Unification and Explanation from a Causal Perspective[*]

Alexander Gebharter Christian J. Feldbacher-Escamilla

Spring 2023

Abstract

[28] We discuss two influential views of unification: mutual information unification (MIU) and common origin unification (COU). We propose a simple probabilistic measure for COU and compare it with Myrvold's (2003; 2017) probabilistic measure for MIU. We then explore how well these two measures perform in simple causal settings. After highlighting several deficiencies, we propose causal constrains for both measures. A comparison with explanatory power shows that the causal version of COU is one step ahead in simple causal settings. However, slightly increasing the complexity of the underlying causal structure shows that both measures can easily disagree with explanatory power. The upshot of this is that even sophisticated causally constrained measures for unification ultimately fail to track explanatory relevance. This shows that unification and explanation are not as closely related as many philosophers thought.

Keywords: unification, explanation, causation

1 Introduction

A hypothesis' ability to unify and systematize different and diverse pieces of evidence is generally seen as an epistemic virtue in philosophy of science. Unification is also often associated with other core concepts of philosophy of science such as abduction, confirmation, causation, prediction, and explanation. In this paper, we bracket abduction, confirmation, and prediction and rather focus on two different views of unification and their connection to explanation from a causal perspective. Taking such a causal perspective will allow us to see that causal structure matters for how probabilistic measures of unification perform and relate to explanatory relevance. Many philosophers of science hold the view that the better a hypothesis h unifies a body of evidence $e_1, \ldots, e_n$,

[*][This text is published under the following bibliographical data: Gebharter, Alexander and Feldbacher-Escamilla, Christian J. (2023). "Unification and Explanation from a Causal Perspective". In: *Studies in History and Philosophy of Science* 99, pp. 28–36. DOI: 10.1016/j.shpsa.2022.12.005. All page numbers of the published text are in square brackets. The final publication is available at https://doi.org/10.1016/j.shpsa.2022.12.005. For more information about the underlying project, please have a look at http://cjf.escamilla.academia.name.]

the better it can explain it. So, for example, unification in the general context of explanation is discussed by Kneale (1949, pp.91f), Hempel (1965, p.444), and Friedman (1974). For a discussion of the latter, see, for example, J. Woodward (2003). Also, Whewell's (1840/2014) *consilience* is sometimes considered a form of unification. Some even go as far as proposing that explanation can be reduced to unification (see, e.g., Kitcher 1981, 1989). In this paper, however, we rather focus on the more general question of whether unification is a good indicator for explanatory relevance.

Which account of unification gets things right and how exactly unificatory power can be measured is still controversial. In this paper, we are especially interested in the following two prominent approaches to unification:

Mutual information unification (MIU): A hypothesis h has more unificatory power with respect to pieces of evidence $e_1, \ldots, e_n$ the more it renders these pieces of evidence (more) informative about each other.

Common origin unification (COU): A hypothesis h unifies a body of evidence $e_1, \ldots, e_n$ in so far as it posits a common origin for these pieces of evidence.

MIU has been defended by Myrvold (2003, 2017), and COU by Lange (2004). Both authors have criticized each other's account. Myrvold (2017) stressed that according to a Bayesian decomposition, only "MIU contributes to incremental evidential support, and that there is no scope, within Bayesian updating, for COU to add to the evidential support of the theory" (p. 93). And Lange (2004) claimed that "genuinely to unify [pieces of evidence], a theory must reveal them to have some deep common explanatory basis" (p. 208) and that Myrvold's account is inadequate because it "sets the bar too low to distinguish genuine from bogus unification" (ibid.). So, whereas Myrvold claims that COU is inadequate in terms of confirmation, Lange claims that MIU is inadequate in terms of explanation. Against Myrvold's claim about COU, Niiniluoto (2016) argued that COU also plays an important role for confirmation if more broadly conceived as including abductive confirmation. [29] However, Niiniluoto's observation does not automatically make COU a better candidate to account for the different roles typically associated with unification.

The main goal of this paper is to shed new light on how MIU and COU relate to explanation.[1] To this end, we take a causal perspective which, as we will see, makes it easier to explore how well the two measures perform when it comes to indicating explanatory relevance. We draw on Reichenbach's (1956) insight that common causes screen off their effects (or render them less informative about each other in the presence of additional causal connections) that has become a crucial assumption in causal modeling. Based on this simple

[1] We acknowledge that Myrvold (2017) makes it clear that, according to MIU, a hypothesis can unify a body of evidence without explaining it. But we believe that it is nevertheless an interesting project to examine whether MIU or a modification of it more suitable for causal settings might provide a measure of explanatory unification, especially since MIU's lack of tracking explanatory relevance was one of Lange's (2004) main objections against MIU.

