certain symmetry properties. Connectivity is the most important structural aspect. The crucial issue in complex systems is the interaction between the system's parts, and neither their detailed behavior nor their minute spatio-temporal organization in the whole system. What really matters is the dynamical interactive organization of a complex system, and even there only certain structural aspects. A conventional mechanistic explanation shows how the often sequential interactions of the different parts, which fulfil specific functions, produce a certain behavior. In complex systems it is usually impossible or at least not helpful to distinguish parts with different functions that play specific stable roles in the mechanism. Mostly all parts have identical properties and behavior. What is essential instead are the structural features of their interaction. Moreover, the parts of a complex system usually all interact simultaneously. Thus in contrast to conventional mechanisms, one could say that structural mechanisms in complex system standardly have an egalitarian set-up: All parts are governed by the same behavioral rules (and they may freely switch from one behavior to another), no external force tells them what to do, and they all interact at the same time. The relevant start and boundary conditions for structural mechanisms in complex systems do not describe a specific configuration that already allows imagining what will happen if we let the system run. This is a crucial difference to conventional mechanisms. In complex systems, for interesting things to happen, it suffices to have, in a sense, an amorphous set-up with very general structural properties that apply to the whole system.

The second class of structures in structural mechanisms concerns the emerging dynamical structures. Robust complex behavior lives on the fact that non-trivial long-range effects (e.g. phase transitions) arise dynamically purely on the basis of short-range interactions. The entire system behaves as if there were some external coordination, while in fact there is none. And I take it that this is one of the deep ideas behind the notion of mechanisms: Once it is set up in the appropriate way it runs largely by itself without the need for

any further coordination. But there is one pivotal issue that distinguishes conventional mechanisms from structural mechanisms in robust complex behavior—which is also the reason for the term “self-organization”: In complex systems the “organization” that is crucial for the system behavior is not already present in the initial set-up but only arises through the dynamics of the system, namely by the interaction of its parts.

Coming back to the initial issue of the identity criteria for mechanisms, it is only possible to say that there is a common mechanism in diverse systems such as ferromagnets and financial markets if one stays on the structural level. For example, one doesn’t want to claim that market traders actually sit on a grid and only interact with their spatially nearest neighbors. Rather, the crucial point is more abstract or structural: Large changes, be it phase transitions or financial-market crashes, and other related phenomena, can arise purely from the local interactions of the systems’ parts without any external coordination. One task of complex systems theories is to identify the underlying structural mechanism.

### References

- Bechtel, William, and Robert C. Richardson 2010. *Discovering complexity: Decomposition and localization as strategies in scientific research*. Second Edition. Cambridge, MA: MIT Press/Bradford Books.
- Glennan, Stuart S. 2002. “Rethinking mechanistic explanation.” *Philosophy of Science* 69: S342–S353.
- Mantegna, R. N., and H. Eugene Stanley 2000. *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge et. al.: Cambridge University Press.

---

*Extrapolation in Basic Research (cancelled)*

TUDOR BAETU

Universidade do Vale do Rio dos Sinos

tbaetu@hotmail.com

---

The use of surrogate models in clinical research raises a well known difficulty, known as the problem of extrapolation: given the differences between the surrogate systems in which the findings are actually documented and the target to which the findings are extrapolated, there is no guarantee that what is true of the surrogate must also be true of the target within a tolerable degree of approximation. The fact that many signaling, developmental, and metabolic mechanisms and pathways are conserved provides general insights about what may hold true, with some variation, of a large number of phylogenetically related organisms. Nevertheless, while such insights play an important role in guiding the discovery process, clinical researchers seek more precise answers. The goal is to figure out whether a particular result documented in the surrogate also holds true, ideally down to minute qualitative and quantitative details, of the target system. When aiming at this degree of precision, the suitability of a surrogate model needs to be evaluated on a case by case basis, where the evaluation procedure relies on an assessment of the relevant similarities between surrogate and target. It has been often argued that the most relevant similarities are those concerning mechanisms causally productive of the phenomenon of interest. Surrogate models that share mechanistic features with their targets are more likely to generate the phenomena of interest via the same causal pathways and respond in similar ways when these pathways are disturbed. By contrast, surrogates that do not share causal features might generate similar phenomena by means of different mechanisms, and therefore behave differently when subjected to similar experimental interventions.

In this paper, I want to draw attention to the largely overlooked fact that basic science is equally indebted to the epistemic practice of extrapolation from surrogate models, with all the benefits and risks this practice entails. The discovery process in basic science faces a multitude of experimental hurdles, and it is seldom the case that a single experimental setup succeeds in addressing all difficulties. The usual way around is to systematically trade one experimental difficulty for another by changing experimental setups, in the hope that a more complete knowledge of the phenomenon and its underlying mechanisms can ultimately be attained by conducting studies in a multitude of setups. One of the most striking consequences of this practice is that descriptions of phenomena and explanatory accounts—such as the diagrams of mechanisms in cell and molecular biology, including their extensions in other fields of investigation, basic or applied—are

in fact mosaic jigsaws reconstituted from bits and pieces of data gathered in distinct experimental setups. How adequately this knowledge reflects reality depends in no negligible part on the extent to which results can be safely extrapolated from one experimental setup to another.

The widespread use of extrapolations in basic science raises new challenges to anyone interested in tackling the problem of extrapolation. One immediate difficulty stems from the fact that previously proposed solutions work on the premise that sufficiently is known about mechanistic similarities and differences between surrogate and target. While such knowledge is available in later stages of research, it cannot be assumed in the initial stages, when nothing is known about the mechanistic basis of a phenomenon. Thus, if the standard solution to the problem of extrapolation relies on prior mechanistic knowledge, then the new challenge is how to address the problem in the absence of mechanistic knowledge. A second challenge stems from the fact that solutions to the problem of extrapolation are framed in relation to the justification of isolated extrapolations given an already available background knowledge, the origin of which is not put into question. This is the typical scenario in clinical research, where the task at hand is figuring out how to justify a particular extrapolation, usually one of immediate practical import, given substantial mechanistic knowledge already provided by basic science. The situation changes radically when it comes to justifying the very background knowledge from basic science, which, as it turns out, combines data from tens or even hundreds of distinct experimental setups. If there is a probability of error associated with any given extrapolation, then as knowledge increases, the number of extrapolations increases as well, and with it the probability of error. Thus, what may constitute a reasonable strategy for addressing the problem of extrapolation in the context of clinical research is unlikely to provide an equally workable way around the problem in basic research.

An analysis of scientific practice reveals that scientists treat extrapolation as a matter of taking an epistemic risk. I argue therefore that it is misleading to think that there is such a thing as a definitive, universally applicable solution to the problem of extrapolation. The relevant question is rather how to plan scientific research in such a way as to keep the possibility of error under control. As I will illustrate and discuss in the following pages, researchers in the life sciences deploy a surprisingly varied array of tactics, from model validation protocols aimed at minimizing the possibility of error for prospective extrapolations, to holistic confirmation strategies aimed at the retrospective testing of previously made extrapolations and fallback positions aimed at providing a stable epistemic basis for troubleshooting anomalies, all of which are carefully orchestrated in such a way as to ensure the overall viability of a research project. This vindicates a more mitigated view about the epistemic status of extrapolations. Extrapolations are used in the context of an overall research strategy that combines both a bottom-up process of inferring mechanistic accounts based on experimental data and a subsequent top-down testing of predictions made by these accounts. On the one hand, defenders of extrapolative practices are right in claiming that the use of surrogate models is a reliable experimental practice, and the tens of thousands of monthly articles publishing results gathered in surrogate

models certainly demonstrate that extrapolations from validated surrogate models are treated as legitimate evidence supporting the bottom-up construction of explanatory mechanistic accounts. On the other hand, however, the equally well documents efforts deployed to retrospectively test extrapolative inferences also demonstrate that findings from any given surrogate model are not considered 'definitive' or 'sufficient' evidence, but simply one line of legitimate evidence, subsequently strengthened and corroborated by means of top-down testing of predictions from mechanistic models.

---

*Model Organisms and Explanation*

ULRICH STEGMANN

University of Aberdeen

u.stegmann@abdn.ac.uk

---

Much research in the life sciences and biomedicine is organized around model organisms. Philosophers of science have identified several theoretical roles of model organisms, both epistemic and non-epistemic. The epistemic role of model organisms is invariably seen as licensing inferences to other organisms. Some authors explicitly understand the role of model organisms in terms of concrete, theoretical models (Weisberg 2013). Even Levy and Currie (2014), who have put some distance between model organisms and theoretical models, maintain that model organisms are models in the sense of proxies that serve to make inferences about other organisms. When it comes to the epistemic role of model organisms, most attention has therefore focused on determining how model organisms license inferences to other species.

This paper argues that the standard view of model organisms as proxies ignores, or takes for granted, a crucial epistemic role, i.e. explaining general biological processes. I start by questioning an assumption about what is being extrapolated when model organisms are used as proxies. This leads me to articulate the explanatory role. In the remainder of the paper I will identify an investigative function unique to the explanatory role. Furthermore, I argue that the generality of epistemic results, which are gained by employing the explanatory strategy, depends not only on the degree to which underlying mechanisms can be generalized.

Steel (2008) maintains that from a phenomenon and its underlying mechanism in the model, scientists infer either (1) only the phenomenon in the target or (2) both the phenomenon and its mechanism. I will argue, however, that the phenomena often are already known to occur in the target organism(s) and therefore only the underlying processes are inferred from the model. This can be seen in inferences concerning action potentials (in *Loligo* sp.), transcription control (in *E. coli*), floral morphogenesis (in *Arabidopsis*), and pathophysiology of schizophrenia (in mice).

The question of how higher-level features relate to their underlying factors has been at the center of much work on explanation and reduction in the biological sciences, and model organisms figure prominently in this literature. These case studies have demonstrated, albeit implicitly, that research on model organisms generates accounts or explanations of some features in terms of other factors. This is an epistemic role that is largely taken for granted in the literature of model organisms. The point here is not simply that research on model organisms can be aimed at exploring biological processes, as emphasized by Hubbard (2007) and Bolker (2009). Nor is the point that such research yields accounts (or models) of some biological process, as



highlighted by Meunier (2012) and Leonelli and Ankeny (2011). Instead, I draw attention to the fact that, once we distinguish between the investigated process and its underlying factors, research on model organisms can be seen to play an *epistemic* role that is largely independent of its inferential role.

I will argue that the independence of the explanatory from the inferential role manifests itself in at least three ways. First, the extent to which a model organism satisfies its explanatory role is independent of the extent to which it satisfies its inferential role. Second, a model organism's explanatory role is a precondition for the model playing an inferential role. Without identifying some factor that explains the phenomenon in the model, there is nothing to extrapolate (in the kind of instances mentioned above).

Most importantly, I will argue the explanatory role is uniquely suited to achieve a particular kind of investigative goal. Suppose scientists are interested in explaining a variety of phenomena that occur in several organisms, including the model organism. Scientists can focus on one of the processes and identify (some of) the factors responsible for it. This strategy can be iterated with the other processes of interest, either in the same or in a different organism. In doing so, scientists only employ the explanatory role of their chosen model organism. Iterations may concern a set of phenomena that are more or less unrelated to one another. For example, research in *Arabidopsis* has delivered explanations for processes as diverse as root development, light perception, disease resistance, cold and freezing resistance, and floral morphogenesis. Rather than exploring fairly unrelated phenomena, scientist can also employ the explanatory strategy in order to explore phenomena across different levels of organization. I will illustrate the latter with examples from the research on learning and memory in rodents (LTP, AMPA receptor trafficking, memory). The inferential role appears unsuited to achieve the same goal. Extrapolation of the factors responsible for a phenomenon from the model organism to the target will, at best, justify the assertion that the same factors are responsible for the same phenomenon in the target organism. But an explanation of *that* phenomenon has already been achieved in model organism. No amount of understanding how of the same phenomenon is realized in the other organisms will advance our understanding of any of the other processes.

## References

- Ankeny, R. A. & Leonelli, S. (2011) "What's so special about model organisms?" *Studies in History and Philosophy of Science* 42: 313-23.
- Bolker, J. A. (2009) "Exemplary and surrogate models: Two modes of representation in biology," *Perspectives in Biology and Medicine* 52: 485-99.
- Hubbard, E. J. A. (2007) "Model organisms as powerful tools for biomedical research," in A. N. H. Creager, E. Lunbeck & M. N. Wise (eds.), *Science without laws: model systems, cases, exemplary narratives*, Raleigh, NC: Duke University Press, pp. 59-72.
- Levy, A. & Currie, A. (2014) "Model organisms are not (theoretical) models," *British Journal for the Philosophy of Science* published online:
- Meunier, R. (2012) "Stages in the development of a model organism as a platform for mechanistic models in developmental biology: Zebrafish, 1970-2000," *Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 522-31.
- Steel, D. P. (2008) *Across the Boundaries: Extrapolation in Biology and Social Science*, New York: Oxford University Press.

Weisberg, M. (2013) *Simulation and Similarity: Using Models to Understand the World*, New York: Oxford University Press.

---

*Modeling Organs with Chips: Design and Representation as Modeling Relations*

MICHAEL POZNIC

TU Delft

m.poznic@tudelft.nl

---

This paper uses discussions on modeling and representation to clarify the question as to how the product of the activity of technological designing should be conceived. Two kinds of modeling relation between vehicles and targets are distinguished which differ in their respective directions of fit. The representation relation has a vehicle-to-target direction of fit and the design relation has a target-to-vehicle direction of fit. A case study in bioengineering shows that a certain product of designing can participate in both, design and representation relations.

The activity of modeling is a core activity of the contemporary practice of scientists: Research at universities, private institutes, and industry often centers around building and using models for a variety of purposes. These purposes include the utilization of mathematical models in order to predict or explain phenomena, the use of concrete models like, e.g., scale models or animal models in order to perform experiments upon them or the application of computational models in order to conduct simulation studies (cf. Weisberg 2013). What unites these different aims of prediction, explanation or simulation is the epistemic aim to learn something about the world. A prerequisite for models to be used in order to infer justified claims about certain phenomena, so-called target systems, is that the models represent these parts of the world. The models are representational vehicles that should adequately represent target systems. For example, one could claim that only if a model adequately represents a target system then this model can be used to reach an understanding of that system. This is reflected in debates that take representation to be the main function of epistemic tools like scientific models. Part of this perspective on models is the implicit focus on vehicle-to-target directions of fit between models and targets. In engineering, including the engineering sciences, the use of models is also a widespread activity. Yet, in technological modeling there are other aims besides the purely epistemic ones at play. An often-mentioned conviction is that engineers want to change the world instead of just understanding it. In engineering, certain vehicles are used in order to construct or change corresponding targets. In such a context, a target-to-vehicle direction of fit is aimed at. Because of that I will speak of a design as a special form of modeling relation. For example, there is a relation between a description of a planned artifact and a realization of that artifact, or the relation between a conceptual design and an artifact that should be built in the future. This relation can be seen as a kind of modeling relation. The relation constitutes a specific kind of representation. It is a modeling relation between a vehicle and an object that has to be modified or between a vehicle and an object that does not exist yet. The activity

of designing – pretty much as the activity of representing – is done with the help of a certain vehicle. However, the goal of designing is not to learn whether certain facts obtain, rather the goal is to bring certain facts about. Either in core designing or in product designing there is a goal to modify an object or to create something that did not exist beforehand. So, the relation of design is about how certain matters should be in contrast to how they in fact are. Here, the condition for the success of the relation is different from the condition for the success of representation in the traditional sense. In the case of design, the target has to be adjusted to the vehicle in order for the technical functions to be fulfilled. In the case of representation, the vehicle, in most cases a model has to be adjusted in order to adequately represent a corresponding target system. A black and white picture of science and engineering would be: scientists seek to understand the world; engineers aim at changing the world. Understood in this way, design can be regarded as the defining characteristic of engineering and, in the philosophy of technology, it is claimed that the main goal of the practice of engineering is to come up with effective and efficient designs (cf. Meijers 2009; Vermaas et al. 2011). I will use a case study in bioengineering in order to show that the notions of, both, design and representation are relevant for an account of modeling in science as well as engineering. The case study will be an analysis of so-called organ on chip models in the field of tissue engineering (Huh et al. 2013). This field of engineering has a great overlap with cell research in biology and particularly with medical research. The models are used to represent human organs and to study inter alia the toxicity or efficacy of certain drugs. This practice is driven by epistemic goals such as to predict risks and benefits of certain treatments with a specific drug or to learn about certain diseases. In this modeling, there are practical goals as well. The engineers have the aspiration of developing certain technical artifacts during the modeling of the organs with the help of their techniques.

This paper proposes that both aims, understanding *and* changing the world, should be accounted for in a philosophical analysis of engineering modeling. The case study shows that modeling organs in bioengineering is guided by both aims. For both, understanding and for changing the world, models are used to reach the respective aim. The goal of the practice of modeling is to establish two different kinds of relation between vehicles and targets. In the case study, these two kinds of modeling relation between vehicles and targets are distinguished which differ in their respective directions of fit. The representation relation has a vehicle-to-target direction of fit and the design relation has a target-to-vehicle direction of fit. The case study further shows that a conception of modeling as involving only relations with a vehicle-to-target direction of fit is too narrow in order to account for models in science and engineering.

## References

- Huh, D., Kim, H. J., Fraser, J. P., Shea, D. E., Khan, M., Bahinski, A., Hamilton, G. A. and Ingber, D. E. 2013. "Microfabrication of human organs-on-chips." In *Nature Protocols* 8 (11): 2135-57.
- Meijers, A., ed. 2009. *Philosophy of technology and engineering sciences* (Handbook of the philosophy of science, volume 9). Amsterdam: North-Holland.
- Vermaas, P., Kroes, P., van de Poel, I., Franssen, M. and Houkes, W. 2011. *A Philosophy of Technology. From*

Technical Artefacts to Sociotechnical Systems. S.l.: Morgan & Claypool.

Weisberg, M. 2013. *Simulation and Similarity. Using Models to Understand the World*. New York: Oxford University Press.

---

*Explanation, Unification, and Mechanisms*

MELINDA FAGAN

University of Utah

mel.fagan@utah.edu

---

This paper motivates and presents a new account of unifying explanation. The standard account (Friedman 1974; Kitcher 1981, 1989) preserves much of the covering-law model of explanation, associating unification with subsumption of many diverse phenomena under a general argument pattern. Kitcher's theory of unifying explanation relaxes the requirement for general laws, focusing instead on general argument patterns from which descriptions of a wide variety of phenomena can be derived. What is unified is the set of all beliefs accepted at a given time, and what does the unifying is a set of argument patterns that most economically justifies all the accepted beliefs. Like the covering law view, Kitcher's theory identifies explanatory power with generality, simplicity and systematicity. The basic idea is that we achieve understanding by identifying a common pattern that unites apparently disparate facts. Revealing these hidden similarities provides a simpler and more intelligible worldview.

Many scientific explanations, however, do not conform to this basic idea, and so fit poorly with the unification or covering-law accounts. An important variety of these, mechanistic explanations in biology, has received considerable philosophical attention (e.g., Machamer et al 2000; Bechtel and Abrahamsen 2005, Craver 2007). The current consensus is that mechanistic explanations explain the behavior of an overall system (a mechanism) by describing how its behavior is produced by the mechanism's component entities, activities, and organization. This stock philosophical description suggests that the phenomenon to be explained is causally produced by its organized components - that the relation between mechanistic parts and the overall system is a causal one. Yet prominent New Mechanists, including Craver and Bechtel (2007) deny this, characterizing the relation as constitutive or functional.

This paper proposes an alternative account: 'unifying mechanistic explanation.' This hybrid account sheds light on recent debates among the New Mechanists and suggests norms for interdisciplinary research programs such as systems biology. The argument is in three parts. The first uses a case study, Jacob and Monod's operon model (1961), to argue that a purely causal theory of mechanistic explanation in biology is incomplete. This approach follows the New Mechanists' method, focusing on scientists' explanatory practices. The operon example, which is representative of many explanations in molecular biology, reveals that symmetric 'combining relations' are as prominent as causal relations. Combining relations, as represented in mechanistic explanations, link multiple parts into single complex whole, and underlie key

causal relations in mechanistic descriptions. Focusing on combining as well as causal relations in mechanistic models extends and deepens the insights of New Mechanists.

In the second part of the argument, I show that this extended conception of mechanistic explanation involves at least three senses of unification. First, combining relations unify lower-level parts by connecting them into a new, complex whole. Second, in molecular biology the connecting relations among mechanism components are limited in number. Diverse molecular entities are connected in virtue of two properties: (i) complementary molecular geometry and (ii) electrochemical charge. These two features ground a wide range of diverse combinations of molecules, affording a kind of simplicity and generality to this aspect of mechanistic explanation. Third, mechanistic explanations unify higher- and lower-level descriptions of the phenomenon of interest. The second sense of unification resembles that of Kitcher's traditional unification account. The first and third, in contrast, can be understood as a combining relation 'writ large,' encompassing a set of causal and combining relations that can be identified with the higher-level behavior to be explained. On this view, mechanistic explanations do not only reveal causal dependency relations underlying a phenomenon of interest, but also provide a multi-level perspective on that phenomenon. I contrast this account of unifying explanation with its predecessors, and summarize the key points of divergence.

The third and final part of the paper examines implications of this new account. I show how this view extends and deepens the insights of New Mechanists. One important consequence bears on recent debates about mechanistic explanation and systems biology. Systems biology is an interdisciplinary research program that combines molecular biology experimentation with mathematical modeling and principles from physics and engineering. Explanatory models in systems biology exhibit the same multi-level structure as mechanistic explanations in molecular biology. But the tools used to construct systems models are different: a combination of mathematical techniques, computer simulation, and concepts from physics and engineering, applied to large experimental datasets. A number of philosophers of science have recently discussed the bearing of systems biology on mechanistic explanation, arriving at very disparate conclusions (e.g., Bechtel 2011, Issad and Malaterre forthcoming, Zednik forthcoming). The main point of dispute concerns the role of mathematical modeling and computer simulation in systems explanations.

It is widely agreed that the role of these systems practices is to connect lower-level interactions among components (of regulatory networks) with higher-level behavior (of cells and organisms). Unlike the situation for classic molecular mechanisms, there is no intuitive way to connect these two levels in systems models. The role of mathematical modeling and computer simulation is to forge a derivational connection between levels. Dispute arises over whether the derivational inter-level relation is incompatible with mechanistic explanation. Here the new account of unifying explanation offers clarification. Briefly, on a purely causal view of mechanistic explanation (i.e., mechanistic models describe causally-linked system- and component-levels), systems explanations are not mechanistic. The latter use mathematical derivation/simulation to link levels in ways that are not easily conceived as causal (at least not without extensive argument). But on the expanded,

unifying view, systems explanations are mechanistic, using different tools to accomplish what all mechanistic explanations do – link two or more levels intelligibly, as parts and whole of a complex system.



## Poster Session

## General Philosophy of Science

Room 34, Thursday / Friday / Saturday 11:30 – 12:00

## Interest Relativity in the Best System Analysis of Laws

MAX BIALEK

University of Maryland, College Park, University of Groningen

mbialek@gmail.com

David Lewis' Best System Analysis (BSA) of laws of nature has it, roughly, that a regularity is a law just in case it appears in the best systematization(s)—where the “best” is the simplest and strongest on balance—of all the particular fundamental matters of fact (Lewis 1973, 1983, 1994, and elsewhere). A common criticism of the BSA focuses on the interest relativity inherent in deciding what it means for a system to be the best. As Armstrong (1985, p. 67) puts it: “The *first* objection which may be made to the [BSA] is that an element of subjectivism remains... May there not be irresolvable conflicts about the exact point of balance [between simplicity and strength]?” If there are such irresolvable conflicts then either there are no BSA-style laws at all or such laws are subjective in the sense that they are laws only relative to the particular balancing of simplicity and strength. Proponents of the BSA (or variations of it) since Lewis have tended to embrace relativity despite Armstrong's objection, with the standard refrain being that the best system is “the best *for us*” given our interests and cognitive limitations (examples of this include Loewer 2007, Cohen and Callender 2009, and, in a related analysis of probabilities, Frigg and Hoefer 2010).

