Unification and Explanation: A causal perspective

Christian J. Feldbacher-Escamilla Alexander Gebharter

Autumn 2021

Project Information

Publication(s):

- Gebharter, Alexander and Feldbacher-Escamilla, Christian J. (submitted). "Unification and Explanation: An unhappy marriage". In: *manuscript*.

Talk(s):

- Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2021-11-10/2021-11-14). *Unification and Explanation: A causal perspective*. Conference. Presentation (contributed). PSA2020. Baltimore: Philosophy of Science Association.
- Feldbacher-Escamilla, Christian J. and Gebharter, Alexander (2020-08-24/2020-08-28). *Unification and Explanation: A causal perspective*. Conference. Presentation (contributed). ECAP 10. European Congress of Analytic Philosophy. University of Utrecht: University of Utrecht.

Project(s):

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

Introduction

A hypothesis' ability to unify and systemise different and diverse pieces of evidence is generally seen as an epistemic virtue.

Unification is related to: confirmation, causation, prediction, explanation

$\Longrightarrow$ **Causation matters!** $\Longleftarrow$

Focus on two influential views about unification:

- Lange's common origin account
- Myrvold's mutual information account

We use causal Bayesian networks and go through different basic causal structures and highlight limitations of both accounts.

We then show that adding structural constraints overcomes these problems.

However, we note that this fix does not generalise to complex structures.

So, unification does not track explanation.

Contents

Two Views of Unification

Common Origin Unification: COU

According to (COU), a hypothesis h unifies pieces of evidence e_1 and e_2 by positing a common origin.

This approach is famously defended by Lange (2004) and Janssen (2002).

An example:

- h: patient suffers from an influenza
- e_1: typical symptom headache
- e_2: typical symptom fever
- There is a correlation between e_1 and e_2.
- Positing as common origin h renders e_1,e_2 less informative about each other.

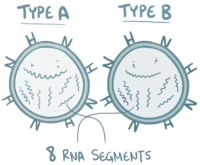
Mutual Information Unification: MIU

According to (MIU), a hypothesis h unifies some evidence e_1 and e_2 insofar as it renders the evidence more informative about each other.

This approach is famously defended by Myrvold (2003, 2017).

An example:

- h: patient suffers from influenza
- e_1: virus of type A is present in patient
- e_2: virus of type B is present in patient
- *Per se*, e_1 and e_2 are independent.
- Knowing h renders e_1,e_2 informative about each other.



Two Probabilistic Measures

Myrvold (2003) suggested the following measure for MIU:

- Mutual information: $I(e_1, e_2) = log_2 \left(\frac{P(e_1, e_2)}{P(e_1) \cdot P(e_2)} \right)$
- Relative mutual information: $I(e_1, e_2 | h) = log_2 \left(\frac{P(e_1, e_2 | h)}{P(e_1 | h) \cdot P(e_2 | h)} \right)$

$$MIU(e_1, e_2; h) = I(e_1, e_2 | h) - I(e_1, e_2)$$

As a first take on COU we suggest:

$$COU(e_1, e_2; h) = I(e_1, e_2) - I(e_1, e_2 | h)$$

Relation:

$$MIU(\cdots) < 0 < COU(\cdots) \;\; \dot{\vee} \;\; MIU(\cdots) = 0 = COU(\cdots) \;\; \dot{\vee} \;\; MIU(\cdots) > 0 > COU(\cdots)$$

The two measures are opposites w.r.t. rendering evidence un-/informative.

Unification and Causation

Causal Bayesian Networks

In accordance with Wheeler and Scheines (2013, p.157) we think that ...

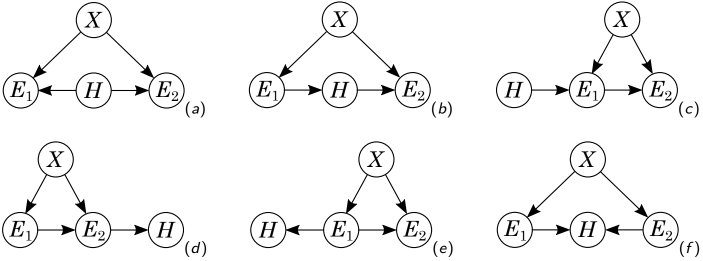
> *"it is necessary to take into consideration the causal structure that might regulate the relationships between evidence and hypothesis".*

We represent causal structure via causally interpreted Bayesian networks combining a directed acyclic graph with a probability distribution.