idea, we propose a probabilistic measure for COU that in some sense complements Myrvold's probabilistic measure for MIU: According to this first probabilistic take on COU, a hypothesis has the more unificatory power the more it renders pieces of evidence uninformative about each other. As a next step, we will use causal Bayesian networks (Pearl 2000; Spirtes, Glymour, and Scheines 1993) to represent different patterns of how a hypothesis can be causally connected to a body of evidence. As we will see, already focussing on the simplest causal patterns suffices to make some relevant observations. We apply Myrvold's (2003; 2017) measure for MIU and our novel take on measuring COU to each of these structures. The upshot of this will be that causal structure heavily constrains the performance of these probabilistic measures. Next, we use the basic causal structures and our results about how the measures for MIU and COU behave to shed new light on the connection of unification and explanation in causal settings. It will turn out that both probabilistic measures of unification do a bad job as indicators for explanatory power. While the measure for MIU underperforms when applied to the elementary causal structures we discuss, the probabilistic measure for COU is too permissive. Based on this observation, we further develop the probabilistic measures for MIU and COU by adding a causal constraint. We then compare the modified measures in the context of the elementary causal structures again. On first glance, COU will seem to have the upper hand, but, as we will see, increasing the complexity of the underlying causal structure only slightly can mess things up easily. This shows that even sophisticated and causally constrained measures for unification ultimately fail to track explanatory relevance. The upshot of this is that unification and explanation do not go hand in hand: They are not as closely related as authors such as Kitcher (1981, 1989) and Lange (2004) proposed.

The structure of the paper is as follows: In section 2, we introduce Myrvold's (2003; 2017) measure for MIU. We then propose a complementary probabilistic measure for COU and discuss its relation to Myrvold's measure. In section 3, we investigate how these measures for MIU and COU perform given different elementary causal structures. In section 4, we explore what we can learn from our earlier results about the relation between unification and explanation and how we could try to account for these insights by adding a causal constraint to our probabilistic measure for COU and how this causal measure for COU performs compared to a similarly modified version of the measure for MIU. It will turn out that though adding these causal constraints increases the performance of both causal measures in the elementary structures discussed so far, the causal COU-measure takes the upper hand. In section 5, however, we show that even the causal version of the COU-measure fails as an indicator for explanatory relevance once one increases the complexity of the underlying causal structure only slightly. We conclude in section 6. For the proofs of the observations made throughout the paper, see the Appendix.

2 Two kinds of unification and two probabilistic measures

In this paper, we are interested in two prominent but somewhat opposed views of unification. According to the more classical view (COU), unification consists in identifying a common origin of the pieces of evidence to be unified. COU has been prominently defended by authors such as Lange (2004) and Janssen (2002). The other view (MIU) says that unification consists in rendering pieces of evidence informative (or more informative) about each other. It has been prominently defended by Myrvold (2003, 2017).[2]

Before introducing Myrvold's (2003; 2017) measure, let us illustrate the basic idea by means of an intuitive example: Let us assume that the price of a good on the market is determined by supply and demand. More precisely, we suppose that it is very likely for a good to have a high price if demand is very high and supply is very low. To the background of any other combination of supply and demand, in contrast, a high price is way less likely. Now assume that the price of a certain good is comparably high. In this situation we can infer that if supply is very low, then it is more likely that demand is very high (and vice versa). Thus, the price being high would render low supply and high demand more informative about each other. If the hypothesis says that a good's price is high and low supply and high demand are the pieces of evidence, then, according to MIU, the hypothesis would unify the body of evidence.

Myrvold (2003; 2017) defines the notion of mutual information between two pieces of evidence e_1 and e_2 that is involved in this form of reasoning as follows:[3]

Definition 2.1 (mutual information and relative mutual information)**.**

$$I(e_1, e_2) = log_2\left(\frac{Pr(e_1, e_2)}{Pr(e_1) \cdot Pr(e_2)}\right)$$

$$I(e_1, e_2|h) = log_2\left(\frac{Pr(e_1, e_2|h)}{Pr(e_1|h) \cdot Pr(e_2|h)}\right)$$

The pieces of evidence e_1 and e_2 are independent if $Pr(e_1, e_2) = Pr(e_1) \cdot Pr(e_2)$. Since $Pr(e_1, e_2)$ is compared to the product $Pr(e_1) \cdot Pr(e_2)$ in Definition 2.1, a positive value of $I(e_1, e_2)$ indicates that the pieces of evidence are

[2] A similar account has been put forward by McGrew (2003). Schupbach (2005) showed that both accounts are ordinally equivalent, so we can bracket McGrew's account for our endeavor.