I argue in this paper that a proponent of the BSA can accept the interest relativity of its laws *and* provide (or at least make progress towards) an answer to Armstrong's objection. The central idea is that interest relativity comes in degrees (in much the same way that moral relativism does). On the most subjective end of the spectrum, every individual scientist might pick out different laws. On the other (most objective) end, every scientist picks out the same laws. A lot happens between those extremes. Every human scientist might get to the same laws, and every Martian scientist will get to the same laws, but the human and Martian laws are different. It may be that the laws are identified dependent on the scientist's field of study; economists all get to the same laws, and those are different from the laws that are arrived at by each physicist. Maybe Martian and human physicists have the same laws, but the respective economists do not.

To answer Armstrong's first objection would seem to require an assurance that the BSA will locate the laws on the most objective end of the spectrum where everyone finds the laws to be the same. In general, the more agreement that can be assured by the BSA—the less interest relative the laws are—the closer it is

to answering Armstrong's first objection. And, insofar as the BSA cannot assure complete agreement on the laws, the objection may be answered in full when what disagreement remains is shown to be defensible as a part of a proper analysis of laws. Thus the strategy for a BSA proponent to answer Armstrong's first objection works in two directions that will hopefully meet. In the *pro-objectivity* direction, BSA proponents should try to identify rules for picking out the best system that are broadly agreeable. In the *pro-relativity* direction they should try to defend particular ways in which the laws may be interest relative. In what remains I offer an example of what might be said in each direction, with a focus on the pro-objectivity direction.

Consider briefly the pro-relativity direction. One strategy here would be to work to include special science laws as a part of broader (than Lewis') BSA-style analysis. The laws of a special science are such plausibly because they are the laws *relative to the interests of that special science*. This idea has already been pursued by Callender and Cohen (2009, 2010) who relativize laws to sets of kinds that are treated as basic in the competition whose winner is the titular best system. For example: The laws of biology are drawn from the best system coming out of the competition that is run when treating biologically interesting kinds as basic. Not only does this provide a BSA-style analysis of special science laws, but it also provides those working in the pro-objectivity direction with a principled excuse for not having a universally acceptable procedure of picking out what kinds are to be treated as basic in the laws.

Pursuing the pro-objectivity direction should begin with a retreat from the standard refrain of "best is the simplest and strongest on balance". These are fraught notions and it is precisely because of uncertainty about there being agreement to be had with regards to them that Armstrong raised the first objection. We might begin instead with the idea that, whatever laws are, they should support inductive inferences. If there is disagreement on this point then I would claim there is disagreement about the very object of study—someone looking for things that don't support inductive inferences is someone not looking for laws.

We seek in light of the above a rule for the best system competition that will yield laws that are better able to support inductive inferences. I propose that such a rule may be of the form "score laws in proportion to the mutual information that they allow for between spatio-temporal regions of the world". By mutual information (MI) I mean here the measure on two relatable random variables that is found in Information Theory. I argue that, paired with laws that indicate how it may be calculated, MI can function as a measure of how well induction will work in a given world. Also, given its source in Information Theory, MI is well placed to accommodate our intuitions that laws should be simple (via the theory's concern with efficient encoding) and strong (directly via the interest in informativeness). These last points suggest specific work to be done in the future that exists as a part of the larger strategy developed earlier in the paper for answering Armstrong's first objection to the BSA.

## References

- [1] Armstrong, D. M. (1985). *What is a Law of Nature?*. Cambridge University Press.
- [2] Beebe, H. (2000). The non-governing conception of laws of nature. *Philosophical and Phenomenological*

*Research*, 571-594.

- [3] Callender, C., & Cohen, J. (2010). Special sciences, conspiracy and the better best system account of lawhood. *Erkenntnis*, 73(3), 427-447.
- [4] Cohen, J., & Callender, C. (2009). A better best system account of lawhood. *Philosophical Studies*, 145(1), 1-34.
- [5] Frigg, Roman and Carl Hoefer (2010). Determinism and Chance from a Humean Perspective. In Friedrich Stadler, Dennis Dieks, Wenceslao Gonzalez, Hartmann J., Uebel Stephan, Weber Thomas and Marcel (Eds.), *The Present Situation in the Philosophy of Science*. Springer.
- [6] Lewis, D. (1973). *Counterfactuals*. John Wiley & Sons.
- [7] Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4), 343-377.
- [8] Lewis, D. (1994). Humean supervenience debugged. *Mind*, 103(412), 473-490.
- [9] Loewer, B. (2007). Laws and natural properties. *Philosophical Topics*, 35(1), 313.

---

What Good is Realism about Natural Kinds?

ANA-MARIA CRETU

University of Edinburgh

d.cretuanamaria@gmail.com

---

Natural kinds realism can be understood as a series of views put forward by those scientific realists committed to kinds that 'latch onto' the real structure of the world. Natural kinds are believed to be the best explanatory tool in that they explain why theories featuring those kinds prove inductively and predictively successful. This is, in a nutshell, what I take to be the epistemological argument for natural kinds. A main proponent of this argument is Richard Boyd whose account has become the received view of realism about kinds. In a series of papers ([1991], [1999a] [1999b]) Boyd articulates a realist account of natural kinds: homeostatic property cluster kinds (HPCK). Natural kinds are, on Boyd's view, necessary to establish the reliability of successful epistemic practices. The idea behind the epistemological argument is the following: we are to some extent justified in giving a posteriori definitions of natural kinds in certain ways that reflect the actual causal structure of the world because we cannot make projectible generalizations otherwise (Boyd [1991], p.138).

Boyd's HPCK account is designed to explain how kinds used in successful epistemic practices 'latch onto' natural divisions in nature. His main motivation for defending such an account is that of defeating skepticism about the success of science, which is reminiscent of the Lockean nominalist tradition. According to this tradition, "we must classify substances according to arbitrary nominal essences instead of according to microstructural real essences" (Boyd, [1991], p. 131). This is because we cannot know the 'real essences' of kinds. Hence, things are classified in virtue of some arbitrary nominal essences. But kinds that are the result of such arbitrary classifications cannot support successful inductive generalizations and make knowledge of such kinds in general seem impossible. The Lockean tradition of kinds thus gives rise to a "tension between empiricist nominalism and the task of accounting for induction" (Boyd, [1991], p. 130). Hence, this tradition is largely responsible for opening the doors to skepticism about the ability of science to use kinds to ground epistemic practices. Boyd's aim is to revoke such skepticism by showing that "in induction and explanation we must refer to kinds whose definitions are specified a posteriori, in deference to nature, rather than nominally" (Boyd, [1991], p. 131).

In this talk I argue that a realist account of natural kinds à la Boyd is neither necessary nor sufficient to explain success in science. In analyzing Boyd's account I distinguish between the constitutive factors of HPCK and the individuation conditions of HPCK: these distinctions are crucial for understanding Boyd's epistemological argument for kinds; they also constitute the basis for the subsequent objections that are

aimed at establishing whether or not Boyd's HPC account is a genuinely realist account of natural kinds. I argue that Boyd's HPC account is neither necessary nor sufficient for grounding epistemic practices in science because: i) individuating the constitutive factors of HPC is more often than not a matter of human decision; ii) the HPC account falls short of accommodating successful scientific kinds that cannot be described in terms of clusters of properties and underlying homeostatic mechanisms; and iii) the HPC account includes as kinds things that by the lights of our present science, failed to latch onto the causal structure of the world (for related arguments see Ereshevsky&Reydon [2015], Slater [2014], Khalidi [2013]). Failing to deliver on their epistemic potential, the commitment to HPC proves not to be the best available tool in the scientific realists' toolbox. I conclude that the epistemological argument should not be made dependent on natural kinds carving nature's joints in some strong realist sense. We can still think of natural kinds as explaining why theories featuring those kinds prove inductively and predictively successful whilst not having just one account of natural kinds. There is not one notion of 'natural kind' that best serves science; in fact the notion of 'natural kind' changes and matures with scientific progress.

Having shown that Boyd's realist account of natural kinds is neither necessary nor sufficient for grounding epistemic practices in science, I conclude that a commitment to natural kinds in some strong realist sense is not necessary to establish the reliability of successful epistemic practices. Instead, we should opt for a less ontologically inflationary account. Taking the cue from Quine's discussion in "Natural Kinds" [1969] I aim to rehabilitate the view that there is no account of natural kinds that spans across all sciences. Instead, I want to suggest, following Quine, that there is a sense in which no unique account of natural kinds is suited to account for the epistemic endeavors of all science. However, there is a sense in which a particular notion of kinds pervades all science. This notion, which is (at least implicitly) deployed by each branch of science in its respective epistemic practices, is the notion of scientifically entrenched kinds. It's worth noting that the notion of 'scientifically entrenched kinds' should be understood as a methodological place-holder for all types of kinds that are useful for the epistemic endeavors of the natural and social sciences. The account I propose is a type of pluralism about accounts of kinds that serve some important epistemic role in science, which accommodates the strengths of the HPC account, whilst not sliding into Dupre's 'promiscuous realism'.

### References

- Boyd, R., [1991], "Realism, Anti-Foundationalism and the Enthusiasm for Natural Kinds", *Philosophical Studies*, 61: 127–148.
- Boyd, R. [1999a], "Kinds as the "workmanship of men"": realism, constructivism, and natural kinds'. In J. Nida-Rümelin (ed.) *Rationalität, Realismus, Revision: Proceedings of the Third International Congress, Gesellschaft für Analytische Philosophie*. Berlin: de Gruyter.
- Boyd, R. [1999b] 'Homeostasis, species, and higher taxa'. In R. Wilson (ed.) *Species. New Interdisciplinary Essays*. Cambridge: MIT Press.
- Ereshevsky, M. & Reydon, T.A.C. (2015): "Scientific kinds: A critique of HPC theory and a proposal for an alternative account", *Philosophical Studies* 172

- Khalidi, M.A., [2013] *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*  
Cambridge: Cambridge University Press,
- Quine, W. V. [1969], "Natural Kinds" in *Ontological Relativity and Other Essays*, Columbia University Press.
- Slater, M.H., [2014], "Natural Kindness", *British Journal of Philosophy of Science*, 1–37.

---

*Laws are Conditionals* (cancelled)

TOBY FRIEND

University College London

toby.friend.13@ucl.ac.uk

---

An often assumed, but rarely argued for, view of laws of nature is that they are conditionals. The ubiquitous '*all Fs are Gs*' dominates much philosophical discussion on laws but rarely is it shown how actual laws mentioned and used in science are supposed to fit it. Instead, what can seem embarrassingly toy examples like '*all ravens are black*' are employed for discussion. Ignoring the complexity of real cases has some merit, but it also comes at a price, and today there is increasing literature arguing that laws have been misrepresented: either we have been misguided in providing their logical form (e.g. Maudlin 2007) or else we are misguided in supposing them to be important features of science (e.g. Cartwright 1983, Van Fraassen 1989).

My presentation will offer, in the first part, an argument that laws are indeed conditionals. This conclusion may at first appear rather insignificant, however I will show, in the second part, that it impacts on a number of more weighty discussions.

I begin with an analysis of the expression ' $V = IR$ ', employed to represent Ohm's law. The symbols  $V$ ,  $I$  and  $R$  clearly represent variables rather than properties, otherwise we could not make sense of the expression as an *equation*. But the meaning of Ohm's law cannot simply be an algebraic statement relating numerical variables since this would render it a trivial truth of mathematics, having nothing to do with the physical world. It would also be trivial if the symbols represented functions from objects to numbers; there are infinitely such functions. A step in the right direction is to understand the entire statement ' $V = IR$ ' as determining a set of functional relationships between sets of properties, e.g. the function which takes as input one property in the set of voltages and another in the set of resistances and outputs some third property in the set of currents. However, this manoeuvre renders the statement non-propositional, since functions are operations not statements, and triviality will again ensue if we render the law as an assertion of the mere existence of such functions.

After navigating these steps, I will suggest that what is missing in the formulation of laws expressed as equations is the added clause that the set of functional relationships described by them are *true of things*. For instance, the set of functions determined by ' $V = IR$ ' map voltages and resistances to currents *of one and the same individual*. This suggests that the information bound up in the statement is something to be predicated of an entity. In general, we may understand equations to denote a high-order behavioural

property which entities have when and only when a specific set of relations determined by those equations holds between various (usually quantifiable) lower-order properties of that entity.

Given this interpretation, we must ask how the variables predicated by the behavioural property-term are quantificationally bound. Since laws are supposed to be general, we could understand the variable to be bound by the universal quantifier. However, it is certainly not the case that Ohm's law, or in fact *any* law, says that *everything* bears some higher-order relationship between its properties. Most individuals, for example, do not even have voltages and currents let alone satisfy the relationships determined by the equation  $V = IR$ . What is missing in the representation of Ohm's law as an equation, therefore, is that the functions denoted by the behavioural property-term relate the voltage, current and resistance of an *electrical conductor*, i.e.,

**Ohm's Law:** *If* a system is an electrical conductor *then* its voltage, current and resistance are related by the set of functions entailed by  $V = IR$ .

Hence, a proper understanding of the utility of the expression  $V = IR$  requires us to understand it in the context of a conditional. The conditional says that a certain sort of behaviour (broadly construed) may be inferred if something is a particular type of system, an electrical conductor in this case.

These points apply generally to all equations and functional relationships commonly associated with a law and even some which aren't, such as the Lotka-Volterra and Schrödinger equations. In general the informativity of all these relationships must come from understanding their place in a conditional.

The rest of the presentation consists in highlighting the impact such a finding has on three areas of discussion. First, according to some authors (Beatty 1995, Rosenberg 2001, Schurz 2002, Reutlinger 2011), there is logical difference between laws in the life sciences and those in physics. Whereas the former class of laws only hold of particular types of system, the latter are supposed to hold irrespective of system-type. However, the argument described above implies that *all* laws include an antecedent qualification some type of system. This seems intuitively correct, as even one of our most widely applicable formulae, the Schrödinger equation, fails to apply, e.g., to thoughts, clouds, numbers, and countries. But since the qualification is inside the scope of the universal quantifier, all laws with such antecedents are still fully general. Hence, according to the argument, there can be no strict division on account of logical form of the laws across the sciences in the way suggested.

Second, those who advocate the so-called 'semantic view' of scientific theories have often (e.g. Van Fraassen 1989) played down the respect to which laws play a role in theory, promoting a different picture of theories as a families of models. However, the argument for the conditionality of laws suggests that the information implicit in the employment of a model (e.g. an equation) must be something beyond the model itself. According to the above argument, a model is to be understood as descriptive of a particular behaviour. In order to know how to apply a model, one must know how the description features in the consequent of a conditional in which the conditions for application for the model are specified in the antecedent's 'system



predicate'. Hence, for a scientific theory to be informative it must contain not just models, but conditional laws too.

Third, understanding laws as conditionals of the form  $\forall x(Fx \rightarrow Gx)$  serves to defl much of the concern surrounding their ontology. First, their universal quantification makes it less plausible to see them as specifically *applying to* or *about* anything in particular. Arguably this disperses some of the mystery behind the notion of laws' 'governance'. Second, the single argument-position gives laws the same structure (and arguably, same modality) as statements of essence like 'all gold has atomic no. 79', and 'all enzymes are catalysts'. Although these latter statements have instigated much debate of their own, their inclusion with laws under a single category would allow for a unified treatment. (This would not be the case if, as many have thought, laws have the more complex form  $\forall x(Fx \rightarrow \exists y(Gy))$  e.g. in Pietroski and Rey 1995.)

### References

- Beatty, J.: 1995, The Evolutionary Contingency Thesis, in G. Wolters and G. Lennox (eds), *Concepts, Theories, and Rationality in the Biological Sciences*, Pittsburgh University Press, pp. 45–81.
- Cartwright, N.: 1983, *How the Laws of Physics Lie*, Oxford University Press, Oxford. Maudlin, T.: 2007, *The Metaphysics within Physics*, Oxford University Press, New York.
- Pietroski, P. and Rey, R.: 1995, When Other Things aren't Equal: Saving Ceteris Paribus Laws from Vacuity, *British Journal of Philosophy of Science* **46**, 81–110.
- Reutlinger, A.: 2011, A Theory of Non-Universal Laws, *International Studies in the Philosophy of Science* **25**, 97–117.
- Rosenberg, A.: 2001, How is biological explanation possible?, *British Journal of Philosophy of Science* **52**, 735–760.
- Schurz, G.: 2002, Ceteris Paribus Laws: Classification and Deconstruction, *Erkenntnis* **57**(3), 351–372. Van Fraassen, B.: 1989, *Laws and Symmetry*, Oxford Clarendon Press.

---

Natural Kinds, Causal Profile and Multiple Constitution

MAX KISTLER

Université Paris 1 Panthéon-Sorbonne

mkistler@univ-paris1.fr

---

A natural kind of objects such as electrons or  $K^+$ -ions can be characterized by a characteristic set of properties that always go together. Such a set of properties forms what Boyd (1991) has called a “homostatic property cluster”, and Slater (2014) a “stable property cluster”. It is crucial for biochemical explanations that  $K^+$  ions have several properties. It is not only their positive charge that determines the  $K^+$  ions’ contribution to a mechanism such as the propagation of action potentials, as can be seen from the fact that equally positively charged  $Na^+$  ions play a completely different role in the same mechanism.

The identity of a natural kind is determined by a set of dispositions that constitutes its causal profile. In the case of biochemical kinds, some of their dispositions correspond to functions. Natural kinds are theoretical postulates that provide a framework for the scientific explanation of why certain properties and dispositions systematically go together.

Armstrong (1997), Bird (2007) and Tobin (2013) suggest that NK can be analyzed as being equivalent to sets of 1) several properties and 2) laws or mechanisms that hold these properties together. However, NK are neither equivalent to sets of properties nor to sets of properties, laws and mechanisms. I first argue for this claim in the case of fundamental natural kinds, and then show that this gives us a reason to keep the metaphysical category of natural kinds in general. Electrons are a fundamental kind. Hence the coexistence of their characteristic set of properties cannot be given a mechanistic explanation. However, even non-fundamental kinds require NK as a metaphysical category insofar as their reduction makes reference to more fundamental NK.

A natural kind cannot be reduced either to something that fills a functional role. First, one natural kind can in principle have many functions, and second, different kinds can share a function. So-called moonlighting proteins belong to the first type of case. They are named metaphorically with respect to people with a second job, done by moonlight. Crystallins are one type of biological molecules that can play several functions. In the lens of vertebral eyes they play the structural function of guaranteeing refractive properties and transparency. However,  $\alpha B$ -crystallin has, beyond its function as a lens protein, also the function of protecting cells from elevated temperatures, thus serving as a “small heat shock protein” (De Jong et al. 1993). The fact that crystallins play different functions in different circumstances justifies their existence as a natural kind. If a molecule played exactly one function, it would be more parcimonious to postulate just a dispositional

property. If it played a determinate number of functions, we might avoid the postulate of the kind by referring to a cluster of dispositional properties instead.

A NK is a powerful entity: all objects belonging to an NK have a certain number of dispositions to behave in various circumstances. A part of these dispositions is actually manifested, a part corresponds to biological functions, and only a part is known. The multi-functional profile of a NK shows that its identity is neither determined by a single function nor by a unique *set* of dispositions/functions because it is essential, at least in the case of proteins, that they can acquire new functions. Rather, a NK is what underlies and determines a set of functions that always go together. The set of functions is the NK's "causal profile"; the known part of this profile constitutes our inductive basis for postulating it.

There is a second and complementary reason for which there is no 1:1 relation between biomolecules and functions: There are many examples of biological functions that are shared by different substances. In evolutionary terms, this is explained by "convergent evolution", or more precisely "functional convergence" (Buller and Townsend 2013). Oxygen transport molecules provide a clear case. At least three types of molecules have evolved to play the role of transporting oxygen through a circulatory system to the various tissues in an animal's body. Hemoglobins, hemerythrins and hemocyanins (van Holde and Miller 1995).

The causal profile characteristic of a chemical NK is determined by its microstructure. This follows from physicalism and from the supervenience of higher-level properties of complex objects on the lower-level properties of their constituents. However, this is not enough to establish microstructuralism about natural kinds, which is "the thesis that membership of [a] kind is conferred by microstructural properties" (Hendry 2006, p. 865). What follows from physicalism is that if A and B have the same microstructural composition then a sample of a chemical substance A is of the same chemical substance as B. However, the reverse does not hold. It is not the case that if a sample of a chemical substance A is of the same chemical substance as B then A and B have the same microstructural composition. This is because a macroscopic NK can be determined by different microstructures. If a chemical substance is "multiconstituted", there are two or more microscopic structures that give rise to the same substance. Hemoglobin is a case in point. There is a huge variety of hemoglobins even within the human species (Huisman et al. 1996). However, only a part of the microstructure makes a difference at the level of the chemical and biological functions of hemoglobin. What makes hemoglobin what it is, i.e. what underlies its functional profile, is determined by a very specific part of the microscopic structure, which consists of the "highly conserved amino acid residues in hemoglobins" (Anandhi 2014, p. 3-34).

The microstructuralist thesis according to which there is a 1:1 relation between the global identity of a chemical substance (= type of molecule) and its microstructure (the set of its constituents together with their relations) is incorrect for such proteins as hemoglobin because there is a many-one relation between their "microstructure" (the sequence of amino acids that constitutes their primary structure) and the tertiary and quaternary structure that determines their characteristic functional profile.

**References**

- Anandhi D. (2014), *Introduction to Biochemistry and Metabolism*, Noida (India) : Dorling Kindersley.
- Armstrong, David M. (1997), *A World of States of Affairs*, Cambridge, Cambridge University Press.
- Bird, Alexander (2007), *Nature's Metaphysics. Laws and Properties*. Oxford, Clarendon Press.
- Boyd, R. (1991), Realism, Anti---Foundationalism and the Enthusiasm for Natural Kinds, *Philosophical Studies* 61, p. 127---148.
- Buller, AR; Townsend, CA (2013). Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proceedings of the National Academy of Sciences of the United States of America* 110 (8): E653– 61. doi:10.1073/pnas.1221050110. PMID 23382230.
- de Jong W.W., Leunissen J.A., Voorter C.E. (1993). Evolution of the alpha crystallin/small heat-shock protein family. *Mol. Biol. Evol.* 10 (1): 103–26.
- Hendry R.F. (2006), Elements, Compounds and other chemical kinds, *Philosophy of science* 73, p. 864-875.
- Huisman, Titus H.J., Carver, Marianne F.H. and Efremov Georgi D., *A Syllabus of Human Hemoglobin Variants* (1996), The Sickle Cell Anemia Foundation, Augusta, GA, USA.
- Slater, M. (2014), Natural Kindness, *British Journal for the Philosophy of Science*. doi: 10.1093/bjps/axt033
- Sousa P., Atran S., and Medin D. (2002), Essentialism and Folkbiology: Evidence from Brazil, *Journal of Cognition and Culture* 2, p. 195-223.
- Tobin E. (2013), Are natural kinds and natural properties distinct?, in S. Mumford and M. Tugby (eds.), *Metaphysics and Science*, Oxford University Press, p. 164-182.
- van Holde K.E. and Miller K.I. (1995), Hemocyanins, *Advances in Protein Chemistry* 47, p. 1---81.

*A Frame-Based Approach for Operationalized Concepts*

STEPHAN KORNMESSE

Carl von Ossietzky Universität Oldenburg

stephan.kornmesser@uni-oldenburg.de

According to Barsalou (1992, pp. 45-52), a frame for a superordinate concept is an attribute-value-matrix on the basis of which subordinate concepts of the superordinate concept can be determined. The frame in Fig. 1 represents the subordinate concepts *water fowl* and *game bird* of the superordinate concept *fowl* that are determined with respect to the values *round*, *pointed* of the attribute *Beak*, the values *short*, *long* of the attribute *Leg*, and the values *webbed*, *unwebbed* of the attribute *Foot*.

