Meta-Abduction: Inference to the probabilistically best prediction[*]

Christian J. Feldbacher-Escamilla

Spring 2022

Abstract

[51] In this paper, a taxonomy of abduction in a wide sense and an exact characterisation of probabilistic selective abduction is given. Two important features of such inferences, namely accuracy and simplicity of explanations and predictions, are described in detail. Afterwards, the epistemic merits of simplicity are discussed. They are used to justify probabilistic selective abduction in terms of an inference to the probabilistically best explanation. By help of the theory of meta-induction and its account of induction, this justification is expanded to abduction in terms of an inference to the probabilistically best prediction. This expansion of the theory of meta-induction to a theory of meta-abduction indicates that the framework fruitfully accounts for important inference methods of science.

Keywords: abduction, simplicity, meta-induction, justification of induction, cognitive costs

2.1 Introduction

The theory of meta-induction allows for justifying inductive inferences based on their past successes. Some have argued that this, finally, solves Hume's problem of induction, at least if one considers it in terms of optimisation (cf. Schurz 2019; and Feldbacher-Escamilla under revision). However, there is also another inference method which is widely-used in science and which is also in need of epistemic justification, namely the method of abduction (cf. Peirce 1994b; Harman 1965; and Lipton 2004). The question of justifying abduction will be addressed in the present essay. To be more precise, we are interested in a particular species of abductive reasoning, namely the species of an *inference to the probabilistically best explanation and prediction*. We will show that the

[*][This text is published under the following bibliographical data: Feldbacher-Escamilla, Christian J. (2022). "Meta-Abduction. Inference to the probabilistically best prediction". In: *Philosophy of Computing*. Ed. by Lundgren, Björn and Nuñez Hernandez, Nancy Abigail. Philosophical Studies Series. Cham: Springer Nature, pp. 51–72. DOI: 10.1007/978-3-030-75267-5_2. All page numbers of the published text are in square brackets. The final publication is available at https://doi.org/10.1007/978-3-030-75267-5_2. For more information about the underlying project, please have a look at http://cjf.escamilla.one.]

meta-inductive framework can be interpreted such that it allows for embedding these inferences and applying the meta-inductive optimality results. By this, meta-induction or, as we will call it: *meta-abduction*, can be not only employed to justify induction, but also to justify a species of abduction. [52]

Our investigation proceeds as follows: In section 2.2, we provide a brief outline of abductive inferences. In section 2.3, we characterise in detail the species of abductive inferences we are interested in here, namely selective abduction in form of an inference to the probabilistically best explanation or prediction, where *best* is understood in terms of accuracy and simplicity. Since the epistemic value of accurate explanations and predictions is clear, but that of simplicity is not, the latter needs to be explored further. This is done in section 2.4, where an information theoretical argument in favour of simplicity is discussed and used for fleshing out the notion of the species of abductive inference we have in mind. As we will see, the epistemic justification of accuracy and simplicity of abduction is guaranteed for an inference to the probabilistically best *explanation*. However, abduction in the sense of an inference to the probabilistically best *prediction* needs further steps of justification. As we will argue, the theory of meta-induction allows to perform these steps. For this purpose, we introduce the framework of meta-induction in section 2.5. There we also outline its vindication of induction. In section 2.6, we expand the epistemic justification of abduction as an inference to the probabilistically best *explanation* to the case of abduction as an inference to the probabilistically best *prediction*. We conclude in section 2.7.

2.2 Abduction

There are three major types of inference used in science and philosophy: deduction, induction, and abduction. Deductive inferences are truth preserving. Inductive inferences are not truth preserving, but have conclusions containing terms that occur already in the premisses. Finally, abduction is formally characterised as a non-deductive inference with a conclusion containing also terms that do not occur already in the premisses. It was Charles S. Peirce who first discussed *abductive inferences* as a topic of philosophy of science and logic in the broad sense. This formal aspect of abductive inferences was described by him as follows: "An Abduction is Originary in respect to being the only kind of argument which starts a new idea" (Peirce 1994b, p. 5.145). Clearly, this formal characterisation does not provide much of a restriction. However, it is an important feature that distinguishes it from the other forms, deduction and induction. For this reason, we want to call it a characteristic of *abduction in the wide sense*. To give an example of the different forms of reasoning, one can say that, e.g., $\{\forall x R(x)\} \vdash R(c)$ is a deductive inference, because it is truth-preserving and does not (relevantly) introduce new terms. $\{R(c_1), \dots, R(c_n)\} \mid\!\sim \forall x R(x)$ is an inductive inference because it is ampliative and also does not (relevantly) introduce new terms. And, e.g., the inference from $\exists_n^n R(x), \exists_m^m W(x)$ to $\exists_l^l H_R(x)$, $\exists_k^k H_W(x)$, $\exists_j^j M(x)$ is an abductive one, since H_R, H_W, and M are terms (representing ideas) that do not occur in the premise set ('$\exists_n^n$' stands for 'there are exactly n things such that ...'; in the example below, we will use numerical quantification with sortal variables $x_0, x_1, \ldots$ for expressing absolute frequencies within different generations). [53]

Abductive inferences play a major role in natural science as they are widely used for theory construction. Simplified speaking, by abduction one can infer from empirically accessible data theoretical hypotheses that allow for a more or less simple explanation of the data. Prominent is the example of Gregor Mendel who inferred from phenotypic properties of plants, e.g. colours red R and white W, laws of inheritance, e.g. the inheritance of recessive homozygous white H_W, dominant homozygous red H_R and mixed M traits. He hypothesised about the existence of such an—at his times observationally not accessible—theoretical structure. Based on such a structure, he could explain, e.g., the frequencies of the phenotypic properties in different generations. E.g., that 50% of the plants of the mother generation (let us say 100—Mendel cultivated and tested in the years from 1856–1863 about 5,000 pea plants) were white ($\exists_{50}^{50} x_0 W(x_0)$) and the other 50% of the plants were dominantly red ($\exists_{50}^{50} x_0 R(x_0)$), i.e. resulted themselves from breeding only red plants over several generations, was interpreted by Mendel as having only homozygous white and homozygous red plants in the mother generation: $\exists_{50}^{50} x_0 H_W(x_0)$ and $\exists_{50}^{50} x_0 H_R(x_0)$. He further hypothesised that by interbreeding the white with the red plants 100% mixed plants arose in the first daughter generation ($\exists_{100}^{100} x_1 M(x_1)$) which, by the assumption that red was dominant, allowed for explaining the fact that the first daughter generation was completely red ($\exists_{100}^{100} x_1 R(x_1)$). By the same type of reasoning he could argue that interbreeding the mixed plants results in 25% homozygous red ($\exists_{25}^{25} x_2 H_R(x_2)$), 25% homozygous white ($\exists_{25}^{25} x_2 H_W(x_2)$), and 50% mixed ($\exists_{50}^{50} x_2 M(x_2)$) traits, which predicted or explained that in the second daughter generation, due to the dominance of red, 75% of the plants were red ($\exists_{75}^{75} x_2 R(x_2)$) and 25% were white ($\exists_{25}^{25} x_2 W(x_2)$). This also explained how a phenotypical property that disappeared in one generation, namely white, could return in another generation.

More generally, we can distinguish two aspects or kinds of abductive inferences in the wide sense (cf. Douven 2018; Schurz 2008a; Aliseda 2006, p.46): those generating new hypotheses and those aiming at determining the best hypothesis from a set of available candidates. Abductive inferences of the former kind are sometimes called *creative abductions*, and those of the latter kind *selective abductions* (see, e.g. Magnani 2000; Schurz 2008a; Feldbacher-Escamilla and Gebharter 2019).

