Modeling Creative Abduction Bayes Net Style

Christian J. Feldbacher-Escamilla Alexander Gebharter

Autumn 2019

Project Information

Publication(s):

- Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2019c-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.

Talk(s):

- Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2019a-09-11/2019-09-14). *Modeling Creative Abduction Bayes Net Style*. Conference. Presentation (contributed). EPSA19. Conference of the European Philosophy of Science Association (EPSA). University of Geneva: EPSA.
- Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2019b-08-05/2019-08-10). *Modeling Creative Abduction Bayes Net Style*. Conference. Presentation (contributed). 16th Congress of Logic, Methodology and Philosophy of Science (CLMPST16). University of Prague: Division of Logic, Methodology, Philosophy of Science, and Technology (DLMPST).

Project Information

Talk(s) Continued:

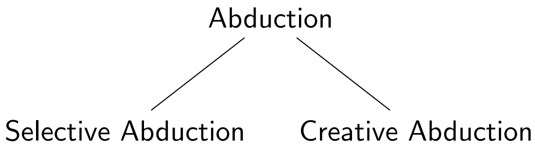
- Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2018b-11-01/2018-11-04). *Modeling Creative Abduction Bayesian Style*. Conference. Presentation (contributed). PSA2018. Seattle: Philosophy of Science Association.
- Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2018a-10-18/2018-10-20). *Abductive Concept Formation Bayesian Style*. Conference. Presentation (contributed). Concept Formation in the Natural and the Social Sciences. University of Zurich: Institute of Philosophy.
- Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2018c-06-11/2018-06-13). *Modeling Creative Abduction Bayesian Style*. Conference. Presentation (contributed). Models of Explanation: MuST 2018. University of Turin: Center for Logic, Language, and Cognition (LLC).

Project(s):

- DFG funded research unit *Inductive Metaphysics* (FOR 2495); subprojects: *Creative Abductive Inference and its Role for Inductive Metaphysics in Comparison to Other Metaphysical Methods* and *Statistical Causation Intervention and Freedom*.

Introduction I

Abduction is an important inference method in the sciences.



- **Selective Abduction (IBE):** aims at determining the best hypothesis from a set of available candidates (Lipton 2004; Niiniluoto 1999)
- **Creative Abduction:** inference method for generating hypotheses featuring new theoretical concepts on the basis of empirical phenomena (Douven 2018; Schurz 2008)

Whereas selective abduction is a commonly accepted inference method, creative abduction (as some kind of *logic of scientific inquiry*) is still controversial.

Introduction II

Schurz (2008) proposed to justify a certain kind of creative abduction on the basis of Reichenbach's (1971) principle of the common cause.

In this talk, we take up this proposal and model cases of successful creative abduction within a Bayes net framework.

This allows us to

- specify general necessary conditions for successful creative abduction,
- describe its unificatory power in a more fine-grained way, and
- to shed new light on several other issues within philosophy of science.

Contents

Abduction, Unification, and Common Causes

Creative Abduction & Unification I

In this talk we concentrate on the approach of Schurz (2008).

Basic idea:

- theoretical concepts and empirical phenomena are intimately connected via dispositions (Carnap 1936, 1937)
- basis for creative abduction: empirically correlated dispositions
- justification via the principle of the common cause (CCP)

Dispositions are understood as test-reaction pairs.

Example: x is soluable in water if x dissolves at some time t, when put into water at t.