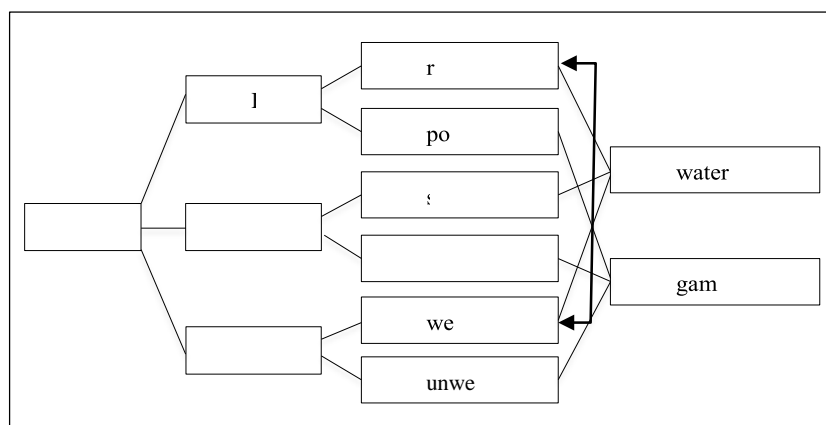


Fig. 1: Fowl-frame that determines two subordinate concepts.

Based on the propositional representation of exemplars by Barsalou (1992, p. 46), I suggest the following representation for the subordinate concept *water fowl* of the frame of in Fig. 1:

$$\text{PropWaterFowl: } \forall x(\text{Fowl}(x) \wedge \text{Beak}(x, \text{round}) \wedge \text{Leg}(x, \text{short}) \wedge \text{Foot}(x, \text{webbed}) \leftrightarrow \text{Water Fowl}(x))$$

Thus, the frame of Fig. 1 *defines* the subordinate concepts *water fowl* (and *game bird*) since the conjunction of all attribute-specific values is sufficient to be a water fowl. Hence, frames are a useful tool to represent defined concepts as well as conceptual taxonomies and can be considered to be an extension of the traditional feature list model for representing the semantic content of concepts (cf. Zenker 2014, pp. 71f.). I call frames of this kind *defining frames*.

Assume there is a nomological relation between the attributes *Beak(x,round)* and *Foot(x,webbed)* expressing that normally all and only fowl that have round beaks have unwebbed feet. In frame-based representations, such nomological relations are called *constraints* (the dotted double arrow in Fig. 1). Constraints represent the world-knowledge contained in a frame. Thus, frames do not only contain semantic

information concerning specific subordinate concepts, but also empirical information about the elements in the extension of the superordinate concept. However, there is no logical relation between the represented definitions of the subordinate concepts and the empirical information about the elements in the extension of the superordinate concepts. That is, the constraint in Fig. 1 is not entailed by the definition of *water fowl* within the defining frame. The definition contains the necessary and sufficient conditions to be a *water fowl*, but it does not say anything about a nomological relation between round beaks and webbed feet *for fowl*.

For reconstructing scientific concepts, the representation of *operationalized* concepts is at least as important as the representation of defined concepts. However, in the relevant literature on frame-based representations, there is so far no approach of representing operationalized concepts by frames. The operationalization of a concept works in another way than the definition of a concept: for a multiply operationalized concept, *each* operationalizing condition is necessary *and* sufficient for being an element in the extension of the operationalized concept. The multiple operationalization of a concept thus gains empirical content, because each operationalizing condition must be satisfied if and only if the other operationalizing conditions are satisfied (Stegmüller 1970, p. 229; Schurz 2011, p. 167f.).

In my talk, I will develop a method for frame-based representations of multiply operationalized concepts in the sense described above. Within a frame-based representation, the empirical implications of a multiple operationalization will be expressed by constraints. Thus, contrary to a defining frame, for an operationalized concept the constraints are *entailed* by the operationalizations of the concept. I call a frame of this kind an *operationalizing frame*.

In order to illuminate the idea of operationalizing frames, I will introduce a multiple operationalized concept of the linguistic theory of generative grammar according to Chomsky (1981, 1986) and provide a frame-based representation of this concept. As we will see, the graphical frame-based representation of an operationalized concept does not differ from the graphical frame-based representation of a defined concept. In order to explicate the difference, I will develop a method for frame-based representations of concepts by means of mathematical graph-theory (Hartsfield and Ringel 1990, Petersen 2007). Proposing that frames are mathematical graphs will provide a frame-based explication of the difference between defined and operationalized concepts including all advantages of frame-based representations in general. Finally, I will give an outlook on challenges for future work on frame-based representations of concepts and theories.

## References

- Andersen, H. and Nersessian, N. J. (2000): Nomic Concepts, Frames, and Conceptual Change, *Philosophy of Science*, 67 (Proceedings), 224-241.
- Andersen, H., Barker, P., and Chen, X. (2006): *The Cognitive Structure of Scientific Revolutions*. Cambridge: University Press.
- Barker, P., Chen, X., Andersen, H. (2003): Kuhn on Concepts and Categorization. In Thomas Nickles (ed.), *Thomas Kuhn*, pp. 212-245. Cambridge: University Press.
- Barsalou, L.W. (1992): Frames, concepts, and conceptual fields. In A. Lehrer and E.F. Kittay (eds.), *Frames, fields, and contrasts*, pp. 21–74. Hillsdale: Lawrence Erlbaum Associates.

- Carnap, R. (1936): Testability and Meaning, *Philosophy of Science*, 3, 419-471.
- Chen, X. (2003): Object and Event Concepts. A Cognitive Mechanism of Incommensurability, *Philosophy of Science*, 70, 962-974.
- Chen, X., and Barker, P. (2000): Continuity through Revolutions: A Frame-Based Account of Conceptual Change During Scientific Revolutions, *Philosophy of Science*, 67, 208-223.
- Chomsky, N. (1981): *Lectures on Government and Binding*. Dordrecht: Foris Publications. Chomsky, N. (1986): *Knowledge of Language*. New York: Praeger.
- Hartsfield, N. and Ringel, G. (1990): *Pearls in Graph Theory. A Comprehensive Introduction*. Boston: Academic Press, Inc.
- Petersen, W. (2007): Representation of concepts as frames. In J. Skilters, F. Toccafondi, and G. Stenberg (eds.), *Complex Cognition and Qualitative Science. The Baltic International Yearbook of Cognition, Logic and Communication*, Vol. 2, pp. 151-170. University of Latvia.
- Schurz, G. (2011): *Einführung in die Wissenschaftstheorie*. 3. edition. Darmstadt: WBG.
- Stegmüller, W. (1970): *Probleme und Resultate der Wissenschaftstheorie und der Analytischen Philosophie. Band II. Theorie und Erfahrung*. Berlin, Heidelberg & New York: Springer.
- Votsis, I., and Schurz, G. (2012): A Frame-Theoretic Analysis of Two Rival Conceptions of Heat, *Studies in History and Philosophy of Science*, 43(1), 105-114.
- Votsis, I., and Schurz, G. (2014): Reconstructing Scientific Theory Change by Means of Frames. In Thomas Gamerschlag, Doris Gerland, Rainer Osswald, Wiebke Petersen (eds.), *Concept Types and Frames. Application in Language, Cognition, and Science*, pp. 93-110. New York: Springer.
- Zenker, F. (2014): From Features via Frames to Spaces: Modeling Scientific Conceptual Change without Incommensurability or Aprioricity. In Thomas Gamerschlag, Doris Gerland, Rainer Osswald, Wiebke Petersen (eds.), *Concept Types and Frames. Application in Language, Cognition, and Science*, pp. 69-89. New York: Springer.

---

*Scientific Representation and Representation-As*

JAMES NGUYEN

London School of Economics and Political Science

j.nguyen1@lse.ac.uk

---

In this paper we propose a novel answer to the question of scientific representation: in virtue of what do scientific models represent their target systems? We make precise the idea that this is a kind of representation-as: scientific models are not only representations *of* their targets, but they also represent them *as thus or so*. The notion of representation-as was introduced by Goodman (1976), and its applicability in the context of scientific modelling has been suggested by various authors (e.g. Elgin (2010), Hughes (1997), van Fraassen (2008)). However, the specific details about how it works has have yet to be developed until now.

Our account introduces the notion of an *interpretation* of a model. Interpreting a model requires using various predicates to describe the model. Sometimes models may be interpreted in terms that literally describe them when considered as objects in and of themselves (i.e. as model-objects): a (model) ship moving through a vat of water is interpreted as a ship moving through a vat of water. But sometimes the predicates involved in the interpretation aren't of the sort usually associated with the model-object: a ball-and-stick model is interpreted as an molecule using predicates from organic chemistry; the Newlyn-Phillips hydraulic machine is interpreted as an economic system using economic predicates; a two-body system is interpreted as a celestial system using predicates from astronomy; and so on.

But an interpretation alone doesn't suffice to establish a representational relationship between a model and a target. Assuming an interpretation, we offer the following four conditions as individually necessary and and jointly sufficient in establishing that a model  $M$  represents a target system  $T$  :

1.  $M$  denotes  $T$  (and parts of  $M$  may denote parts of  $T$ ).
2.  $M$  exemplifies properties  $P_1, \dots, P_m$ .
3.  $M$  comes with a key  $K$ , specifying how  $P_1, \dots, P_n$  are to be translated into a (possibly identical set of properties)  $Q_1, \dots, Q_m$ .
4. A model user imputes at least one  $Q_i$  onto  $T$ .

Each of these conditions are discussed in detail in the paper, but the central ideas are as follows:

**Denotation.** Traditionally denotation has been restricted to the relation that holds between proper names and their bearers. But it's not obvious why this has to be the case. Goodman takes pictures to denote their subjects, and following him we liberalise it to allow that models denote their target systems. What establishes such denotation relations must be investigated on a case-by-case basis. In some scenarios fiat may be



sufficient, in others these stipulations may have to be mediated by more or less elaborate conventions. Perhaps appeals need to be made to causal chains, and so on. These questions are tied up with central questions in the philosophy of language and may require detailed analyses of cases to answer. Although denotation may suffice to establish that a model is a representation *of* its target system, it tells us nothing about how a model represents its target as *thus or so*. That is where the remaining conditions come in.

**Exemplification.** A tailor's swatch of checkered cloth exemplifies the property CHECKEREDNESS by both instantiating, and referring back to, the property in question. This two way relationship distinguishes certain properties of the swatch as epistemically salient. MADE ON A MONDAY isn't exemplified, despite the fact the swatch instantiates it. This is explained by the fact that in the typical scenarios in which the swatch is being used to show a customer the cloth options available, the day on which the swatch was made is of little interest. So it fails to refer back to this property. The same consideration applies to scientific models. As model-objects they instantiate various properties (where the term 'property' is construed liberally to include relations and higher order structural properties). And when a model user applies a particular interpretation to the object, it also refers back to some of them. When the interpretation is in terms that literally apply to the model-object, the notion of exemplification in the scientific context is the same as the one at work in the tailor's swatch case. But not all instances of scientific representation work in this way. A ball-and-stick molecular model exemplifies how many holes each different coloured ball has, but this is interpreted as the valency of different kinds of atoms. The Newlyn-Phillips model exemplifies the flow of water from one reservoir to another, but this is interpreted using economic vocabulary as the movement of money from one area of the economy to another. We investigate the relationship between interpretations and exemplification. Again, what grounds an interpretation of a model-object, and therefore the properties it exemplifies, depend on the background knowledge and assumptions at work in the discipline in question. Therefore this has to be established case-by-case.

**Key.** It's natural to think a model  $M$  represents a target  $T$  as thus or so when  $M$  represents  $T$  as having the properties  $M$  exemplifies. Sometimes this is indeed the case. The tailor could use the swatch to represent a customer's suit as checkered. But the properties need not be carried over directly. A map, which exemplifies distances between points on its surface, represents target locations as being a *scaled* map distance away from one another. And the translation can be much more complicated than a linear scale. The relationship between the forces affecting model ships and those influencing the speed actual ships is highly complex due to the non-linear relationship between the surface area of the ship's hull, and the effect that fluid viscosity has on its movement. In fact, there is much scientific value in coming up with a key that allows us to use such models in a useful manner (Sterrett 2006). Following Frigg (2010) we build the notion of a key into our account of scientific representation. Models come with a key (which again need to be investigated on a case-by-case basis) with which to translate these the exemplified properties into ones that can be imputed onto a target system.

**Imputation.** The final condition required to establish that  $M$  represents  $T$  as thus or so is that the model user actively impute the properties that result from the application of the key onto  $T$  itself. We explain why denotation and (keyed up) exemplified properties don't suffice by themselves to ensure that the model represents the target as having the result of applying the key. Again, like denotation, imputation can be analysed in terms of stipulation but may be mediated by more complicated disciplinary conventions.

Once a model user has adopted an interpretation to a model-object, if the aforementioned conditions are all met, then the model represents the target as having the properties that result from applying the key to the model's exemplified properties. As should have been clear in the above discussion, our account is intended as providing an *abstract* framework in which to think about scientific representation. The notion of an interpretation, and the conditions 1-4, are abstract in the sense that they need 'filling in' in every instance of scientific representation (and we illustrate with various examples drawn from both natural and social science). Our framework provides a novel way of thinking about scientific representation, and expands in detail on the idea that it should be thought of in terms of representation-as.

---

*A Difference-making Account of Causation*

WOLFGANG PIETSCH

Technische Universität München

pietsch@cvi-a.tum.de

---

I propose a difference-making account of causation that conceptually stands in the tradition of counterfactual approaches by adopting a definition along the lines of Hume's famous statement: "we may define a cause to be *an object followed by another, and [...] where, if the first object had not been, the second never had existed*" (Hume 1777, Sec. 7, §60). To this fundamental idea, the difference-making account adds context-dependence and a notion of causal irrelevance: *in a context B, in which a condition C and a phenomenon A occur, C is causally relevant (irrelevant) to A, iff the following counterfactual holds: iff C had not occurred, A would also not have occurred (if C had not occurred, A would still have occurred)*. In the case of causal relevance, the definition is symmetric with respect to the roles of C and A. Thus, there is an issue regarding the direction of causation, which for lack of space cannot be addressed in this talk.

According to the perspective of the difference-making account, causal (ir-)relevance is a three-place relation: a condition C is (ir-)relevant to a phenomenon A with respect to a certain context or background B of further conditions that are allowed to vary only if causally irrelevant. The restriction to a context B is required because there is no guarantee that in a different context B\*, causal (ir-)relevance of C to A will continue to hold. Since the relevant conditions in the background can only rarely be made explicit, causal relations according to the difference-making account have a distinct *ceteris-paribus* character.

A crucial difference in comparison with conventional counterfactual approaches to causation like that of David Lewis concerns the way the counterfactual conditional is evaluated. Lewis, for example, refers to the similarity between the actual and possible worlds, basically: *'If C were the case, A would be the case' is true, if some C-world where A holds is closer to the actual world than is any C-world where A does not hold* (Lewis 1973, 560). Here, a C-world is just a possible world in which C holds. For Lewis, the challenge is to find a proper construal of the notion of possible worlds and of the similarity between them.

By contrast, the difference-making approach evaluates causal counterfactuals in terms of actual instances that differ only in terms of causally irrelevant conditions: *'If C were not the case, A would not be the case' is true with respect to an instance in which both C and A occur in a context B, if there exists at least one instance in which neither C nor A occurs and if there are no instances in which C does not occur but A occurs in contexts that differ from B only in terms of conditions that are causally irrelevant for A (except for conditions that are INUS-conditions for A in virtue of being INUS-conditions for C and conditions for which C is causally relevant and which are themselves causally relevant for A)*. Obviously, this construal of counterfactuals is inspired by

Mill's method of difference. An analogous account can be given for the counterfactual 'If C were not the case, A would still be the case.' One should also define the truth-value for counterfactuals if there are no instances, in which C does not occur and that differ only in terms of irrelevant circumstances. For example, it may be the case that C belongs to a complex of conditions that occur only together. In such a case, nothing can be said about the causal relevance of C alone, only about the relevance of the factors in conjunction.

The definitions of causal relevance and causal irrelevance are the basic building blocks of the difference-making account. All other notions can be construed from these elementary definitions in terms of what one might call a specific causal signature. Consider as an example the notion of a *causal factor* C that together with some conjunct of other condition X causes a phenomenon A with respect to a background B. In terms of causal relevance and causal irrelevance, the causal signature of a causal factor is the following: C is relevant with respect to  $B \wedge X$  and causally irrelevant to A and  $\neg A$  with respect to  $B \wedge \neg X$ ; X is relevant with respect to  $B \wedge C$  and causally irrelevant to A and  $\neg A$  with respect to  $B \wedge \neg C$ .

Similarly, the notion of an *alternative cause*, that either C or some X can cause the phenomenon A with respect to a background B, can be explicated as follows: C is causally relevant to A with respect to a background  $B \wedge \neg X$ , but causally irrelevant to A and  $\neg A$  with respect to a background  $B \wedge X$ ; equally, X is causally relevant to A with respect to a background  $B \wedge \neg C$ , and causally irrelevant to A and  $\neg A$  with respect to a background  $B \wedge C$ .

Based on the notions of causal factor and alternative cause, the difference-making account can identify causation in terms of necessary and sufficient conditions for a phenomenon. More specifically, a cause established by this account can be formulated in terms of an *INUS*-condition (Mackie 1965), i.e. an *Insufficient*, but *Non-redundant* part of an *Unnecessary* but *Sufficient* condition, with the further requirement that these *INUS*-conditions must in general be seen relative to a context. Note that the account can also be extended to include functional relationships.

I will discuss how the difference-making account fares with respect to several classic objections against the counterfactual and other accounts of causation like preemption, overdetermination, or chancy causation. An important issue concerns how to prevent that definitional and other obviously non-causal relations are identified as causal. Another question regards whether pragmatic aspects can be integrated into the account such as the distinction between causes and background conditions. Finally, I will discuss several problems that arise only for the difference-making account, especially with respect to the notion of causal irrelevance.

In the end, it comes down to a matter of taste, if one prefers to classify such an account as a variant of the counterfactual approach to causation or if one considers the differences crucial enough to merit a proper name. To sum up, there are basically three main differences in comparison with conventional counterfactual approaches like the influential account of David Lewis: the construal of counterfactuals, the inclusion of a notion of causal irrelevance, and the introduction of background-dependence. The term 'difference-making' aptly reflects the fact that counterfactuals are evaluated not with respect to possible worlds but with respect

to actual situations that either really occur or are idealized from the actual world. Due to this and since no strong notion of intervention is required, the account is especially suited for contexts of application in which the evidence is mainly of observational nature such as in recent developments concerning a novel data-intensive scientific methodology (compare e.g. Pietsch forthcoming).

**References**

- Hume, David. 1777. *An Enquiry Concerning Human Understanding*. Oxford: Clarendon Press.
- Lewis, David. 1973. "Causation." *Journal of Philosophy* 70:556-67.
- Mackie, John L. 1965. "Causes and Conditions." *American Philosophical Quarterly* 12:245- 265.
- Pietsch, Wolfgang. Forthcoming. "Aspects of Theory-Ladenness in Data-Intensive Science", *Philosophy of Science*. Preprint: [http://philsci-archive.pitt.edu/10777/1/pietsch\\_data-intensive-science\\_psa.pdf](http://philsci-archive.pitt.edu/10777/1/pietsch_data-intensive-science_psa.pdf)

---

*Representation and Explanatory Power*

DEMETRIS PORTIDES

University of Cyprus

portides@ucy.ac.cy

---

One of the theses of scientific realism is that scientific theories represent facts about the world and that we have good reason to believe that our successful theories are approximately true descriptions of these facts. Antirealists would agree with the first part of the conjunct but would disagree with the second. Because of this, a major part of the debate between realists and anti-realists has focused on the latter. This debate brought to light that the thesis that our successful theories are approximately true is linked to issues of reference. The realist argues that if the entities postulated by a theory exist (e.g. electrons, protons, DNA molecules, etc.) then the success of the theory that accounts for how they behave is best explained by its approximate truth. Since the entities a theory postulates are those things its terms refer to, reference becomes a necessary condition for the success and subsequently of the approximate truth of a theory. Antirealists also agree with the latter claim and, in fact, base one of their arguments that undermine the approximate truth of theories on the failure of theories to refer, e.g. when non-existent entities such as the ether are postulated by otherwise successful theories.

In this paper I shall not engage in issues of reference that stem from this debate. I will focus instead on the first part of the realist thesis above and address the question: 'How do our theories represent facts about the world?' In science various means of representation are utilised. A diagram of an electric circuit represents its target; a graph of velocity plotted against time represents the acceleration of a particular body; a material construction of double helical structure represents a DNA molecule; a Feynman diagram represents a neutron decaying to a proton, an electron and an anti-neutrino; and a mathematical model represents the behaviour of a mass-spring system. No doubt, whether our scientific representations are diagrammatic, graphical, material, model-based, or other, they are important aspects of scientific inquiry; they enhance our understanding of the workings of physical systems and also enhance our understanding of abstract theoretical propositions. If one aims for a general theory of scientific representation then surely it must account for all these kinds of representation. The scope of my analysis, in this paper, is confined to mathematical models used in science to represent physical systems and the aim is to give an account for how such models represent their targets. There is something that makes these means of representation stick out from the rest of the group. This in my view is the fact that they are explanatory of their targets and that they yield knowledge about their targets. This claim will be clarified in the paper, where I shall argue that to understand the representational function of models it must be linked to their explanatory and epistemic

functions. This view of representation by scientific models will be contrasted to traditional views of representation.

The view that the assertions of a scientific theory relate to the phenomena in a direct way has for long been explored in the philosophy of science and has been traditionally expressed through either of two ideas: 1) that the deductive consequences of the theory stretch all the way to the phenomena or 2) that the theory represents the phenomena via its semantic models. The first, which was held by the logical positivists, is nowadays abandoned. The second, however, lives on in what has come to be known as the Semantic View of scientific theories. It is true that there are many instances in science where the view that a theory represents via its models is helpful in understanding how the theory relates to experimental measurements. To maintain, however, that this view provides the necessary and sufficient conditions for explicating how theories represent the phenomena in their scope, would restrict the scope of theories and would rule out many scientific models from having representational capacity. It is on the second kind of restriction of the Semantic View that I am interested, and in particular in that it would rule out historically successful quantum mechanical models that are initially constructed by the use of classical considerations and that at some appropriate point the classical functions are quantized. It would do so because such models cannot convincingly be regarded as directly related to quantum mechanics. In this paper I explore how modeling of the preceding kind is achieved by examining the liquid drop model of nuclear structure. I argue that importing such models into the theory rests not on a set of reductive rules by which to assign in a systematic theoretically justified way quantum mechanical properties to classical variables, but on an arbitrary heuristic that is based on a likewise arbitrary assertion that “this is a suitable place to go from the particular (i.e. classical function) to the general (i.e. quantum mechanical operator)”. This arbitrariness in our procedure of importing the model into the framework of quantum mechanics is warranted, I argue, primarily by the fact that the resulting model is successful in explaining the phenomena (i.e. the nuclear properties) and predicting experimental measurements. In view of this arbitrariness, I also argue that, the Semantic View fails to explicate how quantum mechanics -via the liquid drop model-offers a representation of the nuclear structure. On the other hand, the conception of “scientific representation” defended in this paper makes justice to how such models, as the liquid drop, represent their targets.

---

*Theoretical Contingency and Scientific Realism* (cancelled)

LUCA TAMBOLO

University of Trieste

ltambolo@gmail.com

---

The current debate on the contingency of science has been shaped by a question that Ian Hacking asked within the context of his investigations concerning social constructivism: “How inevitable are the results of successful science?” (2000; see also 1999). In other words, if science as a whole is contingent—as social constructivist claim, and scientific realists deny—then it is possible to have alternatives  $S'$ ,  $S''$ ,  $S'''$ , etc., to our current science  $S$  which, as successful as  $S$ , yield results incompatible with those of  $S$  (call this the *contingency thesis*, and call its negation the *inevitability thesis*). The contingency thesis is widely regarded as a serious challenge for scientific realism: for if the results of a successful investigation of a certain subject matter are not inevitable, and alternative results are indeed possible, then the very hard-core of realism (there is a world out there which exists independently of our minds, and which our theories aim—and often manage—to describe, at least approximately) comes under fire.