The account of Peirce is commonly subordinated to creative abduction. Since Peirce coined the term for this form of inference, we call it also *abduction in the narrow sense* here. Peirce provided the following very general inference schema for it (see Peirce 1994b, 5.189):

1. The surprising fact, E, is observed;

2. But if H were true, E would be a matter of course;

3. Hence, there is reason to suspect that H is true.

[54] Since abduction in the narrow sense is about generating new hypotheses and theories, it concerns not only the context of justification of theories, but also their context of discovery. Though most philosophers of science are quite sceptical whether a general approach towards a logic of scientific inquiry can be fruitful, there are accounts that allow for a systematic methodology of abductive hypothesis generation in terms of common cause abduction for generating hypotheses featuring new theoretical concepts on the basis of empirical phenomena (cf. Schurz 2008a; and a generalisation of the approach in Feldbacher-Escamilla and Gebharter 2019; but also Glymour 2018; for a historical case study, cf. Feldbacher-Escamilla 2019, sect.4).

Selective abduction, on the other hand, is often described as an *inference to the best explanation* and most of the philosophical literature on abduction (in the wide sense) focuses on this form of inference (see, e.g., Harman 1965; Lipton 2004; Niiniluoto 1999; Williamson 2016). It is about selecting among a set of possible explanations and predictions that one with most explanatory and predictive virtues.

The exact relation between creative abduction (or abduction in a narrow sense) and selective abduction is a matter of long-lasting philosophical dispute. A reason for this is that Peirce's criteria and their domain of application is not entirely clear and he himself underwent some development regarding his understanding of abduction (Mohammadian 2019, sect.2, distinguishes, e.g., between "the early theory from 1859 to 1890 and the later theory from 1890 to 1914"; and Hintikka 1998, p.511, speaks about "Peirce's early perspective on abduction" and "Peirce's mature view"). Also proponents of the selective abductive camp can be blamed for a similar fault. And a further reason for the long-lasting dispute is that particularly adherents of inference to the best explanation seemed to aim at endowing a long tradition by sloppily equating selective with creative abduction. So, e.g., Harman (1965, p.88) claimed that "'the inference to the best explanation' corresponds approximately to what others have called 'abduction'" and Lipton (2004, p.56) claimed that "Inference to the Best Explanation has become extremely popular in philosophical circles, discussed by many and endorsed without discussion by many more (for discussions see, e.g., Peirce 1994[)]".

In stark contrast to the "identifiers" of abduction with inference to the best explanation, there are strong "separationists" such as Hintikka (1998). According to them, abduction is not at all related to inference to the best explanation. For Hintikka (1998), to explain is to perform a deductive activity, whereas Peircean abduction is an activity of introducing new hypotheses into inquiry, and since deduction and the introduction of new hypotheses are completely different matters, also inference to the best explanation and abduction should be understood as completely different forms of inferences:

> "An explainer's job description is [...] twofold: on the one hand to find the auxiliary facts A and on the other hand to deduce the explanandum from them together with the background theory T."

(cf. p.507)

> "Since the abductive reasoner does not always have at his or her disposal explanations even of the known data, the abductive inference cannot be a step to the known data to a hypothesis or theory that best explains them." (cf. p.509)

[55] Roughly, one can also describe the "separationist's" argument as that of abduction being a matter of the context of discovery, explanation being a matter of the context of justification, and since both contexts are separated, one ought also to keep abduction separated from inference to the best explanation.

There are, however, also positions strictly within the spectrum spanning from "identifiers" to "separationists", namely those who agree that abduction and inference to the best explanation are different, but nevertheless share important components. So, e.g., Mohammadian (2019) provides such a middle ground by arguing for the formula "Abduction - The Context of Discovery + Underdetermination = Inference to the Best Explanation". Similarly, Aliseda (2006, p.46) stresses that the "process side" of abduction consists of both, construction and selection, and highlights the constructive part by help of the formula "Abductive Logic: Inference + Search Strategy" (cf. p.49).

For the purpose of our investigation, we do not need to take a stance on whether and how exactly creative abduction relates to selective abduction. Rather, we need to characterise the kind of abductive inference (in the wide sense) we are interested in and locate it within the general taxonomy. By this, we want to provide a clear picture with sharp boundaries that mark for which form of abductive inference our justification is fit and for which it is not.

Elsewhere, I have provided an approach to creative abduction in terms of common cause/common origin reasoning (Feldbacher-Escamilla and Gebharter 2019). This essay aims at approaching the problem of how to justify selective abduction. As we have outlined above, the main idea of selective abduction is to infer from some evidence that hypothesis or theory from a set of alternative hypotheses or theories, which explains the evidence best. In practice as well as in theory, pretty much everyone agrees on this. However, different accounts of selective abduction result from different ways of fleshing out what is meant with *best explanation*. It is clear that *best* should be spelled out in terms of explanatory virtues, but what exactly are these virtues? Harman (1965, p.89), when introducing the notion of an inference to the best explanation, considered that explanation to be the best, which is "simpler, which is more plausible, which explains more, which is less *ad hoc*, and so forth". In a similar line, Lipton (2004) ranks explanations as better, if they are more plausible and contrastive (in the sense of specifying and producing discriminating evidence), and among those that are equally plausible and contrastive, being better becomes a "question of comparative loveliness" (cf. p.90) which is dependent of "precision, scope, simplicity, fertility or fruitfulness, and fit with background belief" (cf. p.122).

An important problem of inference to the best explanation Harman/Lipton style is the question of its epistemic justification: What epistemic role do simplicity, scope, etc., and loveliness in general have? Lipton (2004, pp.61f) stressed the role of loveliness to be that of increasing understanding and by in a scientific realist spirit, by linking "the search for truth and the search for understanding" also considered the epistemic role of loveliness to be established. However, such a route of loveliness as a reliable indicator of truth or likeliness is not open to anti-realist philosophers of science.

[56] An extreme way out of this solution is to set as a benchmark for an inference to the best explanation something that expresses an epistemic value *per se*, i.e. truth or likelihood. Such an approach would be an account of an *inference to the most likely explanation*. However, as (Lipton 2004) has argued, focussing on likelihood alone would render the account trivial in the sense that we would simply re-state something that we wanted to explain:

> "[There are] two more versions of Inference to the Best Explanation to consider: Inference to the Likeliest Potential Explanation and Inference to the Loveliest Potential Explanation. Which should we choose? There is a natural temptation to plump for likeliness. After all, Inference to the Best Explanation is supposed to describe strong inductive arguments, and a strong inductive argument is one where the premises make the conclusion likely. But in fact this connection is too close and, as a consequence, choosing likeliness would push Inference to the Best Explanation towards triviality. We want a model of inductive inference to describe what principles we use to judge one inference more likely than another, so to say that we infer the likeliest explanation is not helpful." (Lipton 2004, p.60):

So, the problem seems to be that selective abduction Harman/Lipton style is explanatorily potent, but epistemically hard to come by. And selective abduction in the sense of an inference to the most likely explanation is epistemically justified, but explanatorily less (or even im-)potent. However, we think that there is a form of selective abduction that can be epistemically accounted for, i.e. which can be epistemically justified, and which is at the same time explanatorily potent. The idea is to spell out the epistemic value of explanatory virtues such as simplicity in probabilistic terms. Since this account aims at phrasing everything in probabilistic terms, we want to call this form of abduction (in the wide sense) an *inference to the probabilistically best explanation*. It is important to note that this form of selective abduction does take the likelihood into account, however, it is no inference to the most likely explanation, because it also takes other explanatory features into account, though these are also spelled out in probabilistic terms. An approach along these lines is, e.g., outlined in (Williamson 2016).