Creative Abduction & Unification II

General form of dispositional statement:

$$\forall t(\underbrace{T(x,t)}_{\textit{test}} \rightarrow (\underbrace{D(x)}_{\textit{theoretical}} \leftrightarrow \underbrace{R(x,t)}_{\textit{reaction}})) \tag{1}$$

Empirical content of this statement (uniformity assumption):

$$\underbrace{\exists t(T(x,t) \wedge R(x,t))}_{\text{once positively tested}} \rightarrow \underbrace{\forall t(T(x,t) \rightarrow R(x,t))}_{\text{always positively tested}} \tag{2}$$

Not much is gained till now: For every T-R-regularity a distinct disposition D is introduced. $\Rightarrow$ no unification

Creative Abduction & Unification III

Things become more interesting if we focus on regularities among several dispositions $D_1, \dots, D_n$. For example, a strict correlation:

$$D_i(x) \leftrightarrow D_{i+1}(x) \text{ for all } 1 \leq i < n \tag{3}$$

This amounts to assuming that the following statements (crossed uniformity assumptions) have been empirically established (for all $1 \leq i, j \leq n$):

$$\underbrace{\exists t(T_i(x,t) \wedge R_i(x,t))}_{\text{once positively tested with } T_i/R_i} \rightarrow \underbrace{\forall t(T_j(x,t) \rightarrow R_j(x,t))}_{\text{always positively tested with } T_j/R_j} \tag{4}$$

Now the introduction of a higher-level disposition $\mathcal{D}$ allows for unification:

$$\forall t(T_i(x,t) \rightarrow (\mathcal{D}(x) \leftrightarrow R_i(x,t))) \text{ for all } 1 \leq i \leq n \tag{5}$$

n^2 crossed uniformity assumptions $\Rightarrow$ n dispositional statements (5)

Creative Abduction & Common Causes I

Once the crossed uniformity assumptions are established, then disposition $\mathcal{D}$ (as characterized in (5)) is abductively inferred.

In a nutshell:

- Given the uniformity assumptions, lower-level dispositions $D_1, \ldots, D_n$ are introduced. $\Rightarrow$ no unification
- An empirical correlation among $D_1, \ldots, D_n$ is observed (crossed uniformity assumptions).
- **Abductive step**: The higher-level disposition $\mathcal{D}$ is introduced. $\Rightarrow$ unification

Justification of the abductive step via:

(CCP) If two properties D_i and D_j are correlated and neither D_i causes D_j nor D_j causes D_i, then D_i and D_j are effects of a common cause $\mathcal{D}$.

Creative Abduction & Common Causes II

By justifying the abductive step via (CCP) one commits oneself to a realist interpretation of theoretical concepts $(D_1, \ldots, D_n, \mathcal{D})$.

(CCP), in turn, can be justified by subscribing to a realist interpretation of the causal Bayes net framework (Gebharter 2017; Schurz 2016; Schurz and Gebharter 2016).

It is a consequence of the Markov condition, when causally interpreted.

This suggests a causal Bayes net treatment of creative abductive inference (Glymour 2018).

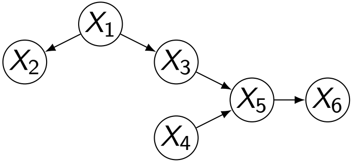
Though the realist interpretation motivates such a framing, a Bayes net treatment also allows for an instrumentalist interpretation.

Modeling Creative Abduction Bayesian Style

Bayesian Networks I

A Bayesian network is a triple $\langle \mathbf{V}, \mathbf{E}, P \rangle$, such that ...

- $\mathbf{V}$ is a set of variables $X_1, \ldots, X_n$.
- $\mathbf{E}$ is a binary relation on $\mathbf{V}$ ($X_i \longrightarrow X_j$).
- P is a probability distribution over $\mathbf{V}$.



Par(X_i) ... the set of X_i's parents
Des(X_i) ... the set of X_i's descendants

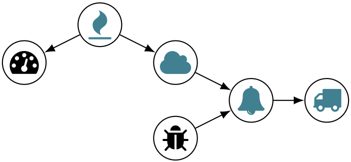
Bayesian Networks II

Definition (Markov condition)

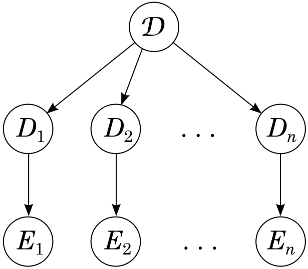
$\langle \mathbf{V}, \mathbf{E}, P \rangle$ satisfies the Markov condition iff every $X \in \mathbf{V}$ is probabilistically independent of its non-descendants conditional on its parents (Pearl 2000, p.16).