[3] Throughout the paper we focus on the simplest case only involving two pieces of evidence. Also note that we leave the specific interpretation of probabilities open. They can be interpreted objectively, for example as capturing the true regularity patterns as they appear in the Humean mosaic or as limiting frequencies of observed regularities. They can also be interpreted subjectively as credences of an agent or a group of agents such as a scientific community. Also the causal settings which we will introduce later on are best interpreted in accordance to one's interpretation of probabilities: Objective interpretations of causation go best with an objective interpretation of probabilities while subjective ones are best associated with a view of causation as a purely conceptual tool useful for structuring data.

positively dependent and a negative value stands for negative dependence, where the logarithm shifts the neutral case to 0. Likewise for relative mutual information $I(e_1, e_2|h)$.

Now, Myrvold (2003, 2017) suggests to identify the degree to which a hypothesis unifies pieces of evidence according to MIU with the following difference: [30][4]

Definition 2.2 (mutual information unification)**.**

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

The idea here is, again, that a unifying hypothesis increases the amount of mutual information between the pieces of evidence it unifies. Accordingly, $MIU > 0$ indicates positive unificatory power, $MIU < 0$ indicates negative unificatory (or disunificatory) power, and $MIU = 0$ indicates no unificatory power at all.

While Myrvold (2003, 2017) suggested a measure for MIU, there is still no measure for the unificatory power of a hypothesis according to COU on the market. So how could COU be measured? As Niiniluoto (2016) remarks, one kind of common origin is a common cause that comes with a specific probabilistic property: It screens off its effects (if no other causal relations are around).[5] Niiniluoto identifies this screening off property as the central feature underlying COU. Taking this observation as a starting point, we suggest the following quite intuitive basic idea for measuring COU in terms of probabilities: Two pieces of evidence e_1 and e_2 accounted for by some hypothesis h are the more unified by h, the more of the dependence between e_1 and e_2 is reduced by assuming h. In other words: h has the more unificatory power w.r.t. e_1 and e_2 the more of the dependence between e_1 and e_2 can actually be accounted for in terms of h.

Let us illustrate the basic idea, again, by help of a simple example. Assume that we are interested in the hypothesis that a patient suffers from an influenza. We are observing two typical symptoms: headache and fever. If the patient has a headache, then also the probability for fever will be higher, and vice versa. But if we learn that the patient actually suffers from an influenza, both symptoms as well as why they are correlated can be explained by that fact, meaning that the hypothesis that the patient suffers from an influenza will render the two pieces of evidence less informative about each other.

Based on the simple intuitive idea above, we propose—as a first take on COU—a probabilistic measure:[6]

[4]Actually, Myrvold (2003, 2017) proposes several different but interrelated measures for MIU and connects them to different measures of confirmation. Since we are not interested in the relation to confirmation in this paper, we will focus on one of these measures only. Also note that "MIU" refers to the measure for mutual information unification while "MIU" refers to the account; likewise for common origin unification.

[5]Other kinds of common origin that share this screening off property with common causes are common supervenience bases (Gebharter 2017a), common constituents (Gebharter 2017b, 2022) and common grounds (Schaffer 2016).

[6]At this point we would once again like to stress that this measure is not Lange's (2004), but

Definition 2.3 (common origin unification)**.**

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

Again, the idea is that $COU(e_1, e_2; h)$ measures how much of the dependence between e_1 and e_2 is reduced by h in probabilistic terms. The higher COU is (given a positive value), the more of the probabilistic dependence between e_1 and e_2 can be accounted for by h. If COU is zero, h does not have any unificatory power, and if it is negative, h disunifies e_1 and e_2.

Since MIU and COU are about opposed probabilistic phenomena, it is no wonder that their respective probabilistic measures MIU and COU are also opposed:

Observation 2.1.

$$COU(e_1, e_2; h) = -MIU(e_1, e_2; h)$$

As we will see in section 4, the relation between unificatory and explanatory power can be discussed most efficiently in terms of inequalities, for which reason we also state the relations between MIU and COU in terms of the following equations:

Observation 2.2.

$$\text{If } MIU(e_1, e_2; h) > 0, \text{ then } COU(e_1, e_2; h) < 0. \tag{1}$$

$$\text{If } MIU(e_1, e_2; h) < 0, \text{ then } COU(e_1, e_2; h) > 0. \tag{2}$$

$$MIU(e_1, e_2; h) = COU(e_1, e_2; h) \text{ iff } I(e_1, e_2) = I(e_1, e_2|h). \tag{3}$$

These equations nicely illustrate the idea that MIU and COU are two opposite views of unification: If a hypothesis h does unify according to MIU, then it disunifies according to COU (Equation 1), and the other way round: If h disunifies according to MIU, then it unifies according to COU (Equation 2). Finally, the only case in which the measures agree is the case when h has no influence at all on the amount of information that e_1 and e_2 bear on each other (Equation 3). In this case, both measures are zero.