Due to the structural identity of explanations and predictions (we subscribe to the so-called *structural identity thesis*, cf. Hempel 1965, pp.366–376), we can distinguish this form of abductive reasoning further into inference to the probabilistically best *explanation* and inference to the probabilistically best *prediction*. Figure 1 provides an overview of our taxonomy of abduction in the wide sense and indicates where to locate the species of abduction we are concerned with in this essay.

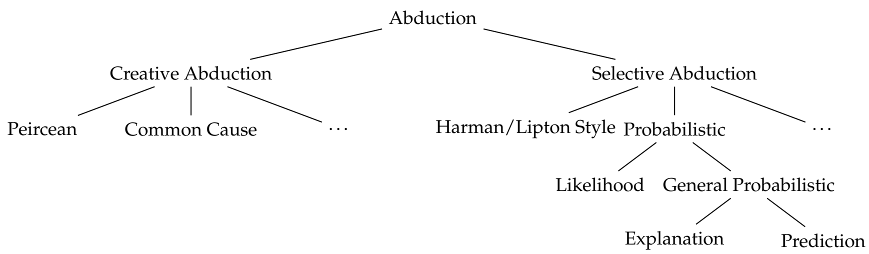

Figure 1: A taxonomy of abduction in the wide sense, where abduction in the narrow sense amounts to Peircean creative abduction; the species we are interested in is abduction in the sense of an inference to the probabilistically best explanation/prediction.

To briefly take stock, we have provided a taxonomy of abductive inferences (in the wide sense). We have seen that in the literature there are accounts that lump the different forms together ("identifiers"), there are accounts that keep them strictly separate from each other ("separationists"), and there are intermediary positions (like the approach of "Abduction - The Context of Discovery + Underdetermination = Inference to the Best Explanation" of Mohammadian 2019). However, since we focus on a particular species of abduction and aim to provide an epistemological justification for it, and only for it, we do not need to take a stance on this. Note, however, whereas Peirce' "early perspective on abduction" seems to support the identifiers' position, his "mature view" rules out such an identification, [57] at least with respect to the species of probabilistic selective abduction, because in his later view he definitely ruled out probabilistic considerations (thanks to an anonymous referee for pointing this out to me):

> "[In my early perspective on abduction] my conceptions of Abduction necessarily confused two different kinds of reasoning. When, after repeated attempts, I finally succeeded in clearing the matter up [i.e. in the mature view], the fact shone out that probability proper had nothing to do with the validity of Abduction[.]" (Peirce 1994a, 2.102)

In the next section, we provide more details on abduction in the sense of an inference to the probabilistically best explanation. In the subsequent section (2.4), we will argue for its epistemic justification. Sections 2.5 to 2.6 are devoted to the case of abduction in the sense of an inference to the probabilistically best prediction.

2.3 Inference to the Probabilistically Best Explanation

Now, what are particular characteristics of selective abductive inferences in the sense of an inference to the probabilistically best explanation? Besides the formal constraint of introducing new (theoretical) vocabulary, materially speaking characteristic for this form of abduction is its validation of explanations. So, usually an abductive inference has no single statement as a conclusion, but laws and regularities that can be used in an explanation or that even form a whole theory. In the case of Mendel's abductive inference, given the premise set outlined in our discussion of the example above, the laws and regularities of a validated explanation might consist of assumptions about the traits (the initial traits being dominant or recessive and the daughter traits being dominant, recessive, or mixed) as well as probabilistic reasoning based on assumptions about the average number of descendants of each possible pair per generation.

What are the constraints for validating such an explanation? Abduction in the sense of an *inference to the probabilistically best explanation* (see, e.g., Williamson 2016) is usually supposed to maximise the data's plausibility (in the sense of the likelihood) [58] and the hypotheses' simplicity. Typically, also further features such as scope, fruitfulness or non-ad hocness are put forward, however, for a lack of space we will focus on simplicity as a proxy for such virtues (the probabilistic reconstruction of the epistemic value of other virtues such as scope, fruitfulness, non-ad hocness etc. is in line with that of simplicity; for details cf. Forster and Sober 1994).

Regarding plausibility/likelihood, the parameter consists in the probability of the premise P (also: the *explanandum*) in the light of laws and regularities used in the explanation (the conclusion C or also: the *explanans*). The simplicity constraint is considered to be necessary in order to rule out ad hoc explanations. For, if one takes, e.g., scientists' conditional degrees of belief Pr of the explanandum P in the light of the explanans C as a measure for plausibility/likelihood: $Pr(P|C)$, which is the likelihood of P given C, then it is clear that choosing a C such that $C \vdash P$ maximises the explanandum's plausibility/likelihood in the light of the explanans. In the simplest case one might set ad hoc: $C = P$. However, what we aim at are not ad hoc explanations that might be even trivial, but universal explanations. Since ad hoc explanations usually turn out to become more and more complex with an increased number of data, abductive validation of an explanation hinges not only on $Pr(P|C)$, but also on C's simplicity. If we assume that there is some way of measuring C's complexity via a non-negative function $c(C)$, then we can characterise the validation procedure of an abductive inference as trying to maximise $Pr(P|C)$ on the one side, and minimise $c(C)$ on the other.

Several remarks are in place: First, in order to remain applicable, in this method the aim of maximising $Pr(P|C)$ and minimising $c(C)$ is to be understood not in absolute terms, but in relative ones. We might aim at $Pr(P|C) = 1$ and $c(C) = 0$, but we will almost never achieve this goal. In particular, it is presupposed that we exclude trivial abductive inferences to P itself ($Pr(P|P) = 1$ is maximal). As we have mentioned above, with an increased number of data P

to be plausibly (in the sense of likelihood) explained by C usually also the complexity of C increases (this is a case where scope opposes simplicity). And on the other hand, reducing the complexity of C usually leads to generalisations of C that are not in full agreement with P, for which reason $Pr(P|C)$ decreases. Since these two measures are intertwined in many applications, often finding a C such that $Pr(P|C) = 1$ and $c(C) = 0$ is not achievable. This was highlighted, e.g., also by Karl R. Popper, who claimed that the aim of increasing $Pr(P|C)$ "inadvertently but necessarily, implies the unacceptable rule: always use the theory which is the most *ad hoc*, i.e. which transcends the available evidence as little as possible [i.e. which sets $C = P$]" (see Popper 2002, p.61). However, what is clearly achievable is a comparative task: Assume that the only available *potential explanantia* are $C_1, \ldots, C_n$. [59] Then it holds:

$$\begin{aligned} &\text{If there is a } i \in \{1, \ldots, n\} \text{ such that for all } j \in \{1, \ldots, n\} \setminus \{i\}: \\ &\qquad c(C_i) \leq c(C_j) \;\&\; Pr(P|C_i) > Pr(P|C_j) \\ &\qquad \text{or} \\ &\qquad c(C_i) < c(C_j) \;\&\; Pr(P|C_i) \geq Pr(P|C_j), \\ &\text{then infer from } P \text{ by probabilistic selective abduction } C_i. \end{aligned} \tag{Abd}$$

(Abd) demands that in case there is an explanans C_i which plausibilises P better (in the sense of increased likelihood), but not at cost of being more complex, or which is simpler, but still not at cost of less plausibilising P than all the other possible explanantia, that in such a case C_i is to be inferred from P. If one generalises this comparative validation to the set of all potential explanantia *one has thought of* (see Williamson 2016, p.267), then one might regain an absolute phrasing of selective abductive inferences in the sense of selecting the probabilistically best explanation that is still applicable.