Markov factorization:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \mathbf{Par}(X_i))$$

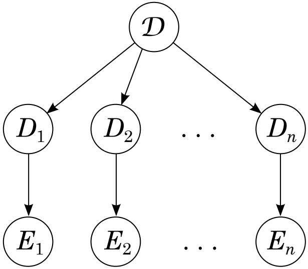


Creative Abduction Bayesian Style I



- $D_1, \dots, D_n$: correlated lower-level dispositions
- $\mathcal{D}$: abduced higher-level disposition
- $E_1, \dots, E_n$: evidence, generalization of test-reaction instances
- $D_i \longrightarrow E_i$: like evidence and hypotheses in Bayesian networks
- $\mathcal{D} \longrightarrow D_i$: justification via (CCP) [with possible intermediate causes]

Creative Abduction Bayesian Style II



Necessary conditions for successful creative abduction:

- N1 $\mathcal{D}$ is not extreme, i.e., $0 < P(\mathcal{D}) < 1$.
- N2 Each D_i depends positively on $\mathcal{D}$, i.e., $P(D_i|\mathcal{D}) > P(D_i)$.
- N3 Each E_i depends positively on D_i, i.e., $P(E_i|D_i) > P(E_i)$.

$\Rightarrow$ probability flow between dispositions and E_i—see (Dardashti, Thébault, and Winsberg 2015)

Unificatory Power I

Recall that in the original approach the crossed uniformity assumptions (4) are to be explained (via introducing $\mathcal{D}$):

$$\exists t(T_i(x,t) \wedge R_i(x,t)) \rightarrow \forall t(T_j(x,t) \rightarrow R_j(x,t))$$

In the Bayesian setting, this amounts to explaining the following empirical correlation statements (via introducing $\mathcal{D}$):

$$P(E_i|E_j) > P(E_i), \text{ where } 1 \leq i \neq j \leq n \tag{6}$$

Numbers of statements to be unified:

- Strict case: n^2 crossed uniformity assumptions
- Probabilistic case: $\binom{n}{2}$ empirical correlation statements

Numbers of unifying statements:

- Strict case: n higher-level dispositional statements (5)
- Probabilistic case: $2n+1$ probabilistic statements (N1–N3)

Unificatory Power II

How to compare the unificatory power of the two approaches?

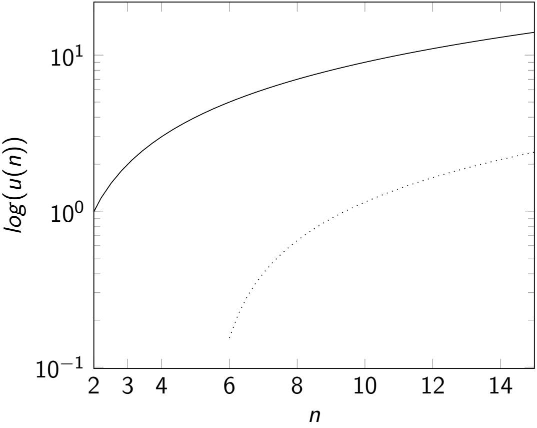
We introduce a simple counting measure $u(n)$ which measures the ratio between $x(n)$ empirical statements to be unified and $y(n)$ unifying theoretical statements, where n is the number of correlated lower-level dispositions $D_1, \ldots, D_n$:

$$u(n) = \frac{x(n)}{y(n)} - 1$$

Note:

- Intervall: $u(n) \in [-1, \infty)$
- $u(n) > 0 \ldots$ theoretical description provides unification
- $u(n) = 0 \ldots$ no gain/cost in providing a theoretical description
- $u(n) < 0 \ldots$ theoretical description more costly than listing the empirical statements