---

our reconstruction of Lange's COU account based on Niiniluoto's (2016) analysis. One might be worried that measuring unificatory power in terms of screening off is an implausible strategy right from the start since, as Sober (1988, ch. 3) argues, separate-cause explanations can screen off as much as common cause explanations. Assume, for example, a patient has headache and fever. Assume further that there are three different diseases: A causes headache but not fever, B causes fever but not headache, and C causes both symptoms. Now, conditioning on $A\&B$ screens headache and fever as much off as conditioning on the common cause C. We believe, however, that such an analysis is far from being a knock-down argument against screening off as a measure for COU. Just to mention one problem, neither A, B, nor the conjunction $A\&B$ classifies as an *origin* in any ordinary meaning of that word. (Labeling $A\&B$ a single origin of headache and fever is nothing over and above a logical trick.) But COU is all about unifiying a body of evidence in terms of a single origin rather than in terms of several independent origins.

3 Unification and causation

In section 2, we motivated our probabilistic measure for COU on the basis of the observation that many (or even all) common origins behave like common causes: They screen off their effects. Note, however, that both measures, *MIU* as well as *COU*, are purely probabilistic in nature; so far no causal considerations or any specific constraints about other types of common origin have been built into these measures. In this section, we explicitly turn to causal settings. In such settings hypotheses and pieces of evidence can stand in all kinds of causal relationships to each other. By taking this causal stance we follow Wheeler and Scheines (2013) who claim that "it is necessary to take into consideration the causal structure that might regulate the relationships between evidence and a hypothesis" (p. 157).

Again we focus on the most simple case involving one hypothesis and two pieces of evidence. In particular, we will consider all the basic possibilities how a single hypothesis and two pieces of evidence can be causally connected and then explore how *MIU* and *COU* perform in these causal structures. We represent causal structures as causally interpreted Bayesian networks (CBNs). A CBN is a structure $\langle \mathbf{V}, \mathbf{E}, Pr \rangle$ in which $\mathbf{V}$ is a set of random variables, $\mathbf{E}$ is a set of directed edges ($\longrightarrow$) connecting variables in $\mathbf{V}$, and Pr is a probability distribution over $\mathbf{V}$. Like all Bayesian networks, CBNs conform to the Markov factorization

$$Pr(x_1, \ldots, x_n) = \prod_{i=1}^{n} Pr(x_i | \mathbf{par}(X_i)), \tag{4}$$

where $\mathbf{Par}(X_i)$ stands for the set of a variable X_i's parents—i.e., the set of variables X_j such that $X_j \longrightarrow X_i$ is in $\mathbf{E}$—and $\mathbf{par}(X_i)$ for the instantiation of the variables in $\mathbf{Par}(X_i)$ to some values. The CBN formalism (Pearl 2000; Spirtes, Glymour, and Scheines 1993) can be seen as a theoretical generalization of Reichenbach's (1956) insight: One implication of the fact that CBNs adhere to the Markov factorization is that the effects of common causes become independent when conditionalizing on these common causes (given no other causal connections are around).

In the following, we will represent the hypothesis and the two pieces of evidence with the binary variables H, E_1, and E_2, respectively. H's value h stands for the truth of the hypothesis and $\bar{h}$ for its falsity. The value e_i of E_i (with $i \in \{1, 2\}$) stands for the fact that the event constituting a piece of evidence occurs and $\bar{e}_i$ for the fact that it does not occur. We limit ourselves to positive probability distributions and to cases where pieces of evidence are independent or depend positively on each other. [31] Since e_1 and e_2 are pieces of evidence for a hypothesis h, we assume that e_1 and e_2 both depend positively on the hypothesis h, which is a quite typical assumption about the relation of hypothesis and evidence (cf. Bovens and Hartmann 2003):

$$Pr(e_1|h) > Pr(e_1|\bar{h}) \tag{5}$$

$$Pr(e_2|h) > Pr(e_2|\bar{h}) \tag{6}$$

Now, there are 24 possible elementary patterns how the three variables H, E_1, and E_2 can be connected by two directed edges. Having a look at these most basic causal structures as listed in Table 1 will give us a good idea how causal structure constrains the performance of the measures.