Second, $Pr(P|C)$ and $c(C)$ can be balanced in several ways. One might consider, e.g., a combination of the form $(1 - Pr(P|C)) \cdot c(C)$ which is the product of the inverse of the likelihood and complexity that is to be minimised, but one might also think of maximising $Pr(P|C) - c(C)$. These possibilities of balancing lead to different inferences in at least some applications. However, what is important to note is that they still satisfy (Abd). This is also the minimal constraint we want to put forward for abduction and as long as a non-deductive inference rule introducing new vocabulary satisfies it, we think it is fine to call it an 'abductive' one. In the next section, we will consider another way of balancing that also satisfies (Abd).

Third, the two parameters $c(C)$ (as a proxy for further explanatory virtues such as scope, fruitfulness, non-ad hocness etc.) and $Pr(P|C)$ are not sufficient for providing a fully adequate account of selective abduction. Usually, also the prior probabilities of the hypotheses used in an explanation are relevant. E.g., if $Pr(C_i)$ is very close to 0 and $Pr(C_j)$ is high, then one will still tend to opt for C_j instead of C_i, although $Pr(P|C_i)$ might be greater than $Pr(P|C_j)$. For simplicity reasons, we restrict the application of (Abd) to cases with close prior probabilities of the alternative hypotheses $C_1, \ldots, C_n$ (other approaches like that of Solomonoff 1964a,b, link prior probabilities to complexity c). Also, as is pointed out in (Schurz 2008a), other theoretical virtues of C as, e.g., use-novelty, unification, etc. are typically considered to be relevant for abductive inferences. Again, we restrict the intended application of (Abd) to cases where these theoretical virtues are considered to be satisfied equally well. The reason for this strong restrictions is twofold. First, some of these further parameters might be reducible to the two we are proposing. So, e.g., regarding unification and use-novelty, Forster and Sober (1994) provide reduction strategies which might be cashed out be allowing for complex P and C (conjunctions of descriptions of phenomena and hypotheses). In principle, one might even think of reducing the prior probability of a hypothesis ($Pr(C)$) as relevant parameter via inversely relating it to the complexity measure $c(C)$ (as mentioned, this would be along the lines of Solomonoff 1964a,b). [60] The second reason is that in this essay we are only interested in an exemplary application of meta-induction to abductive inferences. For this purpose, it suffices to show how the theory can be applied in case one scores not only according to accuracy, but also according to some theoretical value like simplicity/complexity. So, we should mention that our aim is to theorise about a simple model of abduction, and we want to stress that, clearly, this model has lots of limitations in comparison to the full repertoire of abduction in the wide sense.

In this model of an abductive inference there are basically two main ingredients: $Pr(P|C)$ and $c(C)$. One might wonder why $c(C)$ is relevant here. It is not hard to provide an epistemic rationale for maximising $Pr(P|C)$ in an inference of C out of P, since it is a central aim of science and philosophy to provide *good* explanations. In the traditional *deductive nomological* model of explanations, the paradigmatic case of a good explanation consists of a deductively valid argument with true laws and auxiliary assumptions as premisses and the claim to be explained as the conclusion of the argument (see Hempel 1965). Now, a high likelihood of P given C *approximates* deduction of P from C for which reason maximising $Pr(P|C)$ also serves for approximating the paradigmatic case of a good explanation (in this respect an inference to the probabilistically best explanation is along the lines of an inference to the most likely explanation). But what about $c(C)$? In how far does decreased complexity or increased simplicity serve the epistemic goals of science and philosophy? Clearly, without taking into account $c(C)$ we would lack a criterion of selecting among a multitude of potential explanations. But if it were just for reducing the number of potential explanations then also a random choice would serve the aim. According to the argument above, not considering $c(C)$ would allow for ad hoc explanations. But what is the epistemic rationale of excluding ad hoc explanations? One argument which is brought forward quite often is that ad hoc explanations *overfit* the data and so in case there is some error in the data, ad hoc explanations also fit errors. So, the argument is that since P might contain false values or statements, validating explanations that perfectly explain erroneous P are themselves defective and their explanantia C wrong. Since decreased complexity $c(C)$ allows for avoiding overfitting, less complex Cs are also less prone to fit errors. As the literature on model selection shows, this can provide a rationale for also taking into account $c(C)$ in choosing among accessible potential explanations.

To sum up, we are interested in abductive inference in the sense of an inference to the probabilistically best accessible potential explanation. 'Best' is understood as balancing two measures of an explanation of P by help of C: the likelihood $Pr(P|C)$ should be high and the complexity $c(C)$ should be low. An epistemic rationale for the former constraint results from approximating traditional models of explanation. Such a rationale for the second constraint might result from considerations of the literature on model selection showing that $c(C)$ influences C's proneness of overfitting, and by this C's proneness of also fitting errors. In the following section we are going to make this argument in favour of minimising $c(C)$ explicit. [61]

2.4 Simplicity and the Akaike Information Criterion

One way of arguing for minimising $c(C)$ is to postulate as aim of science and philosophy not only to provide true explanations, but also non-ad hoc, universal, simple explanations. In this way, already by convention about the aim of science and philosophy a demand of minimising $c(C)$ follows. However, there is also the possibility of trying to reduce the epistemic value of minimising $c(C)$ to the epistemic value of providing true explanations. The most famous approach in this direction is an application of an information theoretical framework to the problem of how to epistemically justify simplicity. The main line of argumentation is as follows (see Forster and Sober 1994): (i) Data P might be noisy and involve *error*. (ii) An accurate fit of an explanans C to the data P fits also error, it overfits the data. (iii) Whereas a less accurate fit of C to P may depart from error: Closeness to the truth is different from closeness to the data. (iv) Fact: The more parameters an explanans C has, the more prone it is to overfit P. (v) Hence: Simplicity in the sense of having less parameters may account for inaccuracy w.r.t. data P in order to achieve accuracy w.r.t. the truth. So, simplicity is instrumental for truth.

This argument is valid along general lines. But what about the truth of the premisses? Considering applications of the abductive methodology to the natural sciences, premise (i), the assumption of error in the data, is a very natural assumption. But then also premise (ii) and (iii) are straightforward: Assuming that P contains errors one only has a chance of achieving the truth by deviating from P. Intuitively and qualitatively speaking, premise (iv) is also straightforward: If an explanans is complex, it allows for fitting a simple as well as a complex explanandum. If an explanans is simple, it might fit a simple explanandum, but it cannot fit a complex one. But clearly, this is an argument too coarse-grained in order to be convincingly applied for quantitative considerations regarding minimising $c(C)$. However, there is also a much more fine-grained version of premise (iv) stemming from the literature on model selection and curve fitting—here we focus on the latter, since it became a quite influential approach to the epistemic value of simplicity (see Forster and Sober 1994). Note, however, that due to this setup we also focus on a very particular notion of *simplicity*/*complexity* only, namely parametric simplicity (for other notions of and approaches to simplicity and how to generalise the account presented here, cf. Kelly 2007).