Unificatory Power III



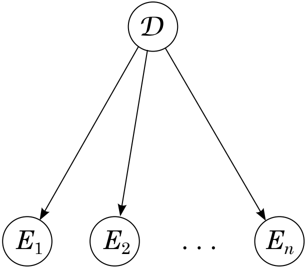
thin solid line: original approach; thin dotted line: Bayesian approach

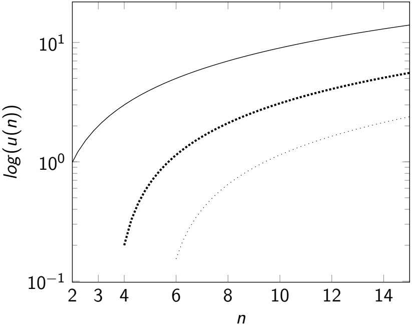
Unificatory Power IV

In the case of strict (unconditional) correlations, the original approach fares better than the Bayesian approach.

⇒ Bayesian framework requires more parametrization

⇒ increased performance by omitting $D_1, \ldots, D_n$



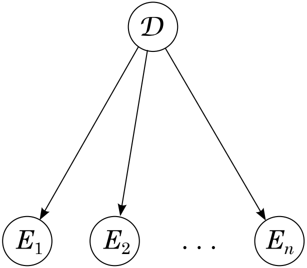


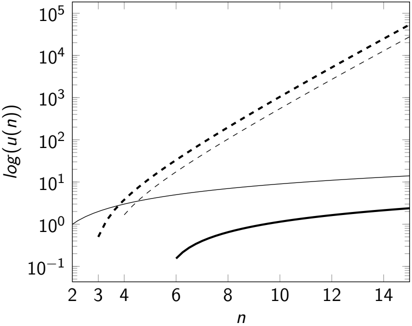
thick dotted line: Bayesian approach (lower-level dispositions omitted)

Unificatory Power V

Bayesian approach gains unificatory power in non-strict probabilistic setting
$\Rightarrow$ up to $2^{n-2} \cdot \binom{n}{2}$ conditional empirical dependencies of the form:

$$P(E_i|E_j, \mathbf{Z}) > P(E_i|\mathbf{Z}), \text{ where } 1 \leq i \neq j \leq n \text{ and } \mathbf{Z} \subseteq \{E_k : 1 \leq i \neq k \neq j \leq n\}. \quad (7)$$

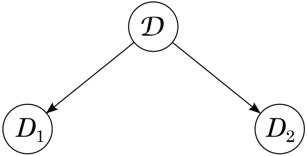




thick dashed line: Bayesian $\mathcal{D}$-approach (maximal conditional dependencies)

Applications

Use-Novel Predictions



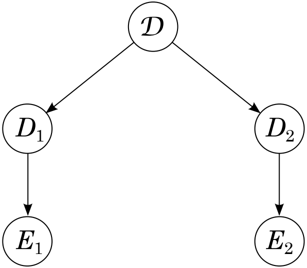
In accordance with (Schurz 2008):

- Variables:

  D_1: attracting iron
  D_2: producing electricity when moved along wire
  $\mathcal{D}$: generating an electro-magnetic field

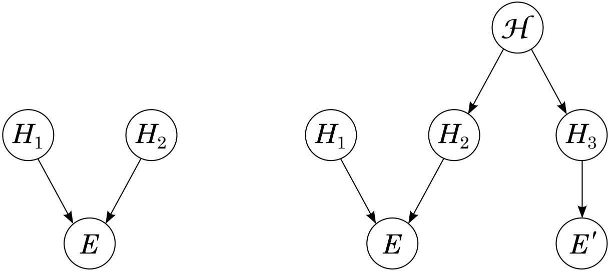
- Empirical finding: A correlation among D_1, D_2 in lodestones.
- Abductive inference: $\mathcal{D}$
- **Novel prediction**: For any x: If $D_1(x)$, then probability increase for $D_2(x)$ and vice versa [with conditions N1,N2]