In this paper we argue that, in spite of the anti-contingency sentiment widespread among realists, certain forms of realism are compatible with qualified versions of the contingency thesis. More specifically, we suggest that certain versions of realism—championed by such authors as Niiniluoto (1999), Kuipers (2000), and Giere (2006)—are compatible with what we call *theoretical contingency*.

As we highlight in Section 2, one very sensible concern raised within the contingency vs. inevitability of science debate is that talk of the contingency of science may well boil down to counterfactual speculations concerning non-existent sciences. Indeed, claiming—as defenders of the contingency thesis do—that the history of science may well have taken a different route from the one it actually took, and that we may well find ourselves embracing the results of  $S'$  instead of those of  $S$ , is one thing. However, actually creating an alternative science  $S'$  from scratch, and ascertaining that  $S'$  is indeed as successful as  $S$ , thereby proving the contingency thesis and rejecting the inevitability thesis, is quite another thing: such an undertaking goes far beyond the possibilities of a single scientist, and in any case it is something that has so far not been accomplished (see Trizio 2008; French 2008).

In Section 3 we argue that, however, the task of addressing the issue of the contingency vs. inevitability of science no longer looks hopeless, as soon as, instead of discussing one all-encompassing contingency thesis, various *contingency theses* are considered. As a glimpse at the relevant literature suffices to reveal, a systematic, wide-ranging analysis of the notion of the contingency of science has not yet been performed, and there is not even an agreed upon terminology to describe the phenomena involved, so that different



authors addressing Hacking's question seem to be after different senses of the notion of the contingency of science. But as suggested by Soler (2008) and Giere (2010), under the heading of the "contingency of science," there is in fact a cluster of distinct although related concepts: one can distinguish, for instance, between metaphysical, methodological, conceptual, evidential, etc., contingency, and consequently, the question to be asked is whether the contingency vs. inevitability of science debate is not only a matter of degrees, but also of kinds (see especially Martin 2013, in which one finds what is to date the only attempt at a comprehensive taxonomy of the contingency of science).

It is well-known that there are probably as many versions of scientific realism around as there are authors discussing the issue. Nevertheless, as we argue in Section 4, there is a recognizable brand of realism, represented for instance by such authors as Niiniluoto (1999), Kuipers (2000), and Giere (2006), which for want of a better word, may be termed "sophisticated." In spite of other important differences among these philosophers, such brand of realism revolves around two claims: that the world can be described by indefinitely many conceptual systems, none of them enjoying a privileged status (call this the *pliability thesis*); and that not all the different conceptual systems that can be used to describe the world will be equally successful, since the world offers resistance to some attempts to describe it (call this the *resistance thesis*).

In Section 5 we propose an extension of Martin's taxonomy and introduce the notion of *theoretical contingency*—a mild version of the contingency thesis which aims at capturing the insight that it is highly contingent what kinds of questions scientists ask in the course of the exploration of the world, and therefore the results of such exploration are contingent. As we argue, in view of the embrace of the pliability thesis and the resistance thesis, sophisticated forms of realism turn out to be compatible with theoretical contingency.

## References

- FRENCH, S. (2008). *Genuine Possibilities in the Scientific Past and How to Spot Them*. "Isis," 99, 568- 575.
- GIERE, R.N. (2006). *Perspectival Realism*. Chicago: The University of Chicago Press.
- (2010). *Contingency, Conditional Realism, and the Evolution of the Sciences*. Manuscript.
- HACKING, I. (1999). *The Social Construction of What?* Chicago: The University of Chicago Press.
- (2000). *How Inevitable Are the Results of Successful Science?* "Philosophy of Science," 67, S58-S71.
- KUIPERS, Th. (2000). *From Instrumentalism to Constructive Realism*. Dordrecht: Kluwer.
- MARTIN, J. (2013). *Is the Contingentist/Inevitabilist Debate a Matter of Degrees?* "Philosophy of Science," 80, 919-930
- NIINILUOTO, I. (1999). *Critical Scientific Realism*. Oxford: Oxford University Press.
- SOLER, L. (2008). *Are the Results of Our Science Contingent or Inevitable?* "Studies in History and Philosophy of Science," 39, 221-229.
- TRIZIO, E. (2008). *How Many Sciences for One World? Contingency and the Success of Science*. "Studies in History and Philosophy of Science," 39, 253-258.

**Philosophy of the Natural Sciences**

Room 34, Thursday / Friday / Saturday 11:30 – 12:00

*Holographic Duality as a Model for Quantum Gravity* (cancelled)

SEBASTIAN DE HARO

University of Amsterdam, University of Cambridge

s.deharo@auc.nl

In this paper I discuss two philosophical questions regarding the main example of a holographic duality, namely, the so-called 'AdS/CFT correspondence': (i) The status of AdS/CFT as a model of quantum gravity, and, in particular, the question of background independence. (ii) The interpretation of the duality, developing an interpretational scheme that should be applicable to other examples of dualities.

Recent developments in string theory are deeply transforming the way physicists think about gravity. In the traditional unification programme, gravity was a force meant to be treated on a par with the other forces: the aim was for a unified description of the four forces, and strings seemed to be of help because different vibration modes of the string give rise to different particles. With the advent of holographic ideas (see 't Hooft (1993) and Maldacena (1997)), however, a slightly different view is emerging that seems to be both more concrete and more modest in its approach. The idea is that gravity admits a holographic reformulation that is generally not available in the absence of gravity. Gravity may therefore be a special force after all. On this view, the goal of understanding gravity at high energies now seems best conceptualised as consisting of two steps: 1) reformulate gravity (holographically) in terms of other forces; 2) use this reformulation to understand how gravity is quantised. Progress on the first step over the past seventeen years has been impressive; the developments on the second step have been more limited.

If holographic dualities are true, they should be of interest to philosophers of physics. For a reformulation of a gravitational theory (even at the classical level, i.e. without taking quantum corrections into account) in terms of a quantum field theory is bound to offer new ways of asking questions, such as the nature of time, questions over conventionalism of space-time, etc., which are of interest to philosophers of physics. In this paper I will be concerned with a study of a question that is a necessary prolegomenon to the use of the duality to typical philosophy of physics questions, viz. the status of the duality itself. I will limit my scope to two aspects of the main example of a holographic duality, the so-called 'AdS/CFT correspondence', or AdS/CFT, for short. I will focus on a number of necessary conditions for AdS/CFT to be:

- 1) a duality;
- 2) a theory of quantum gravity. I will also address the question of:
- 3) the interpretation of AdS/CFT.

Regarding the first question, after providing a definition of a duality, I will give two conditions that must be met in order for AdS/CFT to be an example of a duality: the structures of the observables on the two sides of the duality should be complete and identical. In this paper I will focus on the completeness condition (the condition of identity was discussed in Dieks et al. (2015)). As to the second question, I will concentrate on one of the desiderata for any candidate theory of quantum gravity: background independence. In doing so, I will address some questions that have been raised in the literature as to whether AdS/CFT is background independent. Finally, I will move on to a discussion of the interpretation of the duality, and in particular which of the two theories (if any) should be held to be the more fundamental. A stance to be taken with respect to this question implicitly also indicates a heuristic for the research in theoretical physics, so that it is best to be well aware of the options that are available and the different assumptions they entail.

---

*The Metaphysics of D-CTCs: On the Underlying Assumptions of Deutsch's Quantum Solution to the Paradoxes of Time Travel*

LUCAS DUNLAP

University of Maryland

ldunlap1@umd.edu

---

The possible existence of a closed timelike curve—a path in spacetime that takes a traveler to his own past—gives rise to the possibility of serious paradoxes. The paradoxes of time travel demand a solution if we are to take seriously the possibility of the existence of CTCs. After all, the world can't admit of a physical situation in which the actions of a time traveler prevent the creation of his own time machine. The classical proposal for solutions to the time travel paradoxes simply states that such a situation could not obtain because it is inconsistent. That is to say, the classical solution is to impose a global property of self-consistency on the events in spacetime in order to rule out the possibility of paradoxical situations arising.

David Deutsch argued in his 1991 paper “Quantum Mechanics Near Closed Timelike Lines” that, under certain assumptions, quantum mechanics can solve the paradoxes associated with time travel to the past. What bothered Deutsch about the classical solutions to these paradoxes was the element of superdeterminism implicit in them. Certain initial states of systems are ruled out by these classical solutions, in order to preserve a global consistency. This is at odds with what Deutsch identifies as one of the fundamental principles of the philosophy of science: that global constraints should not overrule our ability to act locally in accord with the laws of physics. He calls this the *autonomy principle*. The classical consistency condition violates this principle by disallowing certain initial trajectories of systems traveling along CTCs.

Deutsch showed that taking quantum effects into account allowed for a solution to the paradoxes of time travel, without disallowing any initial states of the system. He showed that for any initial condition, there is a quantum fixed point solution representing a self-consistent physical state of the system. This is achieved by allowing for mixed quantum states to obtain on the CTC—a strategy to which solutions in the classical setting do not have access.

The Deutsch closed timelike curve (D-CTC) model has been influential in the quantum foundations literature as a plausible candidate for how negative time delays would work in terms of information flow (see e.g. Brun et al 2009, and more recently Ringbauer et al 2014, and Bub and Stairs 2014). The operational features of the model are taken on board, and are considered to be unproblematic additions to the machinery of the quantum information approach.

Presumably, the justification in doing this comes from the assumption that the multiverse on which Deutsch is relying in his description of the D-CTC model is the “multiverse” of the Everett interpretation of

quantum mechanics. If this were true, it could safely be ignored by those preferring an operationalist version of quantum theory, since the Everett interpretation is, after all, unmodified quantum theory.

However, Deutsch is relying on the existence of a more general notion of the multiverse, wherein the universes are not generated as the result of the Schrödinger evolution of the universal wavefunction, leading to the branching-off of macroscopic worlds, as in the standard Everett picture. Rather, the individual universes in this case exist timelessly and in parallel, many identical with one another for at least some period of time. These are not the many worlds of the Everett interpretation.

Part of the reason that this is confusing is that Deutsch refers to both of these objects by the term “multiverse”. For my purposes, I will refer to the collective many worlds of the standard Everett interpretation as the *Many-Worlds Multiverse* (MWM). These are the result of branching of the universal wavefunction via decoherence. I'll refer to this other multiverse concept as the *Mixed-State Multiverse* (MSM). The reason for this will become clear. These universes are a kind of parallelism of existent worlds. They are not generated by the evolution of the quantum state of the universe. They timelessly exist in parallel with one another.

I will argue that a close analysis of the details of Deutsch's model shows that it cannot be so easily separated from his deep metaphysical commitments to the real existence of parallel universes. These parallel worlds are importantly different from the many worlds of the standard Everett interpretation, and as such, Deutsch's key structure is not supported by quantum theory. I will argue that the key mathematical object on which he relies for the success of his model represents a commitment to the existence of an infinite number of parallel worlds identical to the one from which the time traveling system is departing.

I will argue that, for Deutsch this is a natural solution because he is predisposed to accept the existence of parallel worlds. But for those less ready to take on the metaphysics that underlie Deutsch's model, this presents a problem. The system that is confined to the CTC does all of the work to ensure the consistency of the solution. But it is in a state that does not arise from the evolution of the quantum system from its initial conditions. This amounts to another (albeit more localized) form of superdeterminism: the creation of a CTC ensures that there happens to be a system confined to the closed timelike loop, such that its interactions with the “chronology respecting” systems will yield an output that is allowed by the consistency condition.

Finally, I'll address the following question: Is it still possible to adopt the purely operational features of his model? I'll argue that Deutsch uses the existence of MSM in the reasoning about the operation of the model. I'll show, by considering a simple example, that a purely operational acceptance of the D-CTC model would allow for predictions that Deutsch explicitly rules out. That is to say, Deutsch relies on features of the implicit underlying metaphysical picture when defining the effects of his model, and without this influence, different predictions are possible.

---

*Classical Limit from a Quantum Logical Perspective*

SEBASTIAN FORTIN  
CONICET, University of  
Buenos Aires  
sfortin@gmx.net

FEDERICO HOLIK  
CONICET, IFLP  
olentiev2@gmail.com

LEONARDO VANNI  
University of Buenos Aires  
idaeos@gmail.com

---

**1. Introduction**

In the philosophical and scientific domain of knowledge about quantum theory, the study of the classical limit offers a twofold aspect to enquire. In the first place, it leads to a critical analysis of a kind of conceptualization which aims to describe the quantum-classical transition of physical systems. In the second place, it offers a scenario of remarkable philosophical interest for the study of the inter-theoretical relationships at the epistemological and ontological levels.

The rigorous formulation of quantum mechanics was achieved after a series of papers by von Neumann, Jordan, Hilbert and Nordheim (Lacki 2000) where projection operators play a key role in the axiomatization. This is due to the spectral decomposition theorem (Reed and Simon 1972; Redei 1998), which associates a projection valued measure to any quantum observable represented by a self adjoint operator (von Neumann 1996). The set of projection operators can be endowed with an orthomodular (non-Boolean) lattice structure (Birkhoff and von Neumann 1936; Kalmbach 1983) and was named *quantum logic*, in contrast with the distributive structure of classical propositional systems (Dalla Chiara Giuntini 2004). This approach allows to compare quantum and classical systems by putting them in a common mathematical framework. Also, it is possible to use it to provide a solid axiomatic foundation for quantum mechanics and to explain in an operational way many of important features of the Hilbert space formalism (von Neumann 1996; Varadarajan 1968, 1970; Piron 1976; Mackey 1963; Jauch 1968). In this approach, a key role is played by Piron's representation theorems (Piron 1976; Soler 1995; Stubbe and Steirteghem 2007).

But it turns out that the quantum-logical approach has not addressed, up to now, of the dynamical transformation in the logic of a system. But the classical limit process offers the possibility of studying the transition from the Boolean logic of classical systems and quantum logic of quantum systems (Fortin and Vanni 2014).

**2. Decoherence and quantum logic**

In the standard approach (see Zurek 1981, 2003; Schlosshauer 2007), the classical limit is reached by the effect of the environment in the system: a large number of quantum systems interacting with a system is responsible of the selection of a preferred basis and the vanishing of the off-diagonal elements of the density

matrix in this basis. But there exists a more general approach which allows for the possibility of studying the classical limit in terms of the evolution of mean values of relevant observables of the system (Castagnino, Fortin, Laura and Lombardi 2008).

The complete description of a quantum system involves non-commutative operators and as a consequence, the lattice of quantum properties is non-distributive (Bub 1997; Cohen 1989). On the other hand, for classical systems, operators associated with properties commute with each other (the algebra of functions on phase space is commutative); thus, classical properties are distributive (Birkhoff and von Neumann 1936). According with recent works (Kiefer and Polarski 2009; Fortin and Vanni 2014) there are certain quantum systems which, under certain particular conditions, evolve in a special way: although initially the commutator between two operators is not zero, due to the time evolution it tends to zero. In other words, non-Boolean lattices become Boolean.

### 3. Classical limit of algebras and probabilities

On the basis of this observation, in this work we study the classical limit from a logical perspective by studying how the logical structure of quantum properties corresponding to relevant observables acquires Boolean characteristics. Call  $B(H)$  to the set of bounded operators acting on the Hilbert space  $H$ . We start with a definite subalgebra  $A \subseteq B(H)$ , representing the initial accessible observables of the system. In the general case,  $A$  will not be necessarily Boolean. Next, we study the time evolution of  $A$ ,  $U_t A U_t$ , and its associated projection lattice  $P(U_t A U_t)$ , for a unitary time evolution  $U_t$ . We show that depending on the special features of the chosen algebra  $A$  and the Hamiltonian of the system, the limit algebra will be Boolean only if the classical limit is reached. We also discuss the time evolved algebras for general quantum channels, i.e., for evolution operators which are not necessarily unitary. This approach allows us to discuss philosophical aspects of the classical limit by analyzing the projection lattices associated to these time evolved algebras.

At the same time, the study of the projection lattices associated to time evolving algebras allows us to analyze the evolution of the probabilities governing the system. As is well known, the Born's rule, out of which all the relevant probabilities in QM are computed, can be understood (using Gleason's theorem (Gleason 1957)), as a non-commutative probability measure in an orthomodular lattice (Kalmbach 1983). In other words, as a measure in the logic of the system (Rèdei and Summers 2007; Holik, Plastino and Sáenz 2014). By studying the time evolution of the logic, we address the problem of the evolution of the probabilities and the measures of uncertainty, such as entropy. In particular, we find that for all relevant purposes, the von Neumann's entropy becomes the Shannon's one in the limiting case.

An important advantage of this perspective lies in the fact that there is a clear criteria for classicality: if the limit algebra is Boolean, the classical limit will be reached. We will be particularly interested in the time evolution of the actual underlying logic of the system and in obtaining a visual picture of this evolution using Hasse diagrams (Kalmbach 1983; Dalla Chiara and Giuntini 2004). This allows for a clear view of how the

classical limiting process takes place. This is done by discussing the classical limit in simple models, such as a Hamiltonian with continuous spectra in the self induced approach, the two slit experiment and the Mach-Zender interferometer. Despite its simplicity, the study of these models under the light of our approach may be of importance in the understanding of quantum information theory and the open interpretational problems related to the quantum to classical transition (see Schlosshauer 2007, Chapter 8).

### References

- Birkhoff, G. and von Neumann, J. (1936). "The logic of quantum mechanics." *Annals of Mathematics*, **37**: 823-843.
- Bub, J. (1997). *Interpreting the Quantum World*, Cambridge: Cambridge University Press.
- Castagnino, M., Fortin, S., Laura, R. and Lombardi, O. (2008). "A general theoretical framework for decoherence in open and closed systems." *Classical And Quantum Gravity*, **25**: 154002.
- Cohen, D. W. (1989). *An introduction to Hilbert space and quantum logic*, New York: Springer.
- Dalla Chiara, M. L., Giuntini, R. and Greechie, R. (2004), *Reasoning in Quantum Theory*, Dordrecht: Kluwer Academic Publishers.
- Fortin, S. and Vanni, L. (2014). "Quantum Decoherence: A Logical Perspective." *Foundations of Physics*, **44**: 1258-1268.
- Gleason, A. (1957). "Measures on the closed subspaces of a Hilbert space." *Indiana University Mathematics Journal*, **6**: 885-893.
- Holik, F., Plastino, A. and Sáenz, M. (2014). "A discussion on the origin of quantum probabilities." *Annals of Physics* **340**: 293-310.
- Jauch, J. M. (1968). *Foundations of Quantum Mechanics*, Cambridge: Addison-Wesley. Kalmbach (1983). *Orthomodular Lattices*, San Diego: Academic Press.
- Kiefer, C. and Polarski, D. (2009). "Why do cosmological perturbations look classical to us?" *Advanced Science Letters*, **2**: 164-173.
- Lacki, J (2000). "The Early Axiomatizations of Quantum Mechanics: Jordan, von Neumann and the Continuation of Hilbert's Program." *Archive for History of Exact Sciences*, **54**: 279-318.
- Mackey, G. (1963). *Mathematical foundations of quantum mechanics*, New York: W. A. Benjamin.
- Piron, C. (1976). *Foundations of Quantum Physics*, Cambridge: Addison-Wesley.
- Rèdei, M. and Summers, S. (2007). "Quantum probability theory." *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, **38**: 390-417.
- Rèdei, M. (1998), *Quantum Logic in Algebraic Approach*, Dordrecht: Kluwer Academic Publishers.
- Reed, M. and Simon, B. (1972). *Methods of modern mathematical physics I: Functional analysis*, New York: Academic Press.
- Schlosshauer, M. (2007). *Decoherence and the Quantum-to-Classical Transition*, Berlin: Springer.
- Soler, M. (1995). "Characterization of Hilbert spaces by orthomodular spaces." *Communications in Algebra*, **23**: 219-243.
- Stubbe, I. and Steirteghem, B.V. (2007). "Propositional systems, Hilbert lattices and generalized Hilbert spaces." In *Handbook Of Quantum Logic Quantum Structures: Quantum Structures*, K. Engesser, D.M. Gabbay and D. Lehmann (eds.), 477-523. Amsterdam: Elsevier.
- Varadarajan, V. (1968). *Geometry of Quantum Theory I*, Princeton: van Nostrand. Varadarajan, V. (1970). *Geometry of Quantum Theory II*, Princeton: van Nostrand.
- von Neumann, J. (1996). *Mathematical Foundations of Quantum Mechanics*, Princeton: Princeton University Press.
- Zurek, W. (2003). "Decoherence, einselection, and the quantum origins of the classical." *Reviews of Modern Physics*, **75**: 715-76.
- Zurek, W. (1981). "Pointer basis of quantum apparatus: into what mixtures does the wave packet collapse?" *Physical Review D*, **24**: 1516-25.



---

*The Virtuous Climatologist**Neo-Aristotelian Virtues to Address Epistemic Uncertainties in Global Climate Models*

RAFAELA HILLERBRAND

Karlsruhe Institute for Technology

rafaela.hillerbrand@gmail.com

---

Since the 1970s, the ballpark figure of the predicted temperature increase due to manmade greenhouse gas emissions is roughly the same and ranges around 2 degrees Celsius over the course of the 21st century (Maslin/Austin 2012). The details, however, of how the climate reacts to human activities are fraught with great uncertainties (cf. Hanson 1996). Between the Fourth Assessment Report of the IPCC in 2007 and the current Fifth Assessment Report climate scientists have devoted considerable attention to the question as to how to deal with various types of uncertainties. This in particular led to a very rich and nuanced treatment of uncertainties in the current report of working group one of the IPCC, which is concerned with the physical science basis (IPCC 2013). These scientific efforts resulted in an increased understanding and an improved methodological treatment of parameter and model uncertainties in climatology that stands out within the sciences. Therefore climatology provides an excellent study case for epistemological considerations about uncertainty more generally.

In addition to advances in climatology itself, progress as regards dealing with climate-model uncertainties was also made within policy making: Over the last decade or so, there has been an increased understanding amongst governmental bodies and policy advisers that uncertainties are inevitably connected with climate modelling (Smith/Petersen 2014). Both mitigation and adaption measures hinge on input from climate models. Here, the best available scientific knowledge in the field not only comprises the knowledge of the expected increase in global mean temperature, for example, over the course of this century and the anthropogenic impact on global warming, but also understanding of the uncertainty of these so-called climate projections (e.g. Hillerbrand 2010). Uncertainties up to now are, however, mostly recognized in the political discourse in form of probability estimates. However, at least some of the aftermaths of greenhouse gas emissions cannot be captured in terms of probabilities. Policy making hence still needs to catch up with the sciences when it comes to incorporating the uncertainties of scientific findings into the decision process. But even this would not suffice for an adequate approach to uncertainty: Taking up a discussion started by Rudner, Churchman and later Douglas, Winsberg (2012) and others argue that for climate change a clear-cut distinction between normative assessment on the level of policy making and climate modelling in the empirical sciences does not exist. Instead social and moral values also enter the more descriptive analysis.