For illustrative purposes, we will make only very simplified considerations here. The idea of model selection is that, given a data set $X = \{x_1, \ldots, x_n\}$, one is looking for a curve $F = \{f_1, \ldots, f_n, \ldots\}$ that adequately fits X. Now, it is assumed that X might contain errors, so X deviates from the truth $T = \{y_1, \ldots, y_n, \ldots\}$ (see premise (i)). [62] Clearly, the perfect choice would be $F = T$, regardless of X, but since only X is available to us, we have to base our choice of F on X. It is also clear that for any data set X with $n = |X|$ elements, choosing as F a polynomial of degree $n - 1$ allows one to perfectly fit X. One can always find parameters $a_{n-1}, \ldots, a_0$ such that for all $x \in X$ there is a $z \in \mathbb{R}$: $\langle z, x \rangle \in F$, given $F(z) = x = a_{n-1} \cdot z^{n-1} + \cdots + a_1 \cdot z^1 + a_0$. So, n parameters $(a_{n-1}, \ldots, a_0)$ allow for defining an F that perfectly fits X. If F has less parameters than n, then it is possible that there are cases where one cannot fit F perfectly to X. So, the number of parameters of F determines possibilities of perfect fitting. However, fitting X perfectly might deviate from the truth T, whereas fitting X imperfectly might allow for achieving the truth T (see premisses (ii) and (iii)).

Clearly, whether an inaccurate fit brings us closer to the truth or not depends on the exact specifics of error, namely the distance between X and T. If there were no error ($X \subseteq T$), then a more accurate and complex model would be better off than a simpler but less accurate model. However, a famous result of Hirotugu Akaike shows that on average (i.e. in estimating) simplicity matters. Forster and Sober (1994) have transformed Akaike's result to the philosophical debate of problems surrounding curve fitting. The result is the following one (see Forster and Sober 1994, p.10): The estimated predictive accuracy of the family of a model F given some data X, which is also called the *Akaike information measure according to the Akaike information criterion*, is determined by (this criterion serves only as a proxy here and one can draw the same epistemic lesson about the value of simplicity by other information criteria like the Bayes information criterion):

$$AIC(F, X) \propto \log(Pr(X|F)) - c(F) \tag{AIC}$$

Where $c(F)$ is the number of parameters of F (i.e. the degree of the polynomial F plus 1) and F is supposed to be most accurately parameterised regarding X (i.e. it is the/a polynomial of degree $c(F)$ that is closest to X in terms of the sum of squares of the differences).

Note that the idea of the Akaike information criterion is to select an F such that the estimated accuracy regarding the truth T of the family of F given some data X is maximised. Now, as (AIC) tells us, maximising $AIC(F, X)$ is twofold: It consists of maximising the log-likelihood of F given X while at the same time one needs to keep an eye on holding complexity or the number of parameters of F low.

This framework has a wide range of applications (cf. Forster and Sober 1994), however, here we are interested on cashing out (AIC) also for reducing the value of simplicity of the abductive methodology presented before—namely the value of $c(C)$ in (Abd)—to the epistemic value of gaining truths. At least at first glance it is quite straightforward to implement (AIC) into the abductive methodology outlined above: The data set X is to be identified with the premise of the abductive inference P, the explanandum. And the conclusion of the abductive inference C, the explanans, is to be identified with the curve that tries to fit X, i.e. F. The result is an Akaike-motivated characterisation of abductive reasoning: Assume that the only available *potential explanantia* are $C_1, \ldots, C_n$. Then it holds [63]:

$$\begin{aligned} &C_i \text{ can be inferred from } P \text{ by probabilistic selective abduction iff} \\ &\qquad \text{for all } j \in \{1, \ldots, n\}: \\ &\log(Pr(P|C_i)) - c(C_i) \;\geq\; \log(Pr(P|C_j)) - c(C_j) \qquad \text{(AIC-Abd)} \\ &\text{(In case more than one } C_i \text{ satisfy this constraint} \\ &\qquad \text{one might freely choose among them.)} \end{aligned}$$

According to this characterisation, every inference to an explanation is abductively permitted if it manages to get the best balance between likelihood and simplicity. Note that since $Pr(P|C) \in [0,1]$, $\log(Pr(P|C)) \in (-\infty, 0]$. Furthermore, in principle the complexity of C might have no upper limit (F might be a polynomial of arbitrarily high degree), so $c(C) \in [0, +\infty)$. So, in trying to maximise $Pr(P|C)$ and minimise $c(C)$ one also tries to maximise $\log(Pr(P|C)) - c(C)$.

More generally, (AIC-Abd) also satisfies the constraint (Abd), and since (AIC-Abd) is stronger than (Abd), it is also more specific. What is important for our argumentation is that the implementation of (AIC) in agreement with (Abd) in the criterion (AIC-Abd) seems to do the job of reducing the value of simplicity (low $c(C)$) to the epistemic value of gaining truth.

In this sense, selective abduction as an inference to the probabilistically best *explanation* is epistemically justified. However, selective abduction can not only be used as an inference to the probabilistically best *explanation*, but also as an inference to the probabilistically best *prediction*. And the question is how one can epistemically justify such an abduction, if one is missing future data for figuring out what is the best balance between accuracy and simplicity. In the following, we aim at outlining how the epistemic justification of abduction in the sense of inferring an explanation can be expanded to a justification of abduction in the sense of inferring a prediction. For this purpose, we first need to introduce the framework of the theory of meta-induction.

2.5 Meta-Induction and the Justification of Induction

Meta-induction is a theory which overcomes David Hume's problem of induction. It generalises Hans Reichenbach's *best alternatives approach* (cf. Schurz 2008b, sect.2). Reichenbach was the first to propose to consider the problem of induction not with respect to the strong requirement of proving that inductive methods are successful, but with respect to the much weaker, but epistemically still highly relevant, requirement of proving that inductive methods are the best methods accessible to us for making predictions. Since our best methods might be still predictively unsuccessful, this requirement is weaker than the one put forward by Hume. Reichenbach argued that if we cannot realise the sufficient conditions of success, we shall at least realise the necessary conditions. In order to spell out this idea, he defined a world to be *predictable*, if it is sufficiently ordered to enable us to construct a series with a limit. [64] Since the principle of induction leads to the limit, if there is a limit, convergence with the principle of induction is a necessary condition to be successful (cf. Reichenbach 1938, pp.348f; cf. also the explication in Feldbacher-Escamilla 2017, p.421).

This solution to the problem of induction or this vindication of induction is very simple, but also narrow in the sense that it holds only for ordered worlds. Since we do not know whether our world is ordered in this sense or not, the more challenging task is to vindicate induction also for the case of an unordered or—according to Reichenbach's terminology—"unpredictable" world. Exactly this is done within the approach of meta-induction (cf. Schurz 2008b): Here the idea is that if we shift our application of induction from the level of object-induction about the outcomes of an event sequence to the meta level about the success of such methods, one can generalise the Reichenbachean idea also to cases of an unordered world. Reichenbach himself suggested already such a move from the object to the meta level (1938, p.353), but his reasoning was incomplete in the sense that it remained open how an adequate prediction of the limit of a sequence on the meta level of success can be linked to an adequate prediction of the limit of a sequence of event outcomes on the object level (this criticism was put forward particularly by Skyrms 2000, p.49). It was not until 70 years after Reichenbach's proposal that this gap could be closed by the approach of meta-induction of Schurz (2008b).

In the following, we outline foundational parts of the framework of meta-induction (for a comprehensive discussion cf. Schurz 2019; and Feldbacher-Escamilla under revision). Its most central part are so-called *prediction games*. A prediction game consists of the following ingredients:

Events, Predictions, and Truth.