Causal Bayesian networks conform to the Markov factorisation:

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

Elementary Causal Structures



Assumptions:

- E_i is evidence for H: $P(e_i|h) > P(e_i|\bar{h})$
- E_i are independent or positively dependent: $P(e_i|x) \geq P(e_i|\bar{x})$

Intuition:

- Unification in (a)–(c)
- No unification in (d)–(f)

Unification in the Bayesian Network Setup

For these elementary structures we get:

Observation

$$MIU(e_1, e_2; h) < 0 < COU(e_1, e_2; h) \quad \text{for structures (a)–(f)}$$

Problem:

- MIU underperforms (cases (a)–(c))
- COU is way to permissive (cases (d)–(f))

In the following, we back our intuition regarding (a)–(f) by linking it to explanatory constraints.

Afterwards, we show how to improve COU.

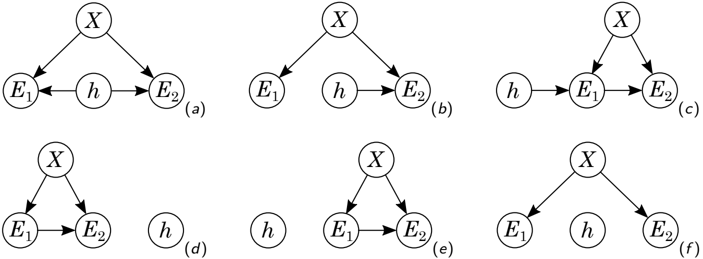
Unification and Explanation

Explanation in the Bayesian Network Setup

In causal settings: Explanation tracks causation!
(cf. J. Woodward 2003; J. F. Woodward and Hitchcock 2003; Hitchcock and J. F. Woodward 2003)

What-if-things-had-been-different questions in our setup ≈ intervention

Structures resulting from setting H to h by intervention ($\hat{h}$):



Explanatory Power

What is h's explanatory power w.r.t. $\mathbf{E} = E_1 \times E_2$?

Intuition from above: information about $\mathbf{E} \uparrow \;\Rightarrow\;$ explanatory power of $h \uparrow$

$\uparrow\mathbf{E}$-information if $\downarrow\mathbf{E}$-uncertainty $\approx$ entropy (Sprenger and Hartmann 2019):

$$\underbrace{\mathcal{H}(\mathbf{E}) = -\sum_{\mathbf{e}\in\mathbf{E}} P(\mathbf{e}) \times log_2 P(\mathbf{e})}_{\textit{unconditional}} \qquad \underbrace{\mathcal{H}(\mathbf{E}|\hat{h}) = -\sum_{\mathbf{e}\in\mathbf{E}} P(\mathbf{e}|\hat{h}) \times log_2 P(\mathbf{e}|\hat{h})}_{\textit{conditional}}$$

Explanatory Power (cf. Gebharter and Eronen ms)

$$EXP(\mathbf{E}; h) = \underbrace{\mathcal{H}(\mathbf{E}) - \mathcal{H}(\mathbf{E}|\hat{h})}_{\text{reduction of uncertainty}}$$

Result: $EXP(\mathbf{E}; h) > 0$ for (a)–(c) and $EXP(\mathbf{E}; h) = 0$ for (d)–(f) ✓ *intuition*

Robustness: also for other measures (cf. Schupbach and Sprenger 2011)

Unification and Explanation in the Bayesian Network Setup

Relationship of unificatory to explanatory power:

#	Model	EXP	MIU	Match	COU	Match
(a)	$E_1 \longleftarrow H \longrightarrow E_2$	> 0	< 0	$\times$	> 0	$\checkmark$
(b)	$E_1 \longrightarrow H \longrightarrow E_2$	> 0	< 0	$\times$	> 0	$\checkmark$
(c)	$H \longrightarrow E_1 \longrightarrow E_2$	> 0	< 0	$\times$	> 0	$\checkmark$
(d)	$E_1 \longrightarrow E_2 \longrightarrow H$	$= 0$	< 0	$\checkmark$	> 0	$\times$
(e)	$H \longleftarrow E_1 \longrightarrow E_2$	$= 0$	< 0	$\checkmark$	> 0	$\times$
(f)	$E_1 \longrightarrow H \longleftarrow E_2$	$= 0$	< 0	$\checkmark$	> 0	$\times$

Again, MIU underperforms and COU is too permissive.