This paper addresses how recent developments in virtue epistemology may help to improve assessment of uncertainties in the sciences themselves and thus advance decision-making in the light of uncertainty. Here the focus is put on two problems, namely: the lack of quantitative probability estimates as well as the intertwining of epistemic and social or ethical issues. Virtue epistemology quite generally searches for epistemic norms by studying (hypothetical) virtuous agent. In contrast to more common approaches that focus on epistemic rules and methods, virtue epistemology looks at certain “inner” qualities of such agents. This paper begins by asking as to how the discourse as to deal with uncertainties between the forth and the current fifth IPCC report (eg. Mastrandrea et al. 2010, Adler/Hirsch Hadorn 2014) can be reconstructed from the viewpoint of virtue epistemology and what is missing from the virtue- epistemic point of view. Special attention is given here to responsiblism and Neo-Aristotelian approaches to virtue epistemology (Zagzenbski 1996). Note that sometimes the latter is seen as special variant of the former. The reason to zoom in on these two strands of virtue epistemology is twofold. Firstly, particularly responsiblism zooms in on the central role of epistemic communities in knowledge generation; the IPCC authors, reviewers and commentators can be seen as such a community. Secondly, in contrast to reliabilism, another current trend in virtue epistemology, both responsiblism and Neo-Aristotelian approaches are more complex: Their evaluative measures include knowledge that combines epistemological and ethical qualities. This seems well suited for complex applied sciences such as climatology where scientific findings directly impact on political decision-making and certain social and moral norms impact on the scientific analysis. Particularly the strong role of the *phronesis* as it can be found in L. Zagzenbski’s work will be investigated. A first task of the *phronesis* is to distinguish certain possible impacts on the climate system as ethically relevant. A second task of the *phronesis* is the application of (general) rules. Thus the *phronesis* arbitrates between, for example, general guidelines on how to deal and a specific decision context. As virtues are character traits that can be acquired and trained, the virtue epistemological view on climatology may help to improve the understanding of uncertainties and the decision making for which the uncertain scientific findings are used. In addition, the epistemically virtuous scientists may also provide a way to increase warranted trust into climate projections among the general public (cf. discussion within medical ethics on epistemic virtues, eg. Marcum 2009).

### References

- Adler CE, Hirsch Hadorn G (2014) The IPCC and treatment of uncertainties: topics and sources of dissensus, *Climate Change* 5(5), 663–676
- Hansson SO (1996) Decision-making under great uncertainty. *Philosophy of the Social Sciences*, 26(3), 369-386
- Hillerbrand R (2010) On non-propositional aspects in modeling complex systems. *Analyse und Kritik* (32), 107-120.
- IPCC (2013) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Government on Climate change*. Cambridge University Press, Cambridge.
- Marcum, JA (2009) The epistemically virtuous clinician, *Theoretical Medicine and Bioethics* 30. 249-265
- Maslin M, Austin P (2012) Climate models at their limit? *Nature* 486:184

- Mastrandrea, M.D., C.B. Field, T.F. Stocker, O. Edenhofer, K.L. Ebi, D.J. Frame, H. Held, E. Kriegler, K.J. Mach, P.R. Matschoss, G.-K. Plattner, G.W. Yohe, and F.W. Zwiers, 2010: Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties. Intergovernmental Panel on Climate Change (IPCC). Available at <http://www.ipcc.ch>
- Smith LA Petersen A (2014) Variations on reliability: connecting climate predictions to climate policy In: Boumans M, Petersen A, Hon G (eds.) Error and uncertainty in scientific practice. History and philosophy of technoscience (1). Pickering and Chatto Publishers, London, UK, 137-156.
- Winsberg E (2012) Values and Uncertainties in the Predictions of Global Climate Models, Kennedy Inst Ethics J, 22(2), 111–137
- Zagzebski, L (1996) Virtues of the Mind. An inquiry into the nature of virtue and the ethical foundations of knowledge. Cambridge: Cambridge University Press.

---

*The Colbeck-Renner Theorem as an Impossibility Theorem for Parameter Independent Hidden Variable Theories*

GIJS LEEGWATER

Erasmus University Rotterdam

gijlsleegwater@gmail.com

---

Recently, Roger Colbeck and Renato Renner (C&R) have claimed that “[n]o extension of quantum theory can have improved predictive power” [3,5]. If correct, this is a spectacular no-go theorem for hidden variable theories, which is much more general than the theorems of Bell [1] and Leggett [7]. The claim implies that if a quantum state is supplemented with hidden variables, the values of these variables have no bearing on the probabilities of measurement outcomes. Such a claim is relevant for the question, famously raised in 1935 by Einstein, Podolsky and Rosen [6], whether quantum mechanics is a complete theory. If it is not possible to introduce additional variables that have any connection to measurement outcomes, it seems such variables are redundant and can be discarded. This suggests that the quantum state itself, without additional hidden variables, gives a complete description of a physical system. Indeed, elsewhere C&R have used the above claim to conclude that the quantum state of a system is in one-to-one correspondence with its elements of reality [4].

Unfortunately, the derivation of C&R suffers from some major issues. Because of this, their work has not been widely accepted; instead, it has mainly been criticized [8,10]. In short, the derivation is hard to follow, additional assumptions seem to be necessary, and mathematically the derivation is inaccurate and incomplete. In our opinion, however, to a large extent C&R’s result can be upheld. Most of the response has been negative, dismissing the result because of the above shortcomings. Instead, in this paper we start from scratch and perform a derivation avoiding these shortcomings, to see what remains of C&R’s claim. The result is a theorem that is less general, but better founded, namely a no-go theorem for hidden variable theories satisfying Parameter Independence. The condition of Parameter independence is a locality condition, first introduced in [12]. In our case it means that, when considering two subsystems of a composite system, probabilities of measurement outcomes for one subsystem are independent on what measurement is being performed on the other subsystem.

Before turning to our own derivation, we name three of the problems of C&R’s derivation. First, it involves an assumption called “Freedom of Choice”. As the name suggests, this is an assumption involving the independence of the experimenter’s choice of settings when performing a measurement. But in the way C&R define this assumption, a no-signaling condition is actually presupposed, making the assumption much less innocent than it sounds. When using this definition, any hidden variable theory violating Parameter

Independence, such as Bohmian Mechanics, is immediately ruled out. The use of this assumption has been criticized previously [8,10]. Second, the derivation of C&R is hard to understand. This is aptly illustrated by quoting Valerio Scarani, of the Centre for Quantum Technologies at the National University of Singapore [11]:

“Beyond the case of the maximally entangled state, which had been settled in a previous paper, they prove something that I honestly have not fully understood. Indeed, so many other colleagues have misunderstood this work, that the authors prepared a page of FAQs [2] (extremely rare for a scientific paper) and a later, clearer version [5].”

The case of the maximally entangled state Scarani refers to corresponds to a more limited result that only concerns measurements on subsystems of a bipartite system prepared in a Bell state. This result states that not only the quantum mechanical outcome probabilities of such measurements equal  $1/2$ , but also the outcome probabilities in any hidden variable theory equal  $1/2$ . Some authors (see for example [9]) appear to have understood C&R's theorem as consisting of only this result, which, as Scarani already points out, had been derived before. Actually, for C&R this is only the first step in proving the much more general theorem that probabilities in hidden variable theories are *always* equal to the quantum mechanical probabilities. Third, mathematically the derivation is far from accurate and complete. This is partly obscured by the fact that the derivation is spread among multiple papers and supplementary sections. For example, more than once in their derivation C&R prove something that only holds approximately, and then subsequently use this as if it holds exactly.

C&R's derivation cannot be easily repaired just by filling in some missing steps. It turns out that at some points, a whole different proof strategy must be used than what is suggested by C&R. Therefore, we start our derivation from scratch. The issue surrounding the “Freedom of Choice” assumption can be circumvented by explicitly assuming Parameter Independence. This is why our result is limited to hidden variable theories that satisfy Parameter Independence. In terms of completeness, this means that quantum mechanics cannot be made more complete by adding hidden variables, without giving up Parameter Independence.

Our derivation is presented in five steps, where the result of each step is stronger than that of the previous step:

1. Redundancy of hidden variables for local measurements on Bell states
2. Redundancy of hidden variables for local measurements on higher-dimensional maximally entangled states
3. Redundancy of hidden variables for local measurements in the Schmidt basis, on states where the Schmidt coefficients are square roots of rational numbers
4. Redundancy of hidden variables for local measurements in the Schmidt basis, on bipartite states with any Schmidt coefficients
5. Redundancy of hidden variables for any measurement

Furthermore, we highlight the main differences with the derivation of C&R and discuss some possible issues in the derivation, as well as some foundational consequences of the result.

### References

- Bell, J. S., The theory of local beables, *Epistemological Letters*, **9**, March 1976; reprinted in *Dialectica*, **39**, 85 (1985)
- Colbeck, R., Frequently asked (and answered) questions for “Quantum theory cannot be extended”, URL: <http://perimeterinstitute.ca/personal/rcolbeck/FAQ.html>, retrieved on November 13, 2014 (2010)
- Colbeck, R. and Renner, R., No extension of quantum theory can have improved predictive power, *Nature Communications*, **2**, 411 (2011)
- Colbeck, R. and Renner, R., Is a system’s wave function in one-to-one correspondence with its elements of reality?, *Physical Review Letters*, **108**, 150402 (2012)
- Colbeck, R. and Renner, R., The completeness of quantum theory for predicting measurement outcomes, arXiv:1208.4123 (2013)
- Einstein, A., Podolsky, A. and Rosen, N., Can quantum-mechanical description of physical reality be considered complete?, *Physical Review*, **47**, 777 (1935)
- Leggett, A. J., Nonlocal hidden-variable theories and quantum mechanics: An incompatibility theorem, *Foundations of Physics*, **33**, 1469 (2003)
- Leifer, M. S., Is the Quantum State Real? An Extended Review of  $\psi$ -ontology Theorems, *Quanta*, **3**, 67 (2014)
- Di Lorenzo, A., Determination of hidden-variable models reproducing the spin singlet, *Physical Review A*, **86**, 042119 (2012)
- Ghirardi, G. and Romano, R., About possible extensions of quantum theory, *Foundations of Physics*, **43**, 881 (2013)
- Scarani, V., Physics and the Bumper Sticker, URL: <http://spreadquantum.wordpress.com/category/academic-advice/paper-writing/>, retrieved on November 13, 2014 (2013)
- Shimony, A., Controllable and uncontrollable non-locality, in *Search for a naturalistic world view*, volume 2, Cambridge University Press (1993)

**Philosophy of the Life Sciences**

Room 34, Thursday / Friday / Saturday 11:30 – 12:00

*Serendipity and Epistemic Merits of Use-Inspired Research*

BAPTISTE BEDESSEM

Laboratoire PPL

baptiste.bedessemp@gmail.com

STÉPHANIE RUPHY

Laboratoire PPL

stephanie.ruphy@wanadoo.fr

Serendipity is usually described as the faculty to make unexpected discoveries by cleverly using surprising facts (Cannon, 1965). A serendipitous discovery is thus characterized by the attention paid to unexpected observations, which leads a well-prepared mind to build an original explicative hypothesis (R-K. Merton, 1949). This notion is widely used nowadays since it has crucial epistemological and practical implications. Notably, serendipity is often cited to defend the ideal of free inquiry and scientific autonomy (Catellin, 2013). The argument presented works as follow: serendipity is a major process leading to scientific discoveries and inventions; scientific autonomy favors serendipity; then the accumulation of new knowledge, and the dynamics of innovation, strongly depend on the liberty given to the scientists in choosing the directions of their research.

As a consequence, all form of limitation of autonomy is seen as epistemologically harmful. A concrete case of such a limitation was described by Vannevar Bush (Bush, 1945), in the frame of an opposition between pure and applied science: the presence of practical objectives exerts a pressure on scientific autonomy. By imposing the resolution of practical problems, applied, or use-inspired research, would destroy the optimal conditions for serendipity to act.

However, it is well known that serendipity often acts into use-inspired science to generate inventions: from Inkjet Printers to Velcro, examples of new technologies derived from a serendipitous process are numerous (P.Thagard, 1999, 2011). From this simple remark, an interesting question arises: in the frame of use-inspired research, can serendipity act to generate fundamental discoveries? This problem leads to a more fundamental one, which questions the epistemological properties of pure and applied science: is serendipity, as a logic of scientific discovery, specific of an autonomous science?

The aim of this presentation is to discuss the relevance of the link intuitively made between serendipity as a tool for increasing scientific knowledge, and a pure science (not directly connected to practical goals). Our answer aims at identifying the optimal conditions for serendipity to act, in order to compare the ability of pure and use-inspired research to fulfill them. We use two examples in medical research (cancerology and psychiatry) to illustrate our views.

Following R-K Merton (R-K. Merton, 2004), we first propose a consensual four steps description of the serendipitous process. An unexpected observation (a) stimulates the curiosity of the scientist (b), who formulates an hypothesis (c) and validates it (d), leading to new knowledge. We show that we can derive from this description three optimal conditions for serendipity to act: the epistemic diversity of the inquiry, the ability of the researchers to pay attention to unexpected facts, and their attraction for fundamental understanding. We then consider each of these points in the frame of a distinction between pure and use-inspired research.

The epistemic diversity plays a role through the first step of the process, that is to say the occurrence of facts which could be the starting point of a serendipitous discovery. It seems reasonable to suppose that the frequency of such initiators increases with the diversity of the paths explored by the scientists. It is an argument often used to defend the epistemic value of free inquiry (Wilholt, 2010; Ruphy, 2016). We discuss this point using the notion of pragmatic pluralism recently developed by H.Longino (2013). In the case of the study of human behavior and mental pathologies, she argues that pure science is not tolerant to pluralism since it creates sterile conflicts focused on incompatible philosophical opposition, such as *nature vs. culture*. We show that a similar situation can be found in cancerology, where an artificial debate between *reductionism* and *holism* structures the search for a fundamental understanding of the disease.

On the contrary, the frame of use-inspired research, by focusing on the treatments, would be more able to accept a plurality of approaches. As a consequence, it multiplies the apparition of surprising facts-that is to say facts which cannot be explained within the current state of knowledge. Examples taken from the history of cancerology are used to illustrate this view.

The second point deals with the psychological state of the scientist: is she able to catch surprising facts? We argue that this question cannot be easily used to defend a privileged link between serendipity and pure science. Notably, the huge number of serendipitous inventions made in the frame of use-inspired research shows that an argument based on the ability to be impressed by surprising facts would certainly be difficult to build.

The third point is more problematic. If, as we concluded, use-inspired research favors the occurrence of unexplained facts, are they effectively used to build fundamental knowledge? It is widely held that serendipity can act as a logic of discovery only if the search for truth is a priority in scientist's mind (Cannon, 1965). To deal with this argumentative line, we adopt Peirce's view: the formulation of a new hypothesis leading to a serendipitous discovery needs an abductive reasoning (Tiercelin, 1993). We argue that it is not required to explicitly search for a fundamental understanding to construct this new abductive hypothesis. It can emerge indifferently from pure or use-inspired research, using a common pool of knowledge built in the frame of autonomous or finalized science.



To illustrate this view, we detail how was built the genetic theory of cancer (Somatic Mutations Theory). In particular, we show how three types of research (pure, use-inspired and applied) contributed to the abductive formulation, in a highly finalized context, of the SMT hypothesis.

This work thus provides interesting tools to compare the epistemic merits of pure and use- inspired research, in the specific frame of medical sciences. By identifying three conditions for serendipity to act efficiently (epistemic diversity, psychological state of the scientist, and desire to build new knowledge), we question the relevance of some *a priori* intuitions which makes pure science to be the privileged space of serendipitous discoveries.

### References

- Cannon, 1965, *The Way of an Investigator: A Scientist's Experiences in Medical Research*, New-York, Hafner.
- Catelin, 2013, *Sérendipité. Du conte au concept*. Seuil, science ouverte.
- Longino, 2013, *Studying Human Behavior. How Scientists Investigate Agression and Sexuality*, The University of Chicago Press.
- Merton, 1949, *Social theory and Social structure*. New-York, Free Press.
- Merton, 2004, *The Travels and Adventures of Serendipity: A Study in Sociological Semantics and the Sociology of Science*, Princeton, Princeton University Press.
- Peirce, 1931-1935, *Collected Papers*, Cambridge, Harvard University Press.
- Ruphy, 2016, *La science doit-elle être autonome pour être utile?* In Science, philosophie, société, Presses universitaires de Franche-Comté, A. Guay et S. Ruphy (eds).
- Thagard, 1999, *Scientific discovery and technological innovation: Ulcers, dinosaur extinction, and the programming language Java*. In L. Magnani, P. Nersessian, and P. Thagard (Eds.), *Model-based reasoning in scientific discovery*, New York, Plenum.
- Thagard, 2011, *Patterns of medical discovery*, in *Philosophy of Medicine*, Edited by Fred Glifford. Tiercelin, 1993, Peirce et le pragmatisme, PUF.
- Wilholt, 2010, *Scientific freedom: its grounds and their limitations*. Studies in History and Philosophy of Science 41.

---

*Mechanistic and Design Explanation in Biology (cancelled)*

DINGMAR VAN ECK

Ghent University

dingmar.vaneck@ugent.be

JULIE MENNES

Ghent University

julie.mennes@ugent.be

---

In this paper we discuss a specific type of functional explanation used in biology to explain the presence of traits, called ‘design explanation’, and compare this model of explanation with mechanistic explanations of traits. We argue that design explanations provide a key explanatory element to construct “individual level” mechanistic explanations of traits, and provide plausibility constraints on the construction of mechanistic “lineage” explanations of the evolution of traits. In-depth analysis of design explanations thus offers means to extend and refine the mechanistic program to the explanation of (adaptive) traits.

It seems uncontested that mechanistic explanation is an important model of explanation in biology (Craver 2007; Barros 2009; Calcott 2009; Illari and Williamson 2010). What are called ‘design explanations’ have received far less attention in the philosophical literature (Wouters 2003, 2007 is a notable exception). Despite this ‘relative neglect’, biologists frequently construct design explanations and consider them a relevant explanatory asset (e.g., Götmark 1987, Karl & Fischer 2012). This invites the question how design explanations relate to mechanistic ones. We take up this issue here.

Wouters (2007) argued for the complementarity of mechanistic and design explanations by showing that the latter spell out constraints on what sort of mechanisms can exist. We expand this comparative work in two ways. We argue that i) design explanations are required to construct “individual level” mechanistic explanations for advantageous traits (Barros 2009), and ii) provide plausibility constraints on the construction of mechanistic “lineage” explanations of the evolution of traits (Calcott 2009).

Design explanations explain why organisms have specific traits (rather than other ones) in terms of the advantages that such traits offer to an organism, where advantage is understood in terms of higher life chances (Wouters 2007). Advantages are often assessed in terms of *counterfactual comparisons* between extant organisms in possession of certain traits and hypothetical organisms which are similar in all respects, except that they lack the trait in question (Wouters 2003, 2007). For instance, the giant eyes of giant deep sea squid are shown to be advantageous by comparing giant squid with giant eyes with hypothetical ones having smaller eyes; large eyes enable the detection of predators (sperm whales), which would not be possible or less efficient if the squid were to have smaller eyes (cf. Nilsson et al. 2012). Such advantages offered by traits are further explained in terms of ‘functional dependence relations’, which relate advantages offered by a trait to contextual conditions (Wouters 2007, p. 76). In the case of giant squid, the advantage of large eye-size is explained, inter alia, by the fact that large eyes reduce diffraction blurring and allow for a

higher flux of photons. This, in turn, allows for smaller contrasts to be detected, thereby making it possible to detect large prey by the bioluminescence they cause in a dark pelagic habitat (Nilsson et al. 2012, p. 683). In such explanations, the presence of specific traits is explained at the organismal level and in a-historic fashion.

Interestingly, “individual level” mechanistic explanations – explanations pitched at the level of specific organisms – have also been invoked to explain the presence of traits in terms of the advantages that they confer. Barros (2009), for instance, argues in terms of a case on crab-snail predation dynamics that characterizations of crab predation-mechanisms on snails with low and high-spired shells explain why low-spired shells confer an advantage on snails that have this trait: they enable better performance of the defense- role than high-spired shells, i.e., they increase the life chances of snail individuals. The difference with design explanations is that biological advantage here gets explained in terms of characterizations of crab predation- mechanisms on snails with low and high-spired shells, rather than in terms of functional dependence relations.

We argue that something important is missing in Barros’ analysis. The conclusion that specific traits offer specific advantages is not warranted, since mere specification of the operation of a mechanism – here, crab predation mechanisms – does not reveal *why* traits, here thick low-spired shells of snails, offer an advantage: *the conditions – functional dependence relations – that need to be in place in order for a trait to confer an advantage are not specified*. In this case, inter alia, contextual relationships between internal conditions of thick shells and external conditions of crab presence – thick shells only offer an advantage when predator crabs are present. Of course such constraints are in the background of crab predation-mechanism characterizations, yet since these are not listed in individual level mechanism characterizations, the conclusion Barros’ draws does not follow. Explicit specification of the relevant functional dependence relations is required. Design explanations therefore provide a necessary complement to construct “individual level” mechanistic explanations for advantageous traits.

Moreover, mechanism characterizations do not specify functional dependencies, since functional dependence relations are *non-causal* synchronic relations (Wouters 2007). They list synchronic constraints, internal and external conditions, that need to obtain for a trait to confer a biological advantage. Yet, the relationships between these constraints are not to be construed causally: e.g., intervening to change the thickness of the shells of given snail individuals has no direct impact on crab presence, and neither vice versa. In other words, they are ‘not invariant change-relating generalizations’ (Wouters 2007, p. 75).

Design explanations can also be heuristically useful for constructing mechanistic lineage explanations. The aim of lineage explanations is to detail plausible trajectories of change in a lineage of biological mechanisms over time: they list incremental, step-by-step, changes to components of mechanisms, i.e., changes in the way mechanisms produce some phenomenon (Calcott 2009).

Both considerations of adaptive value and the existence of morphological variants found in nature may provide plausibility constraints on the modeling of changes in lineages of mechanisms over time (Calcott 2009). Design explanations may offer further plausibility constraints by connecting these two constraints, since they enable showing that variants found in nature, and depicted in a lineage, indeed are advantageous. Design explanations, hence, also provide a relevant complement to mechanistic lineage explanations.

### References

- Barros, B. (2009). Natural Selection as a Mechanism. *Philosophy of Science* 75 (3):306-22.
- Calcott, B. (2009). Lineage explanations: explaining how biological mechanisms change. *The British Journal for the Philosophy of Science*, 60, 51-78.
- Craver, C.F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. New York: Oxford University Press.
- Götmark, F. (1987). White underparts in gulls function as hunting camouflage. *Animal Behaviour*, 35: 1786-1792.
- Illari, P., & Williamson, J. (2010). Function and organization: Comparing the mechanisms of protein synthesis and natural selection. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 41, 279-291.
- Karl, I. & Fischer, K. (2012). Old male mating advantage results from sexual conflict in a butterfly. *Animal Behaviour*, 85: 143-149.
- Nilsson D., Warrant E., Johnsen S., Hanlon R. & Shashar N. (2012). A unique advantage for giant eyes in giant squid. *Current Biology*, 22: 683-688.
- Wouters, A. (2003). Four notions of biological function. *Stud. Hist. Phil. & Biomed. Sci.*, 34: 633-668.
- Wouters, A. (2007). Design explanations: determining the constraints on what can be alive. *Erkenntnis*, 67: 65- 80.

**Philosophy of the Cognitive Sciences**

Room 34, Thursday / Friday / Saturday 11:30 – 12:00

*Natural Kinds and Folk Kinds in the Psychological Sciences* (cancelled)

JOE DEWHURST

University of Edinburgh

joseph.e.dewhurst@gmail.com

This paper will examine the role that natural kinds play in psychology and cognitive science, and ask whether folk psychological kinds are capable of fulfilling this role. I will first specify what I mean by natural kinds and folk psychological kinds, and then argue that the latter are not suitable for the job required of natural kinds in the psychological sciences. Whilst folk psychological kinds constitute what Hacking calls “human kinds”, this is insufficient to qualify them for full natural kind status, even in the limited capacity outlined in this paper. Furthermore, the use of folk psychological kinds threatens to systematically undermine both theoretical and experimental work in psychology and cognitive science. For this reason, I will conclude that a concerted effort is required in order to develop new conceptual categories that more accurately reflect our understanding of the human cognitive system.

Natural kinds *terms* play a central role in scientific discourse and practice, regardless of whether or not they are referred to as such. By this I simply mean that the projectable predicates required for inductive inference resemble what we typically think of as natural kinds (cf. Quine 1970). This fact alone does not entail any stronger claims about the ontological or metaphysical status of natural kinds. It is also important to acknowledge the pragmatic (or perhaps sociological) importance of natural kind terms (Wikforss 2010, Brigandt 2011, and Khalidi 2013 come to similar conclusions), even if one were not interested in the broader philosophical debate.