- Y_t^s: $Y_1^1, Y_2^1, \ldots; Y_1^2, Y_2^2, \ldots$ are infinite series of events.
- $\mathcal{Y} = \langle\langle y_1^1, y_2^1, \ldots\rangle, \langle y_1^2, y_2^2, \ldots\rangle, \ldots\rangle$ are quantified representations (within the interval $[0,1]$) of the true (or actual) outcomes (or values) of the events (event variables) to be predicted: $y_t^s \in [0,1]$.
- F_i: $F_1, \ldots, F_n$ are the prediction or forecasting methods of n predictors or forecasters.
- $\mathcal{F} = \langle\langle\langle f_{i,1}^1, f_{i,2}^1, \ldots\rangle, \langle f_{i,1}^2, f_{i,2}^2, \ldots\rangle, \ldots\rangle \ : \ 1 \leq i \leq n\rangle$ are the predictions or forecasts of the single events within the interval $[0,1]$ of the predictors or forecasters $1 \leq i \leq n$: $f_{i,t}^s \in [0,1]$

Given these ingredients, we define a prediction game by the following 4-tuple:

Prediction Game.
G is a prediction game (with the true values $\mathcal{Y}$ and the predicted values $\mathcal{F}$) about events of type(s) $I \subseteq \mathbb{N}$ iff

$$\begin{aligned} G = \langle\ & \{\langle s, t, Y_t^s \rangle : t \in \mathbb{N} \ \&\ s \in I\}, \\ & \{\langle s, t, y_t^s \rangle : t \in \mathbb{N} \ \&\ s \in I\}, \\ & \{F_i : 1 \leq i \leq n\}, \\ & \{\langle s, i, t, f_{i,t}^s \rangle : 1 \leq i \leq n \ \&\ t \in \mathbb{N} \ \&\ s \in I\}\ \rangle \end{aligned}$$

[65] I is a set of indices of the event types in question ($I \subseteq \mathbb{N}$). E.g., a prediction game with $I = \mathbb{N}$ amounts to a task of predicting everything (assuming that the set of all properties is countably infinite); $I = \{3, 4, 5\}$ might filter out, e.g., a prediction game on weather where events of type 3 might be about sunny, events of type 4 might be about rainy, and events of type 5 might be about cloudy days; one might put forward probabilistic constraints for connecting the predicted values $f_{i,t}^3, f_{i,t}^4, f_{i,t}^5$ as well as for the outcomes y_t^3, y_t^4, y_t^5 such that they are non-negative and sum up to 1 and for the ys one typically assumes that they are $\in \{0, 1\}$ (for all $i, t \in \mathbb{N}$); given these constraints, one can interpret such a prediction game also as a probabilistic one. On the other hand, setting $I = \{3\}$ filters out a simple prediction game on all events of type Y^3 (whether it is sunny or not or to which degree it is sunny etc.). If I is a singleton and the specific event type is irrelevant, then one can just omit super-indices. One can speak then also about a 'prediction game' simpliciter (this notion of a prediction game allows also for probabilistic settings and generalises that of Schurz 2008b). In most parts of what follows we only have in mind such simple prediction games with $|I| = 1$.

Relevant for the evaluation of predictions within a prediction game are especially the values of y and f_i (see figure 2): The closer $f_{i,t}$ is to y_t, the better the prediction of i. And the closer the $f_{i,t}$'s are to the respective y_t's, the better i is a predictor in general.

y_1	y_2	y_3	y_4	y_5	$\dots$
$f_{1,1}$	$f_{1,2}$	$f_{1,3}$	$f_{1,4}$	$f_{1,5}$	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$f_{n,1}$	$f_{n,2}$	$f_{n,3}$	$f_{n,4}$	$f_{n,5}$	$\dots$

Figure 2: Prediction game with event outcomes y and predictions $f_i, \dots, f_n$ of n predictors

How good a prediction $f_{i,t}$ is with respect to the true outcome y_t is measured by help of a loss function ℓ, which is a monotonically increasing function with the arguments $f_{i,t}$ and y_t that is convex. The so-called *natural loss* is, e.g., the absolute difference between $f_{i,t}$ and y_t. The *quadratic loss* consists in the squared difference of these two values, etc. Based on such a loss function, we can define the success of a prediction method f_i up to a specific event or round t as the average of the inverse of the loss $(1 - \ell)$ up to round t:

Success.

$$\mathit{succ}(F_i, t) = \frac{\sum_{u=1}^{t} 1 - \ell(f_{i,u}, y_u)}{t}$$

Given a prediction game G with a prediction method F_i and the predictions f_i, $\mathit{succ}(F_i, t)$ expresses how well F_i scored on average until round t in predicting $f_{i,1}, \ldots, f_{i,t}$ given the true event outcomes $y_1, \ldots, y_t$. The higher $\mathit{succ}(F_i, t)$, the better the prediction method is. [66]

Now, in vindicating induction, one aims to show that employing induction is the best thing one can do, because by doing so one scores, at least in the long run, best among all the competitors of such a prediction game. The theory of meta-induction fleshes out this idea by defining a meta-prediction method, which takes for each prediction round t of a prediction game G as input the success of all the predictors $F_1, \ldots, F_n$ up to round $t - 1$, and makes a prediction for round t by help of success-based weighting of the predictions $f_{1,t}, \ldots, f_{n,t}$. Such a meta-inductive predictor F_{mi} predicts $f_{mi,t}$ for each round t. And the weights used for this predictions might be, e.g., a normalisation of the difference of the successes in the exponent (many more other meta-inductive prediction methods are studied in Schurz 2019):

Meta-Inductive Weights.

$$w(F_i, t) = \frac{e^{\sqrt{8 \cdot ln(n) \cdot (t-1)} \cdot (\mathit{succ}(F_i, t-1) - \mathit{succ}(F_{mi}, t-1))}}{\sum_{j=1}^{n} e^{\sqrt{8 \cdot ln(n) \cdot (t-1)} \cdot (\mathit{succ}(F_j, t-1) - \mathit{succ}(F_{mi}, t-1))}}$$

Given these weights, the meta-inductive prediction consists simply in linear weighting the predictions $f_{1,t}, \ldots, f_{n,t}$ for round t based on the weights calculated for the respective prediction methods' successes until round $t - 1$:

Meta-Inductive Predictions.

$$f_{mi,t} = \sum_{i=1}^{n} w(F_i, t) \cdot f_{i,t}$$

Since the prediction of F_{mi} for round t is based on the successes of the prediction methods until round $t - 1$, the meta-inductive prediction can be considered to be an *inductive* prediction method. It inductively infers future success from past success. Also, this way of employing induction is not about the values of the event outcomes, but about the successes of other methods, it is a *meta*-method. Hence, it is a *meta-inductive* prediction method.