$H \longleftarrow E_1 \longrightarrow E_2$	$H \longrightarrow E_1 \longrightarrow E_2$	$H \longleftarrow E_1 \longleftarrow E_2$	$H \longrightarrow E_1 \longleftarrow E_2$
$H \longleftarrow E_2 \longrightarrow E_1$	$H \longrightarrow E_2 \longrightarrow E_1$	$H \longleftarrow E_2 \longleftarrow E_1$	$H \longrightarrow E_2 \longleftarrow E_1$
$E_1 \longleftarrow H \longrightarrow E_2$	$E_1 \longrightarrow H \longrightarrow E_2$	$E_1 \longleftarrow H \longleftarrow E_2$	$E_1 \longrightarrow H \longleftarrow E_2$
$E_2 \longleftarrow H \longrightarrow E_1$	$E_2 \longrightarrow H \longrightarrow E_1$	$E_2 \longleftarrow H \longleftarrow E_1$	$E_2 \longrightarrow H \longleftarrow E_1$
$E_1 \longleftarrow E_2 \longrightarrow H$	$E_1 \longrightarrow E_2 \longrightarrow H$	$E_1 \longleftarrow E_2 \longleftarrow H$	$E_1 \longrightarrow E_2 \longleftarrow H$
$E_2 \longleftarrow E_1 \longrightarrow H$	$E_2 \longrightarrow E_1 \longrightarrow H$	$E_2 \longleftarrow E_1 \longleftarrow H$	$E_2 \longrightarrow E_1 \longleftarrow H$

Table 1: List of all possible elementary patterns connecting H, E_1, E_2.

Only the six structures in black are relevant for our endeavor. Grey structures are irrelevant for at least one of the following reasons: (i) They are identical to black structures from a graph-theoretic perspective.[7] (ii) They only differ from black structures because E_1 and E_2 swapped places, which will make no difference for the performance of the measures. (iii) They are excluded by Equations 5 and 6. We thus arrive at the six elementary structures in Figure 1.

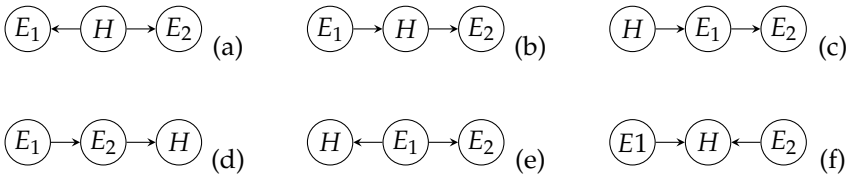

Figure 1: Elementary causal structures connecting H, E_1, and E_2

With the assumptions made earlier in place we can now apply the measures *MIU* and *COU* to the six basic structures in Figure 1. If we do that, we arrive at the following somewhat surprising finding:

Observation 3.1.

$MIU(e_1, e_2; h) < 0 < COU(e_1, e_2; h)$ for structures (a)–(e).
$MIU(e_1, e_2; h), COU(e_1, e_2; h)$ are lesser, equal, or greater 0 for structure (f).

While *MIU* underperforms[8] by indicating disunification in the basic structures (a)–(e), *COU* is way too permissive by indicating unificatory power in all

[7] For example, $H \longrightarrow E_1 \longrightarrow E_2$ and $E_2 \longleftarrow E_1 \longleftarrow H$ formally represent the same graph $\mathbf{G} = \langle \mathbf{V}, \mathbf{E} \rangle$, where $\mathbf{V} = \{H, E_1, E_2\}$ is $\mathbf{G}$'s set of vertices and $\mathbf{E} = \{\langle H, E_1 \rangle, \langle E_1, E_2 \rangle\}$ is $\mathbf{G}$'s set of directed edges.

[8] We bracket Myrvold's (2017) proposal to cover Reichenbachian common cause hypotheses

of these structures. The paradigmatic case of COU advocates such as Lange (2004) have in mind is clearly the one in Figure 1(a). But not only the common cause case can provide unificatory power according to *COU*. Another interesting case is (c) in which the hypothesis directly causes one piece of evidence which then directly causes the other piece of evidence. This case has not been explicitly considered by Lange. In some sense the hypothesis can, however, still be considered as a common origin of both pieces of evidence, namely in the sense that it causes one via causing the other. Also interesting is the structure in (b). Though the hypothesis is clearly not a common origin of the two pieces of evidence, it has unificatory power according to the basic idea underlying COU that a hypothesis' unificatory power consists in the amount of informativeness the pieces of evidence bear on each other reduced by the hypothesis. Even more extreme are (d) and (e). Though the hypothesis is not even an origin of one of the pieces of evidence, it still has unificatory power according to the probabilistic measure *COU*.

Finally, (f) is also interesting because this is then the only case in which both *MIU* and *COU* can be positive. Their behavior depends on the specific way the causes E_1, E_2 interact with each other in bringing about the effect H. If, for example, each piece of evidence alone makes the hypothesis highly probable, then conditioning on h will render e_1 and e_2 negatively dependent on each other. Thus, *MIU* will be negative and *COU* will be positive. If, on the other hand, e_1 and e_2 are both required to increase h's probability, then conditioning on h will render e_1 and e_2 positively dependent. Hence, *MIU* will be positive and *COU* will be negative. Finally, both measures can also be 0. Like for structures (d) and (e), the result about structure (f) is bad news for *COU* since it can be positive though the hypothesis is not even an origin of one of the two pieces of evidence. For exemplary probability distributions, see the Appendix.