It is typically the case that the projectable predicates deployed by a science will, in the first instance, follow the example set by intuitive folk taxonomies (Gopnik & Schwitzgebel 1998: 78--9). In physics and chemistry we began with the observable properties of objects, in biology we began with obvious environmental and physiological groupings, and in psychology and cognitive science we typically begin with folk psychological taxonomies. A key difference here is that whilst physics, chemistry, and biology have all at least partially transcended their folk taxonomical beginnings, in the psychological sciences we are by and large still stuck with folk psychology. We must ask, therefore, whether the folk psychological taxonomy is fit for purpose.

Whilst there is no general agreement as to which account of natural kinds is correct, it is at least broadly acknowledged that to be fit for purpose in the biological and psychological sciences, an account of natural

kinds should allow for a degree of flexibility in membership conditions. Either we find such an account, or we must conclude that the kinds of biology and psychology are not natural kinds. A promising candidate for such an account is some version of the homeostatic property cluster theory, which claims that (at least some) natural kinds consist of regularly co-occurring clusters of properties along with a homeostatic mechanism that explains the co-occurrence of those properties (see e.g. →Kornblith 1993: 35, Boyd 1999, Magnus 2012). Accounts of this kind are fairly liberal, and for the purposes of this paper I will take them as a yardstick against which to measure the success of folk psychological kinds. If they fail here, then they are unlikely under any more stringent account of natural kinds.

There are two reasons to think that folk psychological kinds might not be natural kinds. The first has to do with the extent to which folk psychological explanation and discourse varies across cultures and languages. Given that different cultures draw on different taxonomies when attributing mental states (see e.g. Lillard 1998, Turner 2012), it seems that we cannot simply read off a 'correct' taxonomy that will correspond to the natural kinds of psychological science. Of course it might be the case that genuine psychological kinds will correspond to some folk psychological kinds, but, prior to experimentation, there is no way of knowing which these will be. We certainly cannot assume that the folk psychological kinds of our own culture or language will correspond precisely to the kinds of a finished psychological science.

The second reason for thinking that folk psychological kinds are not natural kinds will apply even if one was able to uncover some cultural universals that were not vulnerable to my first argument. By and large, folk psychological kinds are not suitable for fine-grained scientific enquiry. Consider the archetypal folk psychological kinds, *belief* and *desire*. Whilst they are prevalent in philosophical thought experiments, these terms rarely feature in scientific psychology. When they do appear, they are used to refer to a far more disparate set of concepts than the folk kinds encompass (see e.g. Krueger & Grafman 2013). This means that folk psychological kinds are disjunctive in a way that is ruled out by most contemporary accounts of natural kinds (see e.g. Khalidi 2013: 89--92). Without further refinement, folk psychological kinds are not suitable for the role required of natural kinds in the psychological sciences (i.e. projectability across different domains).

Given that folk psychological kinds appear not be natural kinds, what kind of a thing are they? They certainly appear to be projectable in at least some non-scientific context, such as when they are used to predict the coarse-grained behaviour of conspecifics. It is explanatory power in this sort of context that defenders of folk psychological kinds tend to appeal to. However, it is also precisely this sort of context that introduces the problems raised by Hacking with regard to what he calls "human kinds" (1995, 2006). Folk psychological kinds are only projectable in social contexts, where they are dependant upon the looping effects described by Hacking and more recently explicated by Zawidzki (2013) as "mindshaping". That is to say, folk psychological kinds only have explanatory power when the very act of using them enforces their own validity by shaping the way in which we behave and think. They lose this explanatory power as soon as

we descend below explanation in the social domain, and as such are ill suited for the role required of natural kinds in any more general account of psychology and cognitive science. We must therefore look elsewhere for a psychological taxonomy that is fit for purpose.

### References

- Boyd, R. 1999. "Homeostasis, species, and higher taxa." In Wilson (ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press.
- Brigandt, I. 2011. "Natural Kinds and Concepts: A Pragmatist and Methodologically Naturalist Account." In Knowles & Reidenfelt (eds.), *Pragmatism, Science and Naturalism*. Peter Lang.
- Gopnik, A. & Schwitzgebel, E. 1998. "Whose Concepts Are They, Anyway? The Role of Philosophical Intuition in Empirical Psychology." In De Paul & Ramsey (eds.), *Rethinking Intuition*. Lanham: Rowman & Littlefield.
- Hacking, I. 1995. "The Looping Effects of Human Kinds." In Sperber, Premack, & Premack (eds.), *Causal Cognition, an Interdisciplinary Approach*. Oxford, UK: OUP.
- Hacking, I. 2006. "Kinds of People." *Proceedings of the British Academy* 151: 285-318.
- Khalidi, M. A. 2013. *Natural Categories and Human Kinds*. Cambridge, UK: CUP.
- Kornblith, H. 1993. *Inductive Inference and Its Natural Ground*. Cambridge, MA: MIT Press.
- Krueger, F. & Grafman, J. (eds.) 2013. *The Neural Basis of Human Belief Systems*. Hove, UK: Psychology Press.
- Lillard, A. 1998. "Ethnopsychologies." *Psychological Bulletin*, 123/1: 3-32.
- Magnus, P.D. 2012. *Scientific Enquiry and Natural Kinds*. Basingstoke, Hampshire: Palgrave MacMillan.
- Quine, W.V.O. 1970. "Natural Kinds." In Rescher et al (eds.), *Essays in Honor of Karl G. Hempel*. Dordrecht: D. Reidel.
- Turner, R. 2012. "The need for a systematic ethnopsychology." *Anthropological Theory*, 12/1: 29-42.
- Wikforss, A. 2010. "Are Natural Kind Terms Special?" In Beebe & Sabberton---Leary (eds.), *The semantics and metaphysics of natural kinds*. London: Routledge.
- Zawidski, T. 2013. *Mindshaping*. Cambridge, MA: MIT Press.

---

*Reframing the Problem of Cognitive Penetrability*

ATHANASSIOS RAFTOPOULOS

University of Cyprus

raftop@ucy.ac.cy

---

Under the influence of the work of Hanson (1958) and Kuhn (1962), many philosophers of science espoused the view that cognitive states determine what we perceive. Perception became theory-laden, conceptually modulated, and cognitively penetrated (CP). In parallel, Sellars (1956) attacked one of the main 'dogmas' of classical empiricism, to wit, the view that perception functions independently of concepts and delivers to us the world without any conceptual influences. This 'given', empiricists thought, can be used as a neutral basis on which to provide justification or rational support, and determine the truth, of perceptual beliefs and scientific theories. The rejection of this assumption undermined the justificatory role of perception in grounding beliefs since the fact that prior beliefs can affect perception may introduce a form of confirmation bias that is, epistemically speaking, problematic. Undermining the justificatory role of perception paved the way to constructivism. *Epistemological Constructivism* argues that our experience of the world is mediated by our concepts. Thus, we cannot examine directly which aspects of objects belong to them independently of our conceptualizations. This clashes with epistemological realism's thesis that perception relates mind-independent objects and us.

CP is thought to encompass cognitive influences on perception, where cognition is widely understood so as to include emotive states, such as desires, hopes, etc. Not all cognitive influences are deemed to be cases of CP. One should, therefore, explain which cognitive effects constitute cases of CP. Moreover, discussions concerning the effects of the CP of perception for the epistemic role of perception in grounding beliefs center on whether the cognitive effects diminish the justificatory role of perception by rendering it less sensitive to the data and less reliable. When this happens, CP epistemically downgrades perception. CP downgrades perception because it affects perceptual processing in a way that renders the percept, the output of perception on which perceptual beliefs are grounded, epistemically suspect, by raising concerns about whether the percept reflects more the environment or the contents of the penetrating cognitive states. CP may affect the sensitivity of perception to the data by causing perception to select those data from a visual scene that confirm the contents of the penetrating cognitive states and ignore other data that do not. That is, it may bias perceptual processes to favor the viewer's expectations. It is also accepted that not all cases of CP downgrade perception. The problem is to explain why some cognitive effects downgrade perception, while others do not.



To address these concerns, I propose reframing of the problem of cognitive penetrability (CP) so that it incorporate two factors that largely have been ignored in philosophical accounts of CP. The first factor is that when discussing the CP of perception, one should distinguish between early vision and late vision and examine the cognitive effects on each stage because cognition affects them differently. Also, to examine the effects of cognitive influences for the epistemic role of perception, one should delineate the epistemic contributions of early vision and late vision.

All accounts of CP assume that although CP is a term covering cognitive influences on perception, not all cognitive influences on perception constitute cases of CP. When cognitive factors affect perception by introducing an external link to perceptual processing, as when they make the viewer refocus her attention to some part of the environment changing the percept, this is unanimously considered not to be a case of CP. Moreover, not all instances of CP downgrade perception; some even enhance its justificatory role. To discuss adequately CP and its epistemic effects, one has to examine carefully the way the cognitive states affect perceptual processing. To this end, and this is the second factor that I bring into the discussion, I distinguish between direct or intrinsic and indirect or extrinsic cognitive effects on perception and discuss their function, their impact on defining CP, and the way they affect the epistemic role of perception.

I claim that direct cognitive effects on perception are clear cases of CP because cognition affects perceptual processing on-line, that is, while perceptual processing takes place, and, hence, changes the way visual information is processed. Late vision, which is directly affected by cognition, is CP. This sort of effects influences the epistemic role of perception and *may* lead to harmful epistemic consequences for perception, but it *may* also enhance its epistemic role.

Any other cognitive effects on perception that cannot in principle affect the epistemic role of perception do not count as cases of CP. The main reason for this is because CP was conceived in such a way as to cause various epistemic problems for perception. Since as I have argued (author) there are no direct cognitive effects on early vision, the visual processes that retrieve information from the environment are not affected by cognitive states. There are, however, indirect effects on early vision, such as pre-cueing, which, however, do not affect perceptual processing itself but create the context in which it takes place. I argue that the indirect effects on early vision do not diminish the sensitivity to the distal data and are, epistemically neutral, which entails that early vision is CI.

The fact that early vision is only indirectly affected by cognitive state and, thus, the cognitive states do not influence information retrieval from the visual scene ensures that all data from the visual scene are in the proximal image and are available to late vision processing. I will also show why this fact allows us to explain, first, why not all cases of CP downgrade perception. It also allows us to explain why in most, if not all, cases in which CP downgrades perception, the bad cognitive effects can be mitigated; this is so because the proximal image can be revisited and the right data can be selected from it leading to the formation of the

right percept, that is, of the percept that best reflects the visual scene. Such revisiting is very important in rebutting constructivism, especially in the Philosophy of Science.

**Philosophy of the Social Sciences**

Room 34, Thursday / Friday / Saturday 11:30 – 12:00

*Modeling the Social Organization of Scientific Research: Lessons from Econometrics* (cancelled)

MANUELA FERNÁNDEZ PINTO

University of Helsinki

mfernand3@nd.edu

CARLO MARTINI

University of Helsinki

uni.c.martini@gmail.com

In recent decades, computer simulations have become a central tool for studying the organization of complex systems in a variety of scientific fields, such as physics, cosmology, biology, climate science, economics, sociology, and political science. More recently, social epistemologists have begun to follow this trend, using computer simulations to model the social organization of scientific research (SOSR, for short) (see e.g., Zollman 2007; De Langhe & Greiff 2009; Hegselmann & Krause 2009; Weisberg & Muldoon 2009). Computer simulations are taken to be very valuable tools for modeling the complexity of the internal dynamics of scientific research, and for overcoming some of the limitations of previous analytical models of the social organization of science (e.g., Kitcher 1990; Goldman & Shaked 1991).

Critical assessment of computer simulations in social epistemology, however, is yet to be articulated. In this paper, we take a first step in this direction. Our aim is two-fold. First, we aim at clarifying the relation between computer simulation models in social epistemology and their target systems. Second, through a comparison with computer simulation models in econometrics, we aim to show that for purposes of optimization and prediction social epistemologists ought to connect the results of their computer simulations to their target systems in the real world. We take the latter to be a challenge for SOSR simulation models to become more empirically informed, in the same way that, over time, economic modeling gave way to empirically valuable econometric analysis by linking models with data. We argue that the challenge can be taken as a possible point of criticism for current simulations of SOSR, but also as an avenue for future research into the social organization of scientific research.

The paper has three main sections. The first section looks at recent developments in the literature on the social organization of science, mostly in philosophy of science and social epistemology. Some of the early models by Philip Kitcher, Michael Strevens, and Alvin Goldman tried to show that non-epistemic dynamics in science — for example, individual scientists' pursuit of personal credit for career advancement — may actually contribute to the epistemic aims of science, such as discovering the truth about phenomena. Those models, however, often had to greatly simplify the scenarios due to the limitations of analytical treatment. Accordingly, the recent introduction of computer simulations, which allow more complexity to be added to the described scenarios, has been appraised as a more realistic way of modeling SOSR. But to some critics

even current computer simulation models are in need of more complexity in order to achieve realistic and accurate descriptions of their target systems (see Payette, 2012). More and more papers are being published with finer-grained models and more sophisticated analyses of the social mechanisms of scientific research.

In the second section, we address the main question of the paper, i.e., whether the race to complexity is well-directed or rather misguided. It is hard to resist a comparison of the events in the research fields of SOSR studies with the relatively recent history of economics. Modern economic theory has been focused on developing models of the economy: equilibrium models, trade models, monetary models relating interest rates with macro variables like GDP or employment, just to mention a few examples. But, as soon as computing became available to scientific efforts, economists realized that many of the limits of analytical models could be overcome with computer simulation models, just as social epistemologists realized, later on, that many of the limits of the initial analytical models of SOSR could be overcome with computer simulation models. Complex models of the economy could be connected in a relatively straightforward way to empirical data, and thus parameterized and used for both explaining otherwise puzzling phenomena (e.g., Hobijn and Frances 2000, 2011), or predicting economic trends (e.g. monetary models in use at most central banks). In this section, we use two econometric models, a model of the relationship between expected income and migration decisions (Kennan & Walker 2011) and a model of unemployment memory (van Dijk et al. 2002) to illustrate how economists bridge the gap between computer simulation models and their target systems. We compare these econometric models with three SOSR models in social epistemology: the social network model (Zollman 2007), the epistemic landscape model (Weisberg & Muldoon 2009), and the unified model (De Langhe 2014).

In the third section, we argue that the complexity afforded by simulation models of SOSR is of little value if that research stream doesn't take lessons from economics and econometrics, and think about ways to parameterize their models and link them with data from the target systems. Ultimately, we argue, the empirical challenge of SOSR studies is whether it is feasible to connect simulation models with data, in which case the challenge can be addressed and a new direction of research can be taken, perhaps aided by lessons from econometrics methodology; or whether the game is not worth the candle, and other ways of investigating SOSR may be more fruitful: for example, case studies, historical analyses, conceptual analyses, etc.

To conclude the paper, we present both optimistic and pessimistic scenarios about how the challenge may unfold. For the more optimistic scenario in which the challenge can ultimately be addressed, we present a set of methodological principles that SOSR research should follow. For the more pessimistic scenario, we pose a number of sub-challenges to the idea that SOSR research can proceed by way of modeling; among these the most important one is the problem of measurement of the variables involved in SOSR models. Whether philosophical research on SOSR will be able to face the empirical challenge or not is itself an empirical question, but one that should be informed both by the development of computer simulation

models in the social sciences as well as by rigorous philosophical analysis about the proper aims and feasible achievements of SOSR models.

### References

- De Langhe, Rogier. 2014. "A Unified Model of the Division of Cognitive Labor." *Philosophy of Science*, 81(3): 444-459.
- Goldman, Alvin I. and Moshe Shaked. 1991. "An Economic Model of Scientific Activity and Truth Acquisition." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 63(1): 31-55.
- Hegselmann, Rainer and Ulrich Krause. 2006. "Truth and Cognitive Division of Labour. First Steps towards a Computer Aided Social Epistemology." *Journal of Artificial Societies and Social Simulation*, 9(3)1-10.
- Hobijn, Bart and Philip Hans Franses. 2000. "Asymptotically Perfect and Relative Convergence of Productivity." *Journal of Applied Econometrics*, 15, 59–81.
- Hobijn, Bart and Philip Hans Franses. 2001. "Are Living Standards Converging?" *Structural Change and Economic Dynamics*, 12, 171–200
- Kennan, John and James Walker. 2011. "The effect of expected income on individual migration decisions." *Econometrica* 79(1): 211-251.
- Kitcher, Philip. 1990. 'The Division of Cognitive Labor.' *The Journal of Philosophy*, 87(1):5-22. Kuhn, T. S. 1977. *The Essential Tension: The Selected Studies in Scientific Tradition and Change*. University of Chicago Press: Chicago.
- Payette, Nicolas. 2012. "Agent-Based Models of Science." In *Models of Science Dynamics, Understanding Complex Systems*. A. Scharnhorst et al. (eds.). Berlin Heidelberg: Springer-Verlag (2012).
- Van Dijk, Dick, Philip H. Franses, and Richard Paap. 2002. "A nonlinear long memory model, with an application to US unemployment." *Journal of Econometrics* 110:135-165.
- Weisberg, Michael and Ryan Muldoon. 2009. "Epistemic landscapes and the division of cognitive labor." *Philosophy of Science*, 76(2):225–252.
- Zollman, Kevin. 2007. "The communication structure of epistemic communities." *Philosophy of Science* 74(5): 574–587.

**Formal Approaches to Philosophy of Science**

Room 34, Thursday / Friday / Saturday 11:30 – 12:00

*Italian Pragmatism and De Finetti's Philosophy of Probability*

COLIN ELLIOT

University of Tilburg

c.elliott@uvt.nl

De Finetti's philosophy of probability is highly influential in current philosophy of science: he is one of the founders of subjective Bayesianism. Two main philosophical ideas inform de Finetti's thought: operationalism and pragmatism. The former, however, calls for the identification of a phenomenon with the operations performed to measure it; but de Finetti saw probability as a primitive concept, existing independently of its measurement. A satisfactory explanation of this tension does not exist in the literature. Studying de Finetti's philosophical influence, the pragmatism of Vailati and Calderoni, we are able to give a new detailed and straightforward explanation for the link between de Finetti's operationalism and his pragmatism. Our reading sees the need for the former as dictated by the definition of 'meaning' given in the latter. Our approach also affords an interesting view on other salient aspects of de Finetti's philosophy, such as his verificationism and subjectivism.

The following passage is emblematic of the tension in de Finetti's operationalism:

In order to give an effective meaning to a notion—and not merely an appearance of such in a metaphysical-verbalistic sense—an operational definition is required. By this we mean a definition based on a criterion which allows us to measure it [...]. The criterion, the operative part of the definition which enables us to measure it, consists in this case of testing, through the *decisions* of an individual (which are observable), his *opinions* (previsions, probabilities), which are not directly observable.

In this passage, a footnote refers the reader to Bridgman's book [2]. The following is a quote which can convey, in a 'slogan', the message of that book: "In general, we mean by any concept nothing more than a set of operations [by which it is measured]; *the concept is synonymous with the corresponding set of operations*" [2, p.5] (emphasis in original). As Eriksson and H'ajek note in [3], this sentiment is echoed in the first paragraph, but immediately qualified, perhaps even contradicted, in the second. They suggest that perhaps de Finetti was *not* a strict operationalist. M.C. Galavotti, who has written extensively on the topic, does not go down this path, but seems, in earlier writings, to put more emphasis on the pragmatic (more in the sense of 'practical') role of operational definitions. In Galavotti's accounts, the pragmatic aspects are: the need to put numbers to beliefs in order to guide decisions, and the continuity, in de Finetti's work, between

inductive reasoning and inductive behaviour, and between decisions in everyday life and in science. In more recent work [4], Galavotti examines the broader philosophical perspective that de Finetti borrows from pragmatism.

We think that these existing accounts cannot explain, in detail, the dilemma in the passage above. We argue that a satisfactory answer is to be found in de Finetti's pragmatism, which is, to a great extent, Vailati and Calderoni's pragmatism. Now, according to this school of thought, the *meaning* of a proposition is that set of experiences which we claim will be produced, or would be produced given specific circumstances [6, p.41]. A separate matter is the truth or falsity of the proposition: this is decided by testing the verifiable consequences that the proposition claimed [6, p.44]. But what, then, is the meaning of a probabilistic sentence? Note that a degree of belief of 0.99 that it will rain tomorrow is not proven false by the lack of rain tomorrow. Take this proposition: *S*: "my degree of belief about event *E* is *a*". *If considered as a proposition on E*, this is simply an opinion: Vailati and Calderoni excluded opinions which have no verifiable predictive import from the 'meaningfulness-test', and de Finetti repeats time and again the exact same concept: it is misguided to say that a probability opinion is right or wrong. However, and this is crucial, we do not want propositions like *S* to be meaningless in general. Then, for it to have a meaning in the pragmatist sense, there must be some predictive content which is verifiable. What is this checkable predictive content? It is precisely that given by the operational definitions: *S* contains the claim that I will choose *a* as my betting odds, or display this belief in any other equivalent operational definition. It seems to us that this is a neglected aspect of de Finetti's philosophy: operationalism is needed, because of the 'meaningfulness-test' for propositions required by Vailati and Calderoni's pragmatism. Note how this understanding of meaningfulness favours subjective probability over objectivist interpretations: indeed, if we say 'the objective probability of event *F* is *b*', we have no direct way of verifying this proposition. Furthermore, this understanding of de Finetti's position finds confirmation in the Italian mathematician's own writings (see [5, p.83]).

Vailati and Calderoni also explained that other people's beliefs are not verifiable directly, but can be checked by consequences we can derive from them [6, p.52]. This is exactly the role that decisions play for de Finetti's probability. Other interesting comparisons between Vailati & Calderoni and de Finetti are possible. For example, the former argue that properly speaking, there are no repeated events [6, p.72]. De Finetti echoes this sentiment in a number of his works (see [5, p.7]). Lastly, Berkovitz [1] is right in suggesting that de Finetti would choose a weaker version of verificationism. Vailati and Calderoni say that propositions can be meaningful even if they are only *conceivably* checkable in the future. De Finetti again echoes them ([5, p.6]).

Summing up, we believe that viewing de Finetti's philosophy of probability through a detailed look at his pragmatist influences, we can understand a fundamental but puzzling aspect of his thought, add to the analyses given by Galavotti, and answer questions raised in recent work by Berkovitz and Eriksson & Hájek.

**References**

- [1] Berkovitz, J. (2012). The world According to de Finetti: On de Finetti's Theory of Probability and its Application to Quantum Mechanics. In *Probability in Physics*, Y. Ben Menahem and M. Hemmo (eds.), The Frontiers Collection, Springer, pp. 249-280. Collection, Springer, pp. 249-280.
- [2] Bridgman, P. W. (1960). *The logic of modern physics*. The Macmillan Company, New York.
- [3] Eriksson, L., Hájek, A. (2007). *What Are Degrees of Belief?*, *Studia Logica* 86: 183-213.
- [4] Galavotti, M. C. (2011). Probability and pragmatism. In *Explanation, Prediction, and Confirmation*, The Philosophy of Science in a European Perspective 2, D. Dieks et al. (eds.), Springer, pp.499-510.
- [5] de Finetti, B. (1974). *Theory of Probability: A critical introductory treatment*, Volume 1. John Wiley & Sons.
- [6] Vailati, G. (1976). *Metodo e ricerca*, Lorè, B. (ed.), R. Carabba Edotpre, Lanciano.