Now, we cannot go into many details regarding the performance of meta-induction here, but based on theorems of the machine learning literature, one can show that the difference between the successes of the predictors $F_1, \ldots, F_n$ of G and that of the meta-inductive predictor F_{mi} is bounded as follows (cf. Cesa-Bianchi and Lugosi 2006, sect.2.1f; Schurz 2008b, sect.7; and the proof in the appendix of Feldbacher-Escamilla 2020; and a slight generalisation in Feldbacher-Escamilla and Schurz 2020):

Theorem about Meta-Inductive Success Bounds.
Given the underlying loss function ℓ (which is used for defining succ) is convex, it holds:

$$\mathit{succ}(F_i, t) - \mathit{succ}(F_{mi}, t) \leq \sqrt{const \cdot ln(n)/t} \quad (\forall i \in \{1, \ldots, n\})$$

This bound (*const* is just a particular constant number) is of particular interest, because it shows that the difference between the success rates grows at most sublinearly with t for the meta-inductivist. This means that in the long run, i.e. in the limit, the meta-inductive method cannot be outperformed by any other predictor of the prediction game G. [67] In other words, meta-induction is long run optimal in comparison to any prediction method of G:

Meta-Inductive Optimality.

$$\lim_{t \to \infty} \mathit{succ}(F_i, t) - \mathit{succ}(F_{mi}, t) \leq 0 \quad (\forall i \in \{1, \ldots, n\})$$

It is important to note that this is an analytic result about how we defined F_{mi} and that this holds for any sequence $y_1, y_2, \ldots$ of event outcomes. So, even if the event series in question is not predictable in the sense of Reichenbach, meta-inductive predictions are still long run optimal and in this sense performing meta-inductive predictions is the best one can do. So, we have some form of a priori or deductive justification of meta-induction.

Now, given the meta-inductive optimality result, one can provide also a justification for classical (object-)induction. The general form of reasoning is as follows: As the above-cited result shows, meta-induction for a weighted selection of predictions of any accessible method is proven to be optimal in the long run. If we take for granted the past success of classical inductive methods—something that is clearly a contingent and a posteriori matter of fact and also not scrutinised by Hume himself (cf. Howson 2003, p.4)—it follows that a meta-inductive selection of such classical inductive methods for predictions of future events is guaranteed to be long run optimal. This holds at least as long as there are no alternative methods in G that outperform classical inductive methods. And since, at least up to now, classical inductive methods of science outperformed any other prediction method, our use of them is a posteriori justified.

Having outlined the meta-inductive framework and how it is employed to justify induction, we want to show how this approach can be also employed to justify abductive reasoning in the sense of an inference to the probabilistically best prediction in next section.

2.6 Meta-Abduction: Inference to the Probabilistically Best Prediction

In this section, we apply the theory of meta-induction in order to show how not only induction, but also a species of abduction can be justified. For this purpose, we need to show how abduction in the sense of an inference to the probabilistically best prediction can be embedded into the meta-inductive framework—resulting in a theory of meta-abduction—and then outline how the optimality result of meta-induction can be also employed to argue for the optimality of this kind of abduction.

Given the epistemic relevance of simplicity as outlined above, how should we select among hypotheses, explanations, theories? According to probabilistic selective abduction (AIC-Abd), we should try to maximise the information theoretical balance between accuracy ($Pr(P|C)$) and simplicity ($c(C)$). [68] By choosing that hypothesis, explanation or theory which has the *best* balance, we will be closest to the truth, which might be different from being closest to the data P (see Forster and Sober 1994, p.6). So, given the epistemic aim of *being close to the truth*, (AIC-Abd) seems to be an optimal means to achieve this end. However, this is with respect to explanation. What about predictions? What about choosing the *best* balanced hypothesis or theory for prediction?

The theory of meta-induction can be applied for optimising predictions in any respect, as long as the formal conditions of the framework are satisfied. In our application to Hume's problem of induction we interpreted the framework plainly epistemically: Given a prediction game G with $\mathcal{Y}$ and $\mathcal{F}$, we interpreted $\mathcal{Y}$ as the truth and $\mathcal{F}$ as prediction methods or hypotheses about the truth. However, we can also take in a more pragmatic standpoint and interpret $\mathcal{Y}$ as past, present, and future data, and $\mathcal{F}$ as prediction methods or hypotheses about which data will be gathered in the future. Since data typically contains error and noise, it easily falls apart from the truth, hence this interpretation does not coincide with the former. And in this sense it seems to be perfectly fine that also the criteria for success fall apart: Epistemically speaking, we still aim at predictions that are as closest to the truth as possible. However, given our noisy data, we know that we need to aim at predictions that are best balanced between accuracy (fitting) and complexity (overfitting). Success consists not in minimising the distance from the data, but making a prediction which is *best* balanced between these two parameters. So, in order to achieve this goal in the long run, the idea is to use a normalisation of (AIC). If r is the highest polynomial we are going to consider and Pr is ϵ-regular (i.e. only $Pr(\perp) = 0$ and all other probabilities are $> \epsilon > 0$), then $AIC(C, P) \in [\log(\epsilon) - r, -r]$. Hence, we can normalise $AIC(C, P)$ to $[0, 1]$ by taking

$$\frac{AIC(C, P) - (\log(\epsilon) - r)}{-\log(\epsilon)}$$

which is in $[0,1]$.

Now, let us consider a prediction game G with $\mathcal{Y}$, $\mathcal{F}$. Let G be about predicting the *best* balancing for making explanations or predictions. Assume that $\mathcal{Y}$ is a series of objectively *best* balancing. This will still fall apart from providing a most accurate prediction, i.e. a true prediction, because the degree of an extension of a polynomial predicting up to round $t-1$ might be increased by 1 if one predicts for round t the true value with probability 1, whereas it might not be increased at all by predicting the true value with close to 1 probability and hence deviation from the truth might have a higher AIC (the AIC is higher, if the deviation allows for no change in the degree of the polynomial and the basis of the logarithm in equation (AIC) is $> 1/(1 - Pr(P|C))$). Given such an "objective" best balancing, we can interpret $\mathcal{F}$ as a set of theories or hypotheses which provide predictions of what will be the best balance once new data enters the game (i.e. once one moves forward to the next round). We take the predictions in $\mathcal{F}$ to be the actual AICs of the same methods in predicting some event in another prediction game, let us say G'. [69] So, if for $f'_i \in \mathcal{F}'$, $AIC(f'_i, \{Y_1, \ldots, Y_{t-1}\}) = a$, then the respective $f_i \in \mathcal{F}$ predicts for round t as best balance the normalisation of a, i.e. $\frac{a-(\log(\epsilon)-r)}{-\log(\epsilon)}$. G is, so to say, a meta game where any prediction method of the ordinary game G' predicts that it has the right balance for future predictions. In other words, playing G' and making predictions comes with the commitment of claiming also that one's prediction is right in the sense of best balanced—that is a claim in G.

Now, again by success-based mixing of the forecasts about the best balance to be expected, a meta-inductive learner achieves long run optimality in predicting the best balance in G. If we assume, e.g., that in science creative abductive methods as hinted at in the introduction with high unificatory power had the best balance in the past, then using such creative abduction for inferring theoretical frameworks is epistemically justified, since using them is, at the current state of science, the best thing to do: Following the meta-inductive (or better: meta-abductive) selection allows for optimality in predicting the best balance in G (and actually having the best balance in one's events predictions in G').

Note that given this assumption, anti-abduction fails—which would be to get the worst possible balance of accuracy and simplicity—to be justified: Disunification and theoretically laden hypothesis invention fared suboptimal in past (in G') and hence its predictions of the best balance in G were also wrong. Hence, meta-inductive selection ignores these methods and this is the best thing to do, at least given their past performance. Given this assumption, anti-abduction is by far no optimal means to achieve the epistemic end of being best balanced in G.