Confirmation



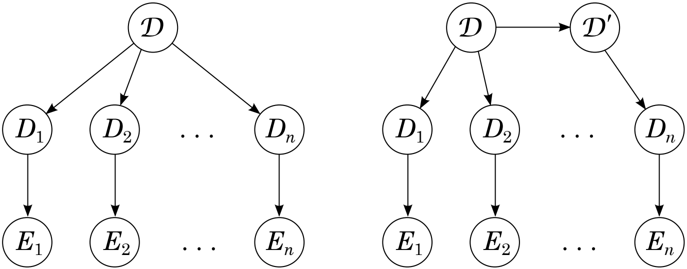
- Empirical finding: correlation among D_1, D_2
- Abductive inference: $\mathcal{D}$
- E_1 confirms not only D_1, but also D_2 [with conditions N1–N3]
- Note: This is structurally similar to *confirmation by analogy*—cf. (Dardashti, Thébault, and Winsberg 2015)

The Problem of Underdetermination



- Assume that choice between H_1, H_2 is underdetermined by E
- Laudan and Leplin (1991) suggest: try to find $\mathcal{H}$ such that $\mathcal{H} \vdash H_2$, but $\mathcal{H} \nvdash H_1$; furthermore: $\mathcal{H} \vdash H_3$, and E' is evidence for H_3
- Their approach can be modeled within the Bayes net framework.
- Creative abduction provides a rationale for their approach: It follows that E' confirms H_2, but not H_1.
  $\Rightarrow$ partial solution to the problem of underdetermination

The Epistemic Challenge: Search



- Note: We aimed at modeling creative abduction, and not at providing an answer to the epistemic questions how and under which conditions it can be applied in practice.
- As Glymour (2018) points out, this problem is tackled in the literature on latent variable search.
- How exactly such approaches fit with the classical literature on abduction has to be investigated in future research.

Conclusion

This talk was about modeling successful cases of creative abduction within a Bayes net framework.

In particular:

- We introduced Schurz (2008) strict approach . . .
- . . . and developed a Bayes net representation.

  ⇒ identifying necessary conditions

  ⇒ more fine-grained investigation of unificatory power
- Finally, we highlighted connections to other issues in PoS: novel predictions, confirmation, underdetermination, search.
- Please see also (Feldbacher-Escamilla and Gebharter 2019c).

References I

Carnap, Rudolf (1936). "Testability and Meaning". In: *Philosophy of Science* 3.4, pp. 419–471. DOI: 10.1086/286432.

— (1937). "Testability and Meaning - Continued". In: *Philosophy of Science* 4.1, pp. 1–40. DOI: 10.1086/286443.

Dardashti, Radin, Thébault, Karim P. Y., and Winsberg, Eric (2015-05). "Confirmation via Analogue Simulation: What Dumb Holes Could Tell Us about Gravity". In: *The British Journal for the Philosophy of Science* 68.1, pp. 55–89. DOI: 10.1093/bjps/axv010.

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

Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2019c-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.

Gebharter, Alexander (2017). *Causal Nets, Interventionism, and Mechanisms. Philosophical Foundations and Applications*. Cham: Springer.

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.

Laudan, Larry and Leplin, Jarrett (1991). "Empirical Equivalence and Underdetermination". In: *The Journal of Philosophy* 88.9, pp. 449–472. DOI: 10.2307/2026601.

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

References II

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

Pearl, Judea (2000). *Causality. Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Reichenbach, Hans (1971). *The Direction of Time*. Ed. by Reichenbach, Maria. Berkeley: University of California Press.

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

— (2016). "Common Cause Abduction: The formation of theoretical concepts and models in science". In: *Logic Journal of the IGPL* 24.4, pp. 494–509. DOI: 10.1093/jigpal/jzw029.

Schurz, Gerhard and Gebharter, Alexander (2016-04). "Causality as a Theoretical Concept: explanatory warrant and empirical content of the theory of causal nets". In: *Synthese* 193.4, pp. 1073–1103. DOI: 10.1007/s11229-014-0630-z.