---

*Explaining Scientific Collaboration: On the Epistemic Efficiency of Groups in a Competitive Environment*  
(cancelled)

CYRILLE IMBERT

CNRS, Université de Lorraine

cyrilleimbert@gmail.com

THOMAS BOYER-KASSEM

CNRS, Université de Lorraine

boyerthomas0@gmail.com

---

## 1. Introduction

Scientific collaboration has kept developing since the 19th century (Beaver and Rosen, 1978, 1979a and 1979b). More and more articles are now produced collaboratively, by larger and larger teams. There is little doubt that this trend can be explained by various factors, which can be non-epistemic (e.g. psychological factors, artifacts of credit or citation, ...) or epistemic (e.g. the possibility to generate new ideas, the need to put together specialists from different fields, ...), or by a combination of such factors.

Our claim is that, to explain scientific collaboration, it can be sufficient to invoke small differences in the efficiency of collaborative groups in passing the various steps of a research project (whatever the origin of these gains in efficiency). We rely on the results of a formal model proposed in Boyer-Kassem and Imbert (forthcoming), but tackle a different question. Whereas their model gives *normative* statements, we aim at providing important results about the ways in which collaboration can be *explained*. For this reason, the core of the present paper is devoted to discussing how such results in terms of efficiency can have explanatory power, and we shall argue that resorting to a functional explanation is appropriate in the present case.

## 2. The model and its results

Boyer-Kassem and Imbert (forth.) propose a model in which the performance of groups are mostly determined by those of individuals. They consider  $n$  scientists, working on a  $l$ -step long research project. Researchers have an individual probability  $p$  to pass a step per unit time. Only the last step is publishable and the priority rule applies. Researchers can collaborate or not. Collaborating only means sharing passed steps and equally splitting the final reward. As the authors note, these hypotheses do not specifically favor collaboration, which boils down to information sharing (e.g., two heads are not supposed to be more than twice better than one, *contra* Thagard), so the gains in step-efficiency are moderate.

Collaboration configurations are simulated up to  $n = 10$ , and individual rewards per unit time are compared. Roughly, the morale that Boyer-Kassem and Imbert draw is that individuals are better off collaborating.

Two other results are of particular interest here:

- (1) There is no such thing as *the* successfulness of a  $k$ -group. This quantity is contextual since it depends

on the collaboration configuration of the other agents in the community.

(2) Small differences in the step efficiency of  $k$ -groups can result in larger differences in overall individual or group rewards per unit time, with sometimes a factor 10 or 20.

Overall, it is not sure that these results, taken at face value, can have general explanatory import, especially if they are too contextual. This is one of the important point that we analyze in the remainder of the paper.

### 3. Explaining collaboration

The results from the model are normative: they describe what is best for a rational individual in some sets of circumstances — roughly: collaborate! A usual economist way of thinking is that providing a normative model about optimal situations can sometimes be explanatory. However, it is unlikely that this is here sufficient since, in the larger social sciences, philosophers are usually more demanding (cf. Kincaid 1996, Pettit 1996).

Functional arguments, when they are well-justified, can provide convincing explanatory patterns in the social sciences: a social practice  $P$  exists “in order to” promote effects  $E$ , or the practice  $P$ ’s function is to bring about effects  $E$  (e.g. initiation rites exist in order to promote social cohesion). Similarly, we shall defend here the claim that scientific collaboration develops because it promotes individual epistemic successfulness. We shall rely on Kincaid’s definition of functional explanation, which says that  $P$  is functionally explained by  $E$  if:

- (a)  $P$  causes  $E$ ,
- (b)  $P$  persists because it causes  $E$ ,
- (c)  $P$  is causally prior to  $E$ .

We then argue that these conditions, and in particular the first two, are met in the present case. The results of the model, taken from an aggregated perspective, can be used to give good grounds for (a), by relying on the above result (1). Importantly, the aggregated perspective shows that (a) is robust under various changes in the size of the group, the size of the community and what other agents do, which answers the above-mentioned worries about contextuality. That collaboration persists because it causes individual epistemic successfulness (condition b) can be plausibly argued independently from the model, by pointing out that the ability to collaborate is usually transmitted between researchers in the process of collaboration itself, especially to students, and that pragmatic rewards accrue to epistemically successful scientists.

### 4. Conclusion

Importantly, our results do not depend on the origin of the epistemic step efficiency: the model shows that differences in the step-efficiency of a group are much amplified at the group successfulness level. Actually, any source of small differences in the epistemic step efficiency of groups (that is, not simply information sharing, as in the model) may also work: epistemic ones (specialization, generation of new ideas, error

checking, ...) or non-epistemic ones (money for hiring technicians/secretaries, for accessing data, for buying instruments, ...). Thus, the argument shows the potential explanatory force of any factor that improves the step efficiency of groups.

Finally, competition, which comes from the enforcement of the priority rule, is a crucial aspect of the model. In this respect, the model involves a social aspect and the corresponding explanation of collaboration is also partly social and departs from the usual purely epistemic explanations. In the same time, as argued just above, the scope of this explanation is large since it can be applied to any case in which a mechanism improves the step efficiency of groups. Therefore the present explanation also departs from purely sociological explanations of collaboration.

### References

- Beaver, Donald deB. and Rosen, Richard (1978), "Studies in Scientific Collaboration: Part I", *Scientometrics* 1(2): 65–84.
- (1979a), "Studies in Scientific Collaboration: Part II", *Scientometrics* 1(2): 133–149.
- (1979b), "Studies in Scientific Collaboration: Part III", *Scientometrics* 1(3): 231–245.
- Boyer-Kassem, Thomas and Imbert, Cyrille (forthcoming), "Scientific collaboration: do two heads need to be more than twice better than one?", forthcoming in *Philosophy of Science*.
- Kincaid, Harold (1996), *Philosophical Foundations of the Social Sciences: Analyzing Controversies in Social Research*, Cambridge University Press.
- Pettit, Philip (1996), "Functional Explanation and Virtual Selection", *The British Journal for the Philosophy of Science*, Vol. 47, No. 2 (Jun., 1996), p. 291–302.

**Historical, Social and Cultural Studies in Philosophy of Science**

Room 34, Thursday / Friday / Saturday 11:30 – 12:00

*The Objectivity of the Humanities: Hempel, Carnap and the Case of Lucien Febvre*

FONS DEWULF

Universiteit Gent

fons.dewulf@ugent.be

One of the most famous criteria to demarcate the humanities from the natural sciences is the method of “Verstehen”. Empathetic understanding of the actors, as “Verstehen” is often understood, is typically considered as a subjective method and thus as a problem for the objective nature of these sciences. This position was famously spelled out by Carl Hempel in his paper “The function of General Laws in History”. Hempel argued that the historian can only explain phenomena from the past through lawful generalization. This notion of explanation also has to account for the objectivity of historiography. Objective checks can be made through empirical testing of the determining conditions, the universal hypotheses and through the logical investigation of the explanations offered. (Hempel 1942, 38) On this view, the method of “Verstehen” is merely a heuristic device that the historian uses to generate hypotheses. (Hempel 1942, 44) Therefore, the objective nature of history relies solely on the possibility of verification of the explanations offered, while the subjective method of “Verstehen” only has a preliminary, heuristic role to play.

Hempel's position, however, has certain problems which are specific to the humanities. I will illustrate these problems with Lucien Febvre's “Le problème de l'incroyance au XVIe siècle”. Febvre's work is a classic of 20th century historiography and has generated a lot of controversy. So in this respect it is interesting to see how these debates can be arbitrated: to what extent and how can the objectivity of his claims be put into question? In the work Febvre argues among other things that certain mockeries of the virgin Mary in the sixteenth century were not critiques of Christianity, but were marks of the contemporary intellectual and religious culture, while in the eighteenth century the same mockeries would have had an atheist meaning. (Febvre 1942, 150–152) Febvre argues for this by relating types of literary discourse to a general intellectual culture of the sixteenth century and showing how it differs from the cultural climate of the eighteenth century. Along the route he criticizes his predecessors, because they lacked a correct sixteenth century understanding of the sources. When one applies Hempel's framework to this case, a problem turns up. The historians whom Febvre is arguing against read the same books and letters as Febvre did. There is no clear way to understand how Febvre is verifying the determining conditions or hypothetical generalizations better than his predecessors, even though he seems to have a better understanding of the texts and the actors involved. This critique on Hempel has already been adequately pointed out by Charles Taylor. (Taylor 1980;

Taylor 1971) Taylor's position, however, leaves the question open whether the humanities' understanding can still be considered as objective. Within current analytic philosophy of science there aren't many elaborate positions that try to address these issues. This is not a very satisfactory situation.

An interesting view that shows how the use of "Verstehen" plays a non-eliminable role in securing the greater objectivity of Febvre's view (as opposed to his opponents), comes from Rudolf Carnap's "Der logische Aufbau der Welt". According to Carnap's structuralist position in the Aufbau the use of concepts embedded in a field of structured logical relations yields objectivity. Carnap believes that every scientific concept can be defined within one system of logical relations. (Carnap 1928, §15) To this end Carnap uses constitutional definitions that stipulate how to translate sentences containing new concepts into sentences containing already constituted concepts of the system. Starting from a set of basic objects (elementary experiences) and a handful of relations Carnap shows how it might be possible to constitute all the scientific concepts, including those of the humanities (Geisteswissenschaften).

In order to constitute the humanities Carnap stipulates two relations: the manifestation-relation and the documentation-relation. (Carnap 1928, §23) These relations should be capable of showing how a physical and psychological object can be understood as an expression of a cultural (geistiges) object, such as a religious or social custom. According to Carnap, the cultural objects form an autonomous level, which is irreducible to the lower levels of the physical or psychological domain. The objectivity of the cultural domain depends solely on the total structure of the system. Carnap also understands "Verstehen" as a heuristic, but it now has a more specific function that goes beyond the mere generation of hypotheses. It helps the historian in stipulating the correct relation between a physical or psychological object and a cultural object. It is a method that guides the scientist to find out how the cultural becomes alive through the merely physical or psychological objects, such as e.g. literary texts. (Carnap 1928, § 55)

One can now see how Febvre's complex argument can be captured within Carnap's framework, since it relates physical objects (books, pamphlets, letters) and psychological objects (feelings, thoughts) to cultural objects (intellectual culture, contemporary theological theory). Febvre relates the physical and psychological objects that his predecessors were studying as well, to a different cultural object, namely contemporary sixteenth century theological theory and humanist, intellectual culture. Using Carnap's framework, we can show why Febvre's better understanding of the actors and texts involved has a direct consequence for the objectivity of his claims. The latter is due not only to possible empirical verification, but also depends on the richer structural relations of the concepts used within the constitutional system.

Although Carnap himself never worked out any of the suggestions concerning the humanities from his Aufbau, I believe his position has a great potential for our understanding of the humanities. At the end of the paper I will reflect on the viability of Carnap's documentation-relation, using Febvre's work. I will show in some detail how Febvre relates the physical books, pamphlets and letters as expressions of the larger intellectual culture and how he tries to validate his relations more objectively than his predecessors.

---

*Determinism and Continuity: Irregular Vibrations of the String (1748)*

IULIA MIHAI

Ghent University

iulmihai@gmail.com

---

In this paper I give an account of the concept of irregular vibrations in order to highlight a deterministic argument in Leonhard Euler's 1748 undertake to solve the vibrating string problem. I show that in the context of this problem determinism is the foundation for Euler's legitimation of the use of arbitrary functions. I contrast Euler's use of this principle of determinism with Jean d'Alembert's appeal to continuity in solving the same problem (1747) and argue that whereas the uses of determinism and continuity are relatively similar in the structure of their respective proofs, the two principles have different foundations: the former is mechanical and the latter is metaphysical.

The thinking and conceptualization of the vibratory motion of the string in eighteenth century analytical approaches to this problem by mathematicians such as Taylor and d'Alembert are centred on the distinction between regular and irregular vibrations. This distinction is about two natures of the vibratory motion, and generally functions as a criterion for the applicability of the calculus to these motions. Mathematical models for the motion of the string are built and equations are expressed only for regular vibrations. Irregular vibrations are not suited for mathematics, but nevertheless are considered to be the vibrations with which the motion of the string starts. A stretched string pinned down at both ends is generally set in motion by plucking it; however, this action can be performed in various ways (by varying the point from which the string is plucked or the distance from which it is set loose) which create irregular motions. These motions are said to be irregular either with respect to the shape of the string, that may not be smooth, or with respect to the differences in speed between portions of the string. Most mathematicians taking up this problem put various arguments as to how irregular vibrations turn into regular vibrations after some time, and proceed to mathematization.

Euler challenges this account by arguing that the "state of vibrations" cannot change nature arbitrarily: that is, the state of the system, or the shape the string makes when a vibration is completed, depends on the preceding states, and is decisive for the next states of the system. This is a description of a conception of determinism which Euler formulates as a mechanical requirement for the system of a string in vibration.

On the basis of this principle of determinism Euler refutes the distinction between regular and irregular vibrations as nonsense: vibrations are to be considered to be of one kind only, without any of the previous restrictions that would make them "regular". Euler expresses this tenet mathematically by assigning arbitrary or "discontinuous" functions as the mathematical expressions of the shapes of the string in motion (functions

that are defined by more than one analytical expression). The scholarship has attributed Euler's trust in the use of arbitrary functions to his optimism regarding the power of mathematics to fit physics; however, in solving specific problems, such as the vibrating string problem, the use of discontinuous and arbitrary functions is an expression of principles derived from mechanics, e.g. his take on determinism.

Euler finds this principle of determinism to be in agreement with his computations in the proof, for the velocities of the points of the string in motion do not come as relevant for the equations, and only accelerations do. This is not the case for d'Alembert's proof of the same problem, since in his reasoning the velocities of the points of the string need to be preserved on a continuous set of values, and to satisfy this requirement, d'Alembert adds an extra condition that the function describing the curve that is to model the shape of the string be even. The notion of continuity at play in d'Alembert's reasoning is not mathematical (mathematical continuity was not available at that time), but metaphysical, holding that nature has no gaps and allows no leaps. This Leibnizian notion of continuity is present in d'Alembert's distinction between irregular and regular vibrations, and he appeals to it when claiming that any irregular motion of the string eventually turns into a regular motion. It is also the main ground for criticizing Euler's approach to the problem of the string.

Euler's version of determinism is not grounded in the Leibnizian notion of continuity. Moreover, determinism and continuity in Euler and d'Alembert's respective proofs play similar roles: to give a framework for the refusal or the acceptance of the irregular/regular vibrations distinction and to give theoretical support to the respective mathematical reasoning and the respective mathematical tools they use. Determinism has the advantage of minimizing the types of motions to be considered.

This paper discusses the proofs which started d'Alembert and Euler's controversy over the vibrating string by identifying the concept of determinism in Euler's work and assessing its relationship to the principle of continuity.

---

*Some Misconstructions of Similarity and a Practice-based Defence of its Value*

JULIA SANCHEZ-DORADO

UCL

juliasanchezdorado@gmail.com

---

The problem of scientific representation has become a booming topic in contemporary philosophy of science. Specifically in the past years numerous authors in the field have tried to explain the role played by similarity –between scientific models and the systems of the world they refer to– in the obtaining of fruitful scientific representations.

This paper draws on the premise that, against some of the strictures formulated on the value of similarity, it is epistemically advantageous to conserve the concept to be able to explain how scientific representations advance understanding about the world. But to succeed in the attempt, it will be indispensable, first of all, to respond to the main criticisms raised against it; and secondly, to develop a specific account of similarity that goes beyond the constrained descriptions of it often proposed in the field.

One of the best acknowledged criticisms to similarity is the already classic one raised by Nelson Goodman in the sixties. Following him, some contemporary philosophers of science have also rejected the role of similarity while generally endorsing explanations of representation based on denotation or inferentialism. These criticisms share the claim that there are significant logical incompatibilities when we try to reduce the relation of representations to a relation of similarity. In other words, they argue that while representations entail asymmetrical, non-reflexive and non-necessarily-transitive relations, similarity comprises symmetrical, reflexive and transitive relations between objects. The way I would like to respond to this kind of analysis is by recognizing that similarity does not exhaust representation indeed, precisely because similarity is not a condition of possibility of something to be a representation of something else. By contrast, similarity should be rather understood as a central feature of those practices of representing in science that generate fruitful and genuine understanding of the world.

In addition, there is another remarkable criticism to similarity that needs to be confronted—and that is key to develop my own approach in opposition to it-. I refer to the historical bias that identifies similarity with the old metaphor of the “mirror of nature” or with the ideals of “perfect imitation” or “copy”. The pertinent response to this fallacy is that similarity is not the notion that should be disregarded, but these obsolete and epistemically useless definitions of it instead.

If my proposal is right so far, then we should move away from the previously referred misconstructions of the notion of similarity. Nevertheless, it is not enough with answering to –or avoiding the criticisms made to it. It is also indispensable to provide an accurate approach that goes beyond the constrained explanations



of it already proposed in the field. Similarity is a many-sided term, it can mean different things, and accordingly diverse and usually conflicting accounts have been endorsed in contemporary philosophy of science -namely, accounts of isomorphism, homomorphism, similarity as resemblance, etc. So it is fundamental to define how exactly we are going to understand it and in which particular sense it will play a role in the construction of scientific representations.

In the second part of my paper, I will not venture to advance a completely defined new approach to similarity, but I will point out several fundamental elements that have to be present in an accurate and useful explanation of it. The main features of my approach will be: i) the location of the idea of similarity in the context of scientific practices, instead of in a binary relation of representation; that is, similarity would not refer to a set of fixed features on the objects of the representation (vehicle and target) but to a characteristic of the practices of representing; ii) the development of a more integrating approach that takes different types of similarity (i.e. isomorphism, homomorphism, appearance resemblance) as compatible in principle to each other; the predominant type of similarity would be defined by the goals of the specific research; iii) the characterization of similarity as inseparable -and compatible with- distortions of different kinds (idealizations, generalizations, simplifications, abstractions), all interlaced in the same creative practice of representing.

To be able to develop the former arguments, an important strategy of enquiry I would like to suggest is the establishment of a dialogue with the field of art, given that it has a much longer tradition discussing the problems of representation and similarity. In particular I shall refer to some stimulating debates in art that took place in the Avant-gardes period, at the beginning of the twentieth century. At first sight it can seem that allusions to similarity were radically rejected to explain the nature of artworks at that time. But quite the opposite, very interesting reflections on how to reinterpret and reconsider similarity can be found in writings on depiction and artistic practices of that period. Perhaps, artworks could not be explained exclusively in terms of "similarity of appearance" since then anymore. But far from disappearing, other kinds of similarity (perceived similarity, structural similarity, conceptual similarity) will still have presence in the theorizing of modern art. Or for example, Kandinsky's treatise *Concerning the Spiritual in Art* (1910), or Klee's and Mondrian's theoretical-applied writing on representation put the emphasis on the idea that genuine representations are characterized by the presence of singular kinds of similarity that go hand in hand with distortion and with changes of the features of the object represented. Accordingly, I will endorse that selected similarities and selected distortions together, leading to a particular goal, are essential part of our creative practices of representing.

### References

- ELGIN, Catherine. 2002. "Creation as reconfiguration: art in the advancement of science". *International Studies In The Philosophy Of Science*, Vol. 16, No. 1.
- FRENCH, Steven. 2003. "A Model-Theoretic Account of Representation (Or, I Don't Know Much about Art... but I Know It Involves Isomorphism)". *Philosophy of Science*, 70, Dec. 2003.
- GIERE, Ronald. 2004. "How Models Are Used to Represent Reality". *Philosophy of Science*, 71, Dec. 2004.

- GOODMAN, Nelson. 1972. "Seven Strictures on Similarity". In Douglas, Mary, and Hull, David (eds.). *How Classification Works*. Edimburgh: Edimburgh University Press, 1992.
- KANDINSKY, Wassily. 1911. *Über das Geistige in der Kunst. Insbesondere in der Malerei*. Translation by M. H. Sanders. *Concerning the Spiritual in Art*. The Floating Press, 2008.
- KLEE, Paul. 1920. "Schöpferische Konfession". Translation: Thames and Hudson. *Creative Confession*. Tate Publishing, 2013.
- MARCOS, Alfredo. 2011. "The Creative Discovery of Similarity". In MARCOS, Alfredo, and CASTRO, Sixto (eds.) *The Paths of Creation*. Peter Lang, 2013.
- SUÁREZ, Mauricio. 2015. "Deflationary Representation, Inference and Practice". *Studies in History and Philosophy of Science*. Vol. 49, Feb. 2015, pp. 36-47
- VAN FRAASSEN, Bas C. 2008. *Scientific Representation. Paradoxes of Perspective*. Oxford: Oxford University Press, 2008.
- WEISBERG, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. Oxford: Oxford University Press.

## Pre-events

### **Normative Social Science after the Great Recession**

Organizer: European Network for the Philosophy of the Social Sciences (ENPOSS)

Room 5E, Wednesday 09:00 – 12:00

---

*Two Ways in which Economics has been Normative*

CATHERINE HERFELD

LMU Munich

c.s.herfeld@googlemail.com

---

*Well-Being in Post-Crisis Economics. Should We Shift Attention from Preference Satisfaction Theory to Objective List Theories?*

TOMASZ KWARCINSKI

Cracow University of Economics

tomasz.kwarcinski@uek.krakow.pl

---

*On the Normative Uses of Social Science*

JOSÉ A. NOGUERA

Universitat Autònoma de Barcelona

jose.noguera@uab.cat

---

*Confirmation Meets Social Epistemology: A Theory of Inferential Judgement*

JULIAN REISS

Durham University

julian.reiss@durham.ac.uk

---

### **General Description**

One important aspect of the social sciences traditionally discussed from the philosophical or methodological points of view is the relatively clear distinction between the 'positive' and the 'normative'. Though this distinction has been subjected to strong criticisms during the last decades, mainly due to the lost of credibility

of the 'value-fact distinction' within analytic philosophy of science, it is clear that the normative concerns are paramount in many of the critics that social science in general, and economics in particular, are receiving in connection to their having failed to contribute both to the prediction of the current economic recession and to the scientifically-grounded political responses to the increasing inequality and the decreasing levels of welfare that many societies are experiencing as a consequence of the crisis. We think it is important, hence, as a contribution from philosophical analysis of the social sciences, to reconceptualise the positive-normative dimension, both with respect to the analytic instruments that the social sciences deploy to describe, understand, and evaluate normatively-laden facts, and with respect to the ways in which the search of scientific knowledge on social facts can be put to normative uses by the citizens or their representatives. Participants in this workshop range from a wide specter of fields within the philosophy of the social sciences.

## Abstracts

### 1. Catherine Herfeld: *Two Ways in which Economics has been Normative*

The normative/descriptive divide in economics has been of large concern for economists and philosophers of economics alike. For a long time, economists have tried to maintain a strict dichotomy between positive and normative. While the former is meant to tell us what is the case in science, the latter tells us what ought to be the case (e.g. Hands 2012). At the same time, a lot of confusion and disagreement exists regarding the way in which the normative and the positive dimension of economics are related. This paper is meant to further clarify the issue at hand. By tracing a set of key episodes in 20th century economic scholarship, I investigate into how the normative/descriptive distinction became interpreted and reinterpreted in economics. Much of the debate has been focused on the status and usefulness of rational choice theories, such as subjective expected utility theory. Hence, I particularly focus on the changing methodological status of rational choice theories, the ways in which economists have justified those theories, and the different interpretations of rationality underlying them. To illuminate this history, I adopt a three-fold distinction between normative, prescriptive, and descriptive models (e.g. Baron 2014): Normative models offer standards for evaluation that are defended independently of observation. They are usually considered objective and justified by philosophical and mathematical argument. Descriptive models psychologically explain human judgment and decision-making. They might for example be used to explain departures from normative models of decision-making. Finally, prescriptive models are used to foster improvement and, as such, are located in the domain of design, engineering, or practice (ibid.). Given this distinction, I argue that there are at least two ways in which economics has been normative. While both ways have frequently been closely intertwined, I urge to keep them separate; not only in economics itself, but also when tracing the history of rational choice theories, as well as when assessing current research programs such as libertarian paternalism.