Another note is in place: In the argument above, we implicitly made the assumption that success in the meta game G and success in G' are synchronous: Whenever one was relatively successful in choosing the right balance for theory and hypothesis invention (G), one was also likewise successful in predicting events (G'). The problem with this assumption is that in principle nothing hinders an adversary in letting fall things apart from each other. So, in principle one could allow for a method performing good in G', but at the same time failing in G. However, we can argue for our assumption of a parallel development of G and G' by hinting at or assuming a past correlation between abductive theory building and predictive success, and then employ induction as we justified by help of meta-induction in the preceding section. By this we can inductively transfer this correlation of the past also to the future. And we are epistemically justified in doing so.

Finally, we should also mention that such an approach of meta-abduction as outlined here can be considered more generally as introducing cognitive costs in prediction games. General features of introducing cognitive costs are studied, e.g., in (Schurz 2019, sect.7.6). [70]

2.7 Conclusion

To briefly sum up: In this essay we have provided a taxonomy of abductive inferences (in the wide sense) and an exact characterisation of a species of selective abduction, which aims at inferring the probabilistically best hypothesis, explanation or theory on the basis of data; the two main relevant factors in doing so are likelihood of the data given the inferred hypotheses and simplicity of the hypotheses, where simplicity was used as a proxy for many other explanatory virtues such as scope, non-ad hocness, unification; we have provided an argument for the epistemic value of simplicity and have shown how inferences of *explanations* based on these factors are justified. Finally, we have introduced the framework of meta-induction, outlined its justification of induction, and have also sketched how abductive inferences regarding *predictions* can be justified by employing the framework of meta-induction.

Acknowledgements

This research was supported by the *German Research Foundation* (DFG), research unit *Inductive Metaphysics* FOR 2495. I would like to thank two anonymous referees for very helpful comments on an earlier version of this paper. For valuable discussions of this topic, I would like to thank Atocha Aliseda and the audience of *IACAP 2019* (annual meeting of the *International Association for Computing and Philosophy*, June 5-7, 2019, Mexico City). Thanks a lot also to Nancy Abigail Nuñez Hernandez and Björn Lundgren for organising this volume. [71]

References

Aliseda, Atocha (2006). *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Dordrecht: Springer. DOI: 10.1007/1-4020-3907-7.

Cesa-Bianchi, Nicolo and Lugosi, Gabor (2006). *Prediction, Learning, and Games*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511546921.

Douven, Igor (2018). "Abduction". In: *The Stanford Encyclopedia of Philosophy (Summer 2018 Edition)*. Ed. by Zalta, Edward N.

Feldbacher-Escamilla, Christian J. (2017). "Optimisation in a Synchronised Prediction Setting". In: *Journal for General Philosophy of Science* 48.3, pp. 419–437. DOI: 10.1007/s10838-017-9379-7.

— (2019). "Newtons Methodologie: Eine Kritik an Duhem, Feyerabend und Lakatos". In: *Archiv für Geschichte der Philosophie* 101.4, pp. 584–615. DOI: 10.1515/agph-2019-4004.

— (2020). "An Optimality-Argument for Equal Weighting". In: *Synthese* 197.4, pp. 1543–1563. DOI: 10.1007/s11229-018-02028-1.

— (under revision). *Epistemic Engineering. Uncovering the Logic of Deceivability and Meta-Induction*. book manuscript.

Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2019-01). "Modeling Creative Abduction Bayesian Style". In: *European Journal for Philosophy of Science* 9.1, pp. 1–15. DOI: 10.1007/s13194-018-0234-4.

Feldbacher-Escamilla, Christian J. and Schurz, Gerhard (2020). "Optimal Probability Aggregation Based on Generalized Brier Scoring". In: *Annals of Mathematics and Artificial Intelligence* 88.7, pp. 717–734. DOI: 10.1007/s10472-019-09648-4.

Forster, Malcolm R. and Sober, Elliott (1994). "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions". In: *The British Journal for the Philosophy of Science* 45.1, pp. 1–35. DOI: 10.1093/bjps/45.1.1.

Glymour, Clark (2018). "Creative Abduction, Factor Analysis, and the Causes of Liberal Democracy". In: *Kriterion – Journal of Philosophy*. URL: http://www.kriterion-journal-of-philosophy.org/kriterion/issues/Permanent/Kriterion-glymour-01.pdf.

Harman, Gilbert H. (1965). "The Inference to the Best Explanation". In: *The Philosophical Review* 74.1, pp. 88–95. DOI: 10.2307/2183532.

Hempel, Carl G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press.

Hintikka, Jaakko (1998). "What Is Abduction? The Fundamental Problem of Contemporary Epistemology". In: *Transactions of the Charles S. Peirce Society* 34.3, pp. 503–533. DOI: 10.2307/40320712.

Howson, Colin (2003). *Hume's Problem*. Oxford: Clarendon Press.

Hume, David, ed. (1748/2007). *An Enquiry Concerning Human Understanding*. Oxford World's Classics. (ed. by Millican, Peter). Oxford University Press.

Kelly, Kevin T. (2007). "How Simplicity Helps You Find the Truth without Pointing at it". In: *Induction, Algorithmic Learning Theory, and Philosophy*. Ed. by Friend, Michèle, Goethe, Norma B., and Harizanov, Valentina S. Dordrecht: Springer, pp. 111–143. DOI: 10.1007/978-1-4020-6127-1_4.

Lipton, Peter (2004). *Inference to the Best Explanation*. 2nd Edition. London: Routledge.

Magnani, Lorenzo (2000). *Abduction, Reason and Science. Processes of Discovery and Explanation*. New York: Kluwer Academic Publishers.

Mohammadian, Mousa (2019). "Abduction - The Context of Discovery + Underdetermination = Inference to the Best Explanation". In: *Synthese*. DOI: 10.1007/s11229-019-02337-z.

Niiniluoto, Ilkka (1999). "Defending Abduction". In: *Philosophy of Science* 66, S436–S451. DOI: 10.1086/392744.

Hartshorne, Charles, Weiss, Paul, and Burks, Arthur W., eds. (1994a). *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

— (1994b). "Pragmatism and Abduction". In: *Collected Papers of Charles Sanders Peirce*. Ed. by Hartshorne, Charles, Weiss, Paul, and Burks, Arthur W. Cambridge, MA: Harvard University Press.

Popper, Karl R. (2002). *Conjectures and Refutation*. New York: Basic Books.

Reichenbach, Hans (1938). *Experience and Prediction. An Analysis of the Foundations and the Structure of Knowledge*. Chicago: University of Chicago Press.

Schurz, Gerhard (2008a). "Patterns of Abduction". English. In: *Synthese* 164.2, pp. 201–234. DOI: 10.1007/s11229-007-9223-4.

— (2008b). "The Meta-Inductivist's Winning Strategy in the Prediction Game: A New Approach to Hume's Problem". In: *Philosophy of Science* 75.3, pp. 278–305. DOI: 10.1086/592550.

— (2019). *Hume's Problem Solved. The Optimality of Meta-Induction*. Cambridge, Massachusetts: The MIT Press.

Skyrms, Brian (2000). *Choice and Chance. An Introduction to Inductive Logic*. Fourth Edition. Stamford: Wadsworth.

Solomonoff, Ray J. (1964a). "A Formal Theory of Inductive Inference. Part I". In: *Information and Control* 7.1, pp. 1–22. DOI: 10.1016/S0019-9958(64)90223-2.

— (1964b). "A Formal Theory of Inductive Inference. Part II". In: *Information and Control* 7.2, pp. 224–254. DOI: 10.1016/S0019-9958(64)90131-7.

Williamson, Timothy (2016). "Abductive Philosophy". In: *The Philosophical Forum* 47.3-4, pp. 263–280. DOI: 10.1111/phil.12122.