The upshot of our investigation so far is that causal structure crucially constrains the performance of the probabilistic measures for unificatory power in unexpected ways. As outlined before, Lange (2004) thinks that one of the most important virtues of a hypothesis that unifies according to COU is its purported ability to also explain the body of evidence it unifies. In the next section, we will see in detail in which elementary causal scenarios a hypothesis' ability to unify does actually indicate its explanatory relevance. We will then take this as a basis for adding a causal constraint to *COU* that brings it closer to Lange's original understanding of COU and makes it more suitable for the application to the simple causal settings considered so far.

as instances of MIU in this paper. In a nutshell, Myrvold suggests that such hypotheses should be considered as postulating a common cause. Accordingly, one would have to compare two models: One with H and one without H. Adding H as a common cause would, however, render e_1 and e_2 more informative about each other only if adding H adds something to the positive dependence between e_1 and e_2. This means, in turn, that the distribution over $\{E_1, E_2\}$ needs to be different in the two models, which seems artificial and a bit ad hoc to us. In the end, it is the specific dependence between e_1 and e_2 (in the actual world) that should be accounted for by introducing the hypothesis. Introducing the hypothesis should explain and not change the actual dependence between e_1 and e_2. A more detailed investigation of Myrvold's strategy must await another occasion.

4 Unification and explanation

There are many different measures for explanatory power currently on the market. For an overview see, for example, (Sprenger and Hartmann 2019). Basically all of these measures are probabilistic difference-making measures, meaning that their output increases the more the hypothesis increases the body of evidence's probability. Because of this, we can choose any of these measures as a proxy. For our further investigation we choose Schupbach and Sprenger's (2011) measure. However, the results we provide also hold for any other probabilistic difference-making measure. Their measure is as follows:

$$EXP(\mathbf{e}; h) = \frac{Pr(h|\mathbf{e}) - Pr(h|\bar{\mathbf{e}})}{Pr(h|\mathbf{e}) + Pr(h|\bar{\mathbf{e}})}$$

Note that this measure (as well as related measures) are purely probabilistic and, at least in their standard form, not yet suitable for causal settings. In causal settings, explanation is typically assumed to track causation in order to account for explanatory asymmetries (cf. J. Woodward 2003). One of the paradigmatic examples to illustrate this point is about the solar altitude being causally relevant for the length of a flagpole's shadow, but not the other way round. The two phenomena are probabilistically dependent since a lower solar altitude results in a longer shadow than a higher one does. Let us choose h to stand for low solar altitude or long shadow and $\mathbf{e}$ for the other of the two factors. Regardless of how we choose, $EXP(\mathbf{e}; h)$ will always be positive since each factor is a positive probabilistic difference-maker for the other one. However, while pointing to the low solar altitude results in a good explanation for the long shadow, the other way round does not. The simple reason for this is that causes can explain their effects, but not vice versa.

To get a measure for explanatory power suitable for causal settings, one needs to guarantee that explanation tracks causation. For the classical difference-making measures, this can be achieved with the help of the technical notion of an ideal intervention (see, e.g., Pearl 2000; Sprenger 2018). Recently, Eva and Stern (2019) have proposed such a modification for the Schupbach and Sprenger (2011) measure. [32] The classical probabilistic measures as well as Eva and Stern's (2019) measure capture the intuition that explanatory power consists in the hypothesis' ability to increase the body of evidence's probability. Explanatory power can, however, also be understood in terms of answering *what-if-things-had-been-different questions* or w-questions (Hitchcock and J. F. Woodward 2003; J. F. Woodward and Hitchcock 2003): The greater the range of w-questions a hypothesis can answer, the more explanatory power it has. A measure that captures this idea has recently been proposed by Gebharter and Eronen (forthcoming). However, since probabilistic difference-making measures such as Eva and Stern's modification of the Schupbach and Sprenger (2011) measure are closer to the more traditional unification debate, we go for this measure in this paper. In particular, we use a simplified version of their

account:[9]

$$CEXP(\mathbf{e}; h) = \frac{Pr(\hat{h}|\mathbf{e}) - Pr(\hat{h}|\bar{\mathbf{e}})}{Pr(\hat{h}|\mathbf{e}) + Pr(\hat{h}|\bar{\mathbf{e}})} \tag{7}$$