Common Causal Origin Unification

We can do better. Idea: also COU-unification needs to track causation.

We can guarantee this by adding a causal structural constraint:

Causal Common Origin Unification

$$CCOU(e_1, e_2; h) = I(e_1, e_2) - I(e_1, e_2 | \hat{h})$$

CCOU applied to our elementary causal structures leads to the desired result:

Observation

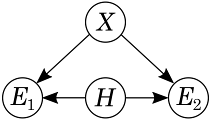
$$CCOU(e_1, e_2; h) > 0 \quad \text{for structures (a)–(c)}$$
$$CCOU(e_1, e_2; h) = 0 \quad \text{for structures (d)–(f)}$$

Common Causal Origin Unification: A Problem

So, we see that the behaviour of *CCOU* ordinally coincides with that of *EXP* with respect to a simple causal setup.

Does this generalise to more complex setups? The answer is: no.

We can construct a counterexample to *CCOU*'s tracking explanation for:



If we take MIU and COU as the key approaches in the field of unification, we conclude that unification and explanation do not go hand in hand as claimed by several authors (Kitcher 1981, 1989; Lange 2004).

Upholding such a relation comes at the cost of an increased need of modification and parametrization ⇒ degenerative research programme.

Summary

- We have provided a probabilistic measure for Lange's account of common origin unification.
- We then checked for a relation between unification and causal structure and found that both measures have some deficiencies.
- We could verify the deficiencies also with respect to the relation between unification and explanatory power.
- We showed that by implementing the same structural constraint that is relevant for explanatory power into our measure of common origin unification one can overcome these deficiencies.
- However, we also noted that this solution does not work for more general structures.
- So, the relation between unification and explanation is a "troubled" one.

References I

Gebharter, Alexander and Eronen, Markus I. (ms). "Quantifying Proportionality and the Limits of Higher-Level Causation and Explanation". In: *manuscript*.

Gebharter, Alexander and Feldbacher-Escamilla, Christian J. (submitted). "Unification and Explanation: An unhappy marriage". In: *manuscript*.

Hitchcock, Christopher and Woodward, James F. (2003). "Explanatory Generalizations, Part II: Plumbing explanatory depth". In: *Noûs* 37.2, pp. 181–199. DOI: 10.1111/1468-0068.00435.

Janssen, Michel (2002). "COI Stories: Explanation and Evidence in the History of Science". In: *Perspectives on Science* 10.4, pp. 457–522. DOI: 10.1162/106361402322288066.

Kitcher, Philip (1981). "Explanatory Unification". In: *Philosophy of Science* 48.4, pp. 507–531. DOI: 10.1086/289019.

— (1989). "Explanatory Unification and the Causal Structure of the World". In: *Scientific Explanation*. Ed. by Kitcher, Philip and Salmon, Wesley C. Minneapolis: University of Minnesota Press, pp. 410–505.

Lange, Marc (2004). "Bayesianism and Unification: A Reply to Wayne Myrvold". In: *Philosophy of Science* 71.2, pp. 205–215. DOI: 10.1086/383012.

Myrvold, Wayne C. (2003). "A Bayesian Account of the Virtue of Unification". In: *Philosophy of Science* 70.2, pp. 399–423. DOI: 10.1086/375475.

— (2017). "On the Evidential Import of Unification". In: *Philosophy of Science* 84.1, pp. 92–114.

References II

Schupbach, Jonah N. and Sprenger, Jan (2011). "The Logic of Explanatory Power". In: *Philosophy of Science* 78.1, pp. 105–127. DOI: 10.1086/658111.

Sprenger, Jan and Hartmann, Stephan (2019). *Bayesian Philosophy of Science. Variations on a Theme by the Reverend Thomas Bayes*. Oxford: Oxford University Press.

Wheeler, Gregory and Scheines, Richard (2013-06). "Coherence and Confirmation through Causation". In: *Mind* 122.485, pp. 135–170. DOI: 10.1093/mind/fzt019.

Woodward, James (2003). *Making Things Happen. A Theory of Causal Explanation*. Oxford: Oxford University Press.

Woodward, James F. and Hitchcock, Christopher (2003). "Explanatory Generalizations, Part I: A counterfactual account". In: *Noûs* 37.1, pp. 1–24. DOI: 10.1111/1468-0068.00426.