**References**

- Baron, Jonathan (2014): *Heuristics and Biases*, in: Eyal, Zamir et al. (eds.): *The Oxford Handbook of Behavioral Economics and the Law*, Oxford: Oxford University Press, pp. 3-27.
- Hands, D. Wade (2012): *The Positive-Normative Dichotomy and Economics*, in *Philosophy of Economics*, ed. Uskali Mäki, North-Holland: Elsevir, pp. 219-239.

**2. Tomasz Kwarcinski: *Well-Being in Post-Crisis Economics. Should We Shift Attention from Preference Satisfaction Theory to Objective List Theories?***

Since D. Parfit's *Reasons and Persons* (1984) the philosophers have been indicating three main kinds of well-being theories. Accordingly, there are hedonistic theories, desire fulfillment theories, and objective list theories. In the mainstream of economics well-being is usually associated with preference satisfaction which is one of the variants of desire-fulfillment theory. In my paper I would like to show that reflection on the latest economic crisis can provide persuasive reasons for shifting, at least to some degree, economist's attention from preference satisfaction theory to objective list theories.

At the beginning, the main reasons for defining well-being in terms of preference satisfaction, which are widely shared by contemporary economists, will be presented. Then, I will try to show that objective list theories can cope with economist's expectations too. To do this I will refer to the research in behavioral economics (G. Akerlof, R. Shiller, R. Thaler, C. Sunstein) as well as some reflection on causes and normative implications of recent crisis (D. Ross, J. Roemer). I want to come to the conclusion that in order to prevent similar crisis in the future it is insufficient to set up more business ethics or corporate social responsibility courses in business schools but rather some in-depth reflection on such key economic concepts as well-being.

**3. José A. Noguera: *On the Normative Uses of Social Science***

The relationship between social science and normative theory (including political philosophy and ethical theory) has been most often thought only in one way: how ethical or normative values enter into the discourse and practice of social science. In this paper I will argue that it is philosophically more interesting and promising to address the opposite question: how social facts are relevant for normative theories. The issue of whether a 'value-free' social science is possible, I will claim, is not philosophically as relevant as the issue of whether a 'fact-free' normative theory is possible.

First, I will argue that debates on 'value-free' social science are quite exhausted, and that the conventional Weberian solution to this problem is essentially correct. In the context of these debates, there are two different questions that are often merged: first, is it possible to analytically distinguish between fact and value judgments?; and second, are there necessary logical relationships between some fact judgements and some normative judgments? I will defend an affirmative answer to both questions, and show how different thinkers such as Max Weber, Amartya Sen or John Rawls would subscribe this position. I would also show how the conventional anti-normative stance of empirical social scientists, as well as the mainstream

'normativist' stance of current 'social theory' or 'critical theory' (with special reference to Habermas and Putnam) are both problematic.

Second, I will claim that the role of facts in normative theory is under-theorized in at least two ways (which are different from the usual tasks of welfare economics or social choice theory): 1) assessing the factual assumptions of normative theories and political philosophy, and 2) assessing the feasibility and practical applicability of normative theories. In particular, task 1 suggests that, contrary to what mainstream 'social theory' assumes, the validity of a normative theory may depend on the validity of some social-scientific theory or evidence. Task 2 suggests that informational problems affect how to tell whether some normatively relevant state of affairs is the case or not in a particular context or situation.

Both claims lead to the conclusion that the logical relationship between positive and normative theory is not symmetrical, since the validity of a positive theory does never depend on the validity of a normative-ethical theory. So the focus should be on how facts are relevant for normative theories, and not so much on how values are relevant for positive theories (which is more a methodological and deontological issue).

#### **4. JULIAN REISS: *Confirmation Meets Social Epistemology: A Theory of Inferential Judgement***

There is no universal logic of induction (cf. Norton 2003). Inductive arguments can be more or less compelling or 'cogent' but the extent to which they are so depends on contextual factors that go well beyond what could be described as an inductive logic. William Rehg (in his book *Cogent Science in Context*) focuses in particular on three extra-logical dimensions of scientific argumentation: the dialectical, the rhetorical, and the socio-political. Dialectical elements have to do with the way in which arguments are being produced – for instance with whether or not an argument has had a sufficient chance of being challenged by a sufficiently inclusive choice of experts. Rhetorical elements are features of the presenter of the argument (e.g., 'Is she an honest scientist?') and its audience (e.g., 'Does it mainly speak to other scientists or the public at large?'). Socio-political elements are the institutional realities within which an argument is made (scarce resources may mandate a smaller degree of inclusivity than would ideally be desirable, for instance).

This paper looks at one logic of induction in particular – eliminativism – and, starting from accepting that it is indeed the case that logic and 'the facts' alone indeed underdetermine the conclusion of an inductive argument, develops a richer 'theory of inferential judgement' that includes dialectical, rhetorical, and socio-political elements in Rehg's sense. The theory is then applied to a case from contemporary social science that illustrates the importance of the extra-logical elements in making a good social-scientific argument.

**Recent Trends in the Philosophy of Social Science**

Organizer: Paul A. Roth, Philosophy of Social Science Roundtable

Room 5G, Wednesday 09:00 – 12:00

---

*Republicanism Then and Now*

JAMES BOHMAN  
Saint Louis University  
bohmanjf@slu.edu

---

*Reviving the Philosophy of History*

PAUL A. ROTH  
University of California-Santa Cruz  
paroth@ucsc.edu

---

*Normativity and Social Science*

STEPHEN TURNER  
University of South Florida  
turner@usf.edu

---

**General Description**

The papers in this workshop explore a variety of different topics areas and related core issues in the contemporary philosophy of social science, including topics in political theory (republicanism and democratic theory), the “return” of philosophy of history, and the alleged special status of normative explanations within social science. Discussions of these issues will also make clear how issues in the philosophy of social science connect to some core issues in philosophy of science, and to what extent perhaps philosophy of science ought to take more notice of topics within philosophy of social science.

**Abstracts****1. James Bohman: *Republicanism Then and Now***

Republicanism is an old, but innovative doctrine. Most of all in the work of Pettit, Skinner, and others, republicans have explored new ground, which is forward looking and oriented to the modern age and modern

understandings and a distinctive modern and "international form" of justice and democracy. These questions are now at the core rather than the margins of republicanism, precisely where fundamental dimensions of modern political life takes place, with a "broader and deepened form of freedom." By introducing a new conception of democracy, republicanism has become distinctly modern, where international bodies and states are needed to establish and entrench the basic liberties of citizens and a role for contestation the existing global framework. While I agree with some features of this view, such as its the appeal to a role for "publically delivered resources and protections" as way to entrench freedom. However, the scope of cosmopolitan liberty must, for the sake of justice, include many others.

However, modern republicanism faces a number of problems. Given the complexity of modern procedures (including Bargaining, compromise, majority rule and others, the republican polity requires a healthy forms of contestation. Here democratic legitimacy best captures the modern debates about priorities. Republicans thus need to more clearly develop a genuine role for the participation of citizens in processes of political will formation that is cosmopolitan in character. However, even with constitutional provisions in a multilevel system, if the courts take over too many functions in a multilevel system, the supremacy of the community can be undermined by a dispersed form of popular sovereignty, as Fritz Scharpf has shown. The long term solution is unstable, since it cannot make room for democratic political demands. He argues that "the extreme case" is of a polity conforming only to liberal principles which, at the same time lacks practically all republican credentials. I will show that Scharpf's argument works only on certain conditions.

## **2. Paul A. Roth: *Reviving the Philosophy of History***

A call to revive philosophy of history will, I expect, quickly prompt at least the following two questions: first, what exactly would this revival revive; and, second, why bother? Those skeptically inclined might counsel indefinite postponement, inasmuch as this subfield has remained mostly deserted since the 1970s. My primary concern will be to outline the current status of key issues raised by the first question, for the purpose of identifying those aspects within philosophy of history that both merit and demand renewed philosophical consideration. In particular, my paper reconsiders questions tied to the use of narrative as a form of explanation. Specifically, I focus on those features that make historical explanation distinctive and yet belonging on any satisfactory catalogue of explanatory strategies. I directly address an epistemic question that I take to be of central philosophical concern, viz., in what respects explanations in narrative form can be said to offer credible *justifications*. Answering this requires a turn away from narrative theory and back to neglected works by Arthur Danto and Louis Mink. For they understood in a manner now lost or forgotten the question of narrative explanation as an epistemic issue. Examination of some recent reflections on narrative explanation reveals how disconnected discussion has become from their concerns. Moreover, their work provides important and still crucial insights that can be deployed to fashion answers to philosophical



concerns about narrative explanation. I conclude with two examples of what I claim to be explanations in narrative form—Raul Hilberg’s *The Destruction of the European Jews* and Michael Friedman’s *A Parting of the Ways*—that should motivate philosophers and others to attend to narratives as a mode of explanation. These examples indicate as well how answering the first question noted at the outset also answers the second of my initial questions.

### **3. Stephen Turner: *Normativity and Social Science***

The concept of normativity poses problems for the philosophy of social science, to the extent that it is a rival to explanations in the social sciences, and to the extent that social phenomena are claimed to be immune to social science explanation and in need of some other, distinctively normative kind of explanation. The issues can be given a long history in the conflict between idealisms of various kinds and naturalistic social explanation, but there is a specific history to the current form of this problem. Wilfrid Sellars, in his early writings on the normative, was specifically concerned with the question of what “extreme anthropology” and behaviorism could explain, and with the question of what vestiges of rationalism could be preserved in face of the challenges posed by behavioral science.

The general form this problem takes is this: defenders of normativity as a special kind fact beyond the reach of social science explanation describe some realm of fact in a normative way, and contend that the apparent social science explanations fail to capture some essential feature of the fact. Social science explanations typically describe the facts in ways which eliminate this feature. Conflicts arise in a wide range of areas, between ethical theorists, for example Joseph Raz, who regard particular “good” reasons in action explanations as complete by virtue of being “good” and therefore regard social science explanations as gratuitous. The same sort of arguments may be made about explanations involving knowledge, which can be taken as a normative concept. Conflicts also take the form of claims that action, intention, and other indispensable notions are intrinsically normative. These conflicts are stubborn, and reach into many areas of philosophy. However, if they are taken seriously, social science becomes a limited enterprise, which fails to explain much of what it has traditionally attempted to explain.

**The Problem of Applicability is not a Problem**

Organizer: Philosophy of Mathematics Association (PMA)

Room 5H, Wednesday 09:00 – 12:00

---

*How to Dissolve the Problem of the Application of Mathematics*

OTÁVIO BUENO

University of Miami

otaviobueno@mac.com

---

*Mathematical Structuralism and Mathematical Applicability*

ELAINE LANDRY

University of California

emlandry@ucdavis.edu

---

*Mathematics and Inference to the Best Explanation*

ØYSTEIN LINNEBO

University of Oslo

oystein.linnebo@ifikk.uio.no

---

**General Description**

Since much of philosophy of science depends on, or at least is informed by, philosophy of mathematics it is crucial that such connections be both highlighted and valued. Again, well-witnessing the varying perspectives and differing investigations of philosophers of mathematics the topics of this session will include: 1) arguments showing that the problem of applicability is a general philosophical problem that can be faced head-on by scientific investigation and so can be dissolved as a specifically mathematical problem; 2) arguments that understanding the proper notion of a mathematical axiom allows us to give an account or how systems, both mathematical and physical, can be said to have a structure, without our having to give a metaphysical or semantic account of what structures are “made of” or “refer to”; and, 3) arguments that investigate the use of inferences to the best explanation, both in mathematical and science, to then reconsider what this might tell us about both the nature of mathematical axioms and the application of mathematics to physical theories. Overall, our aim is to show that mathematical applicability is not

mysterious, not unreasonable, and not really a problem for either the philosopher of mathematics or the philosopher of science.

## Abstracts

### 1. Otávio Bueno: *How to Dissolve the Problem of the Application of Mathematics*

I argue that there is no genuine problem of the application of mathematics. Other problems get confused with the issue of the application of mathematics, creating the impression that there is a special problem here. As it turns out, the problems in question are problems about (i) the (apparently indispensable) use of mathematical techniques in theory construction; (ii) the difficulties of determining the consequences of physical and mathematical principles in a given domain; and (iii) the mechanism of representation of empirical or mathematical phenomena. These are all well-known, and important, problems in philosophy of mathematics, philosophical logic and philosophy of science. But none of them makes the so-called problem of the application of mathematics special (see also Azzouni [2000]). After all, anyone who attempts to provide an account of science or mathematics faces these problems anyway, and these problems can be accommodated without assigning any special role to the problem of the application of mathematics.

### 2. Elaine Landry: *Mathematical Structuralism and Mathematical Applicability*

I argue that taking mathematical axioms as Hilbertian is not only better for our account of mathematical structuralism, but it yields a better account of mathematical applicability. Building on Reck's [2003] account of Dedekind, I show the sense in which, as mathematical structuralists, we ought to dispense with metaphysical/semantic demands. Moreover, I argue that it is these problematic demands that underlie both the Frege/Hilbert debate and the current debates about category-theoretic structuralism. At the heart of both debates is the metaphysical/semantic presumption that structures must be constituted from/refer to some primary system of elements, either sets or collections, platonic places or nominalist concreta, so axioms, as truth about such systems, must be prior to the notion of structure. But what we ought have learned from Dedekind [1888] and Hilbert [1899], respectively, is that we are to "entirely neglect the special character of the elements", and so axioms are but implicit definitions, and, consequently "every theory is only a scaffolding or schema of concepts together with their necessary relations... and the basic elements can be thought of in any way one likes... [A]ny theory can always be applied to infinitely many systems." The first thing to note is that no primitive system is necessary, the second is that any system, be it mathematical or physical, can be said to have a structure. Thus, applicability can be understood as just the Giere claim that a physical system has a mathematical structure, i.e., that it satisfies the axioms, in certain respects and degrees for certain physical purposes.

**3. Øystein Linnebo: *Mathematics and Inference to the Best Explanation***

The method of “inference to the best explanation” is frequently invoked, not only in the empirical sciences but also in the philosophy of mathematics. After reviewing some appeals to this method in the philosophy of mathematics, I examine the conditions under which the method is justified. This examination calls into question the potential of the method as a way of justifying new mathematical axioms.

## Name Index

---

- Abrams, Marshall 247  
 Andersen, Hanne 303  
 Andersen, Lise M. 195, 199  
 Artiga, Marc 87  
 Atanasova, Nina 160  
 Atkinson, David 10, 70  
 Autzen, Bengt 68  
 Azhar, Feraz 221  
 Bacciagaluppi, Guido 213  
 Baetu, Tudor 402  
 Bain, Jonathan 258  
 Bangu, Sorin 243  
 Bartels, Andreas 29, 116, 213  
 Barwich, Ann-Sophie 40  
 Basoukos, Antonios 106, 224  
 Baumgartner, Michael 270  
 Baysan, Umut 195, 199  
 Bedessem, Baptiste 452  
 Beebee, Helen 232, 234  
 Beirlaen, Mathieu 180  
 Bertolaso, Marta 53, 56  
 Biagioli, Francesca 224, 311  
 Bialek, Max 414  
 Bicchieri, Cristina 1  
 Boge, Florian 216, 343  
 Bohman, James 484  
 Bokulich, Alisa 46, 48  
 Boumans, Marcel 53, 54  
 Boyer-Kassem, Thomas 470  
 Bradley, Seamus 238, 240, 361  
 Brössel, Peter 10, 13  
 Brown, James Robert 142  
 Bschr, Karim 40  
 Bueno, Otávio 189, 191, 487, 488  
 Bueter, Anke 79  
 Bzovy, Justin 264, 402  
 Campaner, Raffaella 53, 56  
 Camprubi, Lino 53, 57  
 Carr, Jennifer 238, 240  
 Casini, Lorenzo 270  
 Cassini, Alejandro 73  
 Castellani, Elena 66  
 Cevolani, Gustavo 10, 14  
 Chakravartty, Anjan 3, 6  
 Cheng, Sen 201  
 Chow, Sheldon 172  
 Christian, Alexander 153  
 Colombo, Matteo 153, 155  
 Cresto, Eleonora 73  
 Cretu, Ana-Maria 417  
 Crupi, Vincenzo 10, 14  
 Dardashti, Radin 343  
 Dawid, Alexander P. 328, 330  
 Dawid, Richard 341, 390  
 de Haro, Sebastian 439  
 de Regt, Henk W. 321, 326  
 de Swart, Jaco 54, 57  
 del Corral, Miranda 73  
 Desautels, Lane 124, 300  
 Desmond, Hugh 165  
 Dewhurst, Joe 458  
 Dewulf, Fons 473  
 Dicken, Paul 372, 375  
 Diez, Jose 321  
 Dizadji-Bahmani, Foad 63  
 Dorato, Mauro 379  
 Douven, Igor 1  
 Dunlap, Lucas 441  
 Dupré, John 90  
 Duwell, Armond 275, 278  
 Dziurosz-Serafinowicz, Patryk 183  
 Egg, Matthias 382  
 Eigi, Jaana 226  
 Elder, Jamee 147, 149  
 Elliot, Colin 467  
 Fagan, Melinda 411  
 Faye, Jan 321, 325  
 Fazekas, Peter 195, 198  
 Feest, Uljana 210  
 Feintzeig, Benjamin 119  
 Feldbacher, Christian J. 189, 366  
 Fenton-Glynn, Luke 328, 330  
 Fernández Pinto, Manuela 464  
 Festa, Roberto 10, 480, 484  
 Fink, Sascha Benjamin 204  
 Fortin, Sebastian 443  
 Fraser, James 256  
 French, Steven 232, 235  
 Friederich, Simon 349, 379  
 Friend, Toby 420  
 Frigg, Roman 121  
 Frisch, Mathias 46, 49  
 Gaertner, Wulf 366, 370  
 Gärdenfors, Peter 395

## Name Index

- Gebharter, Alexander 96  
Gelfert, Axel 280, 285  
Gingras, Yves 260  
Gonzalez-Moreno, Maria 158  
Green, Sara 313  
Guay, Alexandre 232, 235  
Gurova, Lilia 85, 201  
Guttinger, Stephan 90, 93  
Gyenis, Zalan 346  
Hanson, John A. 147, 150  
Harbecke, Jens 96, 98  
Harker, David W. 3, 7  
Harr, Quinn 288  
Hartmann, Stephan 1  
Hasse, Hans 305  
Hauswald, Rico 76  
Heesen, Remco 291  
Heilmann, Conrad 363  
Held, Carsten 18, 59  
Henderson, Leah 372, 375  
Hendry, Robin 195, 198  
Herfeld, Catherine 53, 55, 480, 481  
Hillerbrand, Rafaela 446  
Hirsch Hadorn, Gertrude 300  
Hofer, Carl 59  
Hofer-Szabó, Gábor 346  
Hoff, Karla 153, 156  
Hoffmann-Kolss, Vera 96, 99, 195  
Holik, Federico 443  
Hommen, David 79  
Horsman, Clare 275, 278  
Huneman, Philippe 46, 50  
Illari, Phyllis 393  
Imbert, Cyrille 470  
Ismael, Jenann 328, 331  
Jansson, Ida L. S. 46, 51  
Kaiser, Marie I. 90, 250  
Kästner, Lena 267  
Kendon, Viv 275, 278  
Kertész, Gergely 195, 200  
Kindi, Vasso 103, 390  
Kistler, Max 423  
Knuuttila, Tarja 66  
Kornmesser, Stephan 426  
Krickel, Beate 267  
Kuhlmann, Meinard 399  
Kuipers, Theo 68, 294  
Kuorikoski, Jaakko 167, 385  
Kwarcinski, Tomasz 480, 482  
Ladyman, James 189, 192  
Lam, Vincent 352  
Landry, Elaine 487, 488  
Leegwater, Gijs 449  
Lemeire, Olivier 337  
Leonelli, Sabina 280, 284  
Leuridan, Bert 180  
Leuschner, Anna 103  
Linnebo, Øystein 487, 489  
Lisciandra, Chiara 153, 156  
Love, Alan 37, 158  
Ludwig, David 111  
Lutz, Sebastian 189, 192  
Lyon, Aidan 328, 331  
Lyons, Timothy D. 3, 8  
Macleod, Miles 280, 283  
Marcoci, Alexandru 366, 370  
Maroney, Owen 275, 278  
Martini, Carlo 464  
Masterton, George 395  
Mccoy, C. D. 219  
McKenzie, Kerry 35  
Meier, Thomas 189, 192  
Meincke, Anne Sophie 90, 93  
Menke, Cornelis 167  
Mennes, Julie 455  
Mets, Ave 109  
Mihai, Iulia 475  
Miłkowski, Marcin 96, 100  
Morreau, Michael 366, 369  
Morris, Rick 273  
Mose Bentzen, Martin 177  
Muller, F.A. 35  
Müller, Thomas 238, 296  
Nascimbene, Juan 73  
Nersessian, Nancy 280, 283  
Ney, Alyssa 232, 236  
Nguyen, James 366, 370, 429  
Nicholson, Daniel J. 90, 94  
Niiniluoto, Ilkka 245  
Noguera, José A. 480, 482  
Northcott, Robert 61  
Nounou, Antigone M. 321, 325  
Nyrup, Rune 21  
Okasha, Samir 366, 369  
Osimani, Barbara 42  
Padovani, Flavia 175  
Palermos, Orestis 280, 286  
Pashby, Thomas 190, 193  
Peijnenburg, Jeanne 70  
Pence, Charles 44  
Pero, Francesca 66  
Peters, Dean 3, 8  
Pietsch, Wolfgang 432  
Pitts, J. Brian 253  
Pontarotti, Gaëlle 163  
Portides, Demetris 435  
Poznic, Michael 408  
Pradeu, Thomas 232, 235

## Name Index

- Psillos, Stathis 372, 376  
Raftopoulos, Athanassios 461  
Ramsey, Grant 124  
Redei, Miklos 346  
Reiss, Julian 480, 483  
Retzlaff, Nina 147, 328  
Reutlinger, Alexander 46, 49, 361  
Reydon, Thomas 316  
Romano, Davide 213  
Romeijn, Jan-Willem 1  
Roth, Paul A. 484, 485  
Ruphy, Stéphanie 452  
Rus, Richard David 124  
Saatsi, Juha 46, 51, 232, 372, 377  
Saborido, Christian 158  
Sanchez-Dorado, Julia 477  
Schilgen, Hardy 355  
Schindler, Samuel 24  
Schippers, Michael 10, 15  
Scholl, Raphael 243, 311  
Schurz, Gerhard 2, 177  
Sengers, Geurt 388  
Seselja, Dunja 26  
Shagrir, Oron 96, 99  
Solomon, Adriana M. 147, 150  
Sprenger, Jan 10, 15  
Steeger, Jeremy 147, 151  
Stegenga, Jacob 144  
Stegmann, Ulrich 264, 405  
Suárez, Mauricio 328, 332  
Sus, Adán 29, 46  
Tajer, Diego 73  
Tambolo, Luca 437  
Teira, David 158  
Teller, Paul 339  
Thebault, Karim 275, 361  
Thorn, Paul 186, 288, 355  
Timpson, Chris 275, 278  
Toon, Adam 280, 282  
Trout, J. D. 18  
Tulodziecki, Dana 170  
Turner, Stephen 484, 486  
Uebel, Thomas 319  
van Bouwel, Jeroen 358  
van den Herik, Jasper 82  
van der Deijl, Willem 229  
van Eck, Dingmar 455  
van Strien, Marij 229  
Vanni, Leonardo 443  
Vassallo, Antonio 33  
Vervoort, Louis 260  
Vickers, Peter 3, 4, 9  
Votsis, Ioannis 3, 253, 308  
Weber, Marcel 2  
Wells, Aaron 148, 151  
WERNDL, CHARLOTTE 121  
Werning, Markus 201, 280  
Wheeler, Gregory 238, 240  
Williamson, Jon 238, 239  
Wilson, Alastair 31, 114  
Wintein, Stefan 363  
Woody, Andrea 114  
Wüthrich, Christian 352  
Wüthrich, Nicolas 334, 366, 370  
Ylikoski, Petri 321, 326  
Zamora Bonilla, Jesus 127, 334  
Zenker, Frank 395  
Zipoli Caiani, Silvano 207