e stands short for the whole body of evidence to be explained. In our case, this means that **e** stands for e_1, e_2 and that $\bar{\mathbf{e}}$ stands for $e_1, \bar{e}_2 \vee \bar{e}_1, e_2 \vee \bar{e}_1, \bar{e}_2$. The hat symbol above h stands for an intervention that decouples H from the probabilistic influence of its causes. Decoupling H from the influence of its causes allows for isolating the probabilistic influence of h on **e** (and, vice versa, the probabilistic influence of **e** on h) that arises because H is a cause of **E**. Under such an intervention all the possible probabilistic influence the hypothesis might have on the body of evidence (and vice versa) over other paths (e.g., paths featuring a common cause of both the hypothesis and the evidence) is ignored. Considering only the probabilistic influence due to directed paths from the hypothesis to the evidence allows $CEXP$ to track causation. In particular, the post intervention distribution used in $CEXP$ can be computed by applying the Markov factorization (Equation 4) to the truncated structures one gets from the basic structures in Figure 1 by deleting all the arrows pointing at H. We assume that the post intervention distribution over H is identical to the pre intervention distribution.[10] The post intervention structures are depicted in Figure 2.

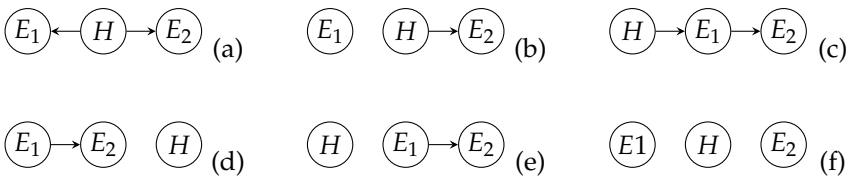

Figure 2: Structures resulting from Figure 1 by intervening on H

Given the assumptions made earlier, we can now draw the following conclusions for our basic causal structures in Figure 1:

[9]Nothing hinges on that particular choice. The other classical probabilistic difference-making measures on the marked can be transformed into causal measures in the same way and the results we provide also hold for their causal versions. Thus, we use Eva and Stern's (2019) measure as a proxy for any causal interpretation of a probabilistic difference-making measure in this paper.

[10]When performing a classical intervention on H, one typically sets H to a particular value h in addition to deleting all the incoming arrows (cf. Pearl 2000). Such interventions are useful for computing the effect the particular value h would have on the body of evidence. Our goals in this paper are different. Because we want to compute $CEXP$, we need to assess the probabilistic impact of **e** on h that arises because H is a cause of **E**. To this end, we must allow for H to change its value after the intervention.

Observation 4.1.

$$CEXP(\mathbf{e}; h) > 0 \textit{ for structures (a)–(c).}$$
$$CEXP(\mathbf{e}; h) = 0 \textit{ for structures (d)–(f).}$$

This result has the following bearing on MIU and COU: Both probabilistic measures do a bad job as indicators for explanatory power. As we saw earlier, MIU is negative for the elementary causal settings (a)–(e). So while MIU correctly indicates a lack of explanatory power for structures (d) and (e), it fails to indicate positive explanatory power for structures (a)–(c). And as we saw earlier, MIU can be positive for structure (f) though h cannot have any explanatory power w.r.t. **e** in this structure. This can be seen as supporting Lange (2004) who claimed that MIU cannot account for explanation. But what about COU? Does it fare any better? The elementary cases in which h unifies according to COU but does not have any explanatory power are (d) and (e). Here, h cannot even explain one of the pieces of evidence. If H, E_1, and E_2 are connected as in (a), (b), or (c), on the other hand, unificatory power according to COU goes hand in hand with explanatory power. A somewhat special case is (b). Here h can unify the body of evidence, but cannot explain the whole body of evidence. In this case h has, strictly speaking, only explanatory relevance w.r.t. e_2, but cannot help in any way to explain e_1. Finally, COU can be positive for structure (f) though h cannot have any explanatory power w.r.t. **e** in this particular setting. Summarizing, MIU indicates too many false negatives (i.e., a lack of explanatory power where there is one), whereas COU indicates too many false positives (i.e., explanatory power where there is none). The results of this comparison are summarized in Table 2.

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

Table 2: Relationship of unificatory to explanatory power. A check mark under "Match" indicates that unificatory power (positive vs. negative) indicates explanatory power (positive vs. zero).

The lesson to be learned from these findings is this: Though Lange (2004) is right in claiming that MIU cannot account for positive explanatory relevance, while COU can, a purely probabilistic measure for COU fails to account for a lack of explanatory relevance. However, Lange explicitly refers to structural constraints when he claims that "genuinely to unify $[e_1]$ and $[e_2]$, a theory must reveal them to have some deep common explanatory basis" (p. 208). So, defining a measure for COU in purely probabilistic terms might seem inadequate

in Lange's view. Contra the measure COU, Lange might argue that not only explanation, but also unification needs to track causation. In other words: A measure for COU should first and foremost be a measure of *causal* unification. Our discussion so far makes his point explicit and shows that a measure for COU in purely probabilistic terms such as COU is inadequate. Our investigation shows, however, not only what is wrong with COU—it also prepares the grounds for modifying COU by adding a causal structural constraints. We can define the following measure for *causal* COU:

Definition 4.1 (causal common origin unification)**.**

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

[33] Like in the causal measure for explanatory power discussed before, assuming that H is decoupled from its causes by a hypothetical intervention guarantees causation tracking. Now $CCOU$ applied to our six elementary causal structures leads to the desired result:

Observation 4.2.

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

For our set of elementary structures, the behavior of $CCOU$ ordinally coincides with that of $CEXP$ and, thus, provides the result intended by (Lange 2004): Whenever a hypothesis (causally) unifies a body of evidence, it also (causally) explains this body of evidence.

In order to have a fair comparison of MIU and COU in causal settings, we introduce a causal version of MIU as well:

Definition 4.2 (causal mutual information unification)**.**

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

If we apply this measure to our six basic causal structures, we get the following as a result:

Observation 4.3.

$$CMIU(e_1, e_2; h) < 0 \textit{ for structures (a)–(c).}$$
$$CMIU(e_1, e_2; h) = 0 \textit{ for structures (d)–(f).}$$

At least for the simple causal settings considered so far, this result speaks in favor of Lange (2004): Though the causal version of MIU turns out to perform slightly better than the purely probabilistic version, the causal version of COU stays one step ahead since it is able to correctly indicate explanatory power in all settings. The results of this comparison are summarized in Table 3.

#	Model	CEXP	CMIU	Match	CCOU	Match
(a)	$E_1 \longleftarrow H \longrightarrow E_2$	> 0	< 0	$\times$	> 0	✓
(b)	$E_1 \longrightarrow H \longrightarrow E_2$	> 0	< 0	$\times$	> 0	✓
(c)	$H \longrightarrow E_1 \longrightarrow E_2$	> 0	< 0	$\times$	> 0	✓
(d)	$E_1 \longrightarrow E_2 \longrightarrow H$	$= 0$	$= 0$	✓	$= 0$	✓
(e)	$H \longleftarrow E_1 \longrightarrow E_2$	$= 0$	$= 0$	✓	$= 0$	✓
(f)	$E_1 \longrightarrow H \longleftarrow E_2$	$= 0$	$= 0$	✓	$= 0$	✓

Table 3: Relationship of causal unificatory to explanatory power.

So far, we have pushed the two views and their corresponding measures as far as we could, always with the goal in mind to make them as fit as possible to go hand in hand with explanatory relevance. But as we will see in the next section, even the sophisticated causal modifications of the original measures are ultimately doomed to failure when it comes to the question of their suitability to correctly indicate explanatory relevance in general.

5 Increasing complexity

To see why *CCOU* does not always coincide with positive explanatory power in causal settings, it suffices to slightly increase the complexity of one of the simple causal structures discussed so far. To this end, let us modify the causal structure in Figure 1(a) by introducing another causal path connecting E_1 and E_2 that does not go through H. Let us further introduce the additional causal variable X that is a common cause of E_1 and E_2 lying on this path. The structure resulting from this is depicted in Figure 3. For convenience, let us label this structure (a*).

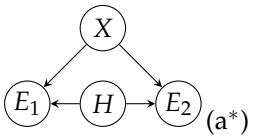

Figure 3: Causal structure resulting from Figure 1(a) by adding another causal path connecting E_1 and E_2

As before, let us assume that X is a binary variable with the two possible values $x, \bar{x}$. To stay true to our earlier assumptions, we still limit ourselves to positive probability distributions and to cases where both pieces of evidence are independent or depend positively on each other. In order to guarantee this, we make an assumption for X similar to the ones we made for H before

in section 3:

$$Pr(e_1|x) > Pr(e_1|\bar{x})$$
$$Pr(e_2|x) > Pr(e_2|\bar{x})$$

If we apply our measures $CCOU$ and $CEXP$ to the slightly more complex structure (a^*) in Figure 3, we get the following as a result:

Observation 5.1.

> $CCOU(e_1, e_2; h) > 0$ *and* $CCOU(e_1, e_2; h) < 0$ *are both compatible with* $CEXP(\mathbf{e}; h) > 0$ *for structure* (a^*).

The part saying that $CCOU(e_1, e_2; h) > 0$ is compatible with $CEXP(\mathbf{e}; h) > 0$ does not come unexpected. $CCOU$ and $CEXP$ would both be positive, for example, if $Pr(e_i|h, x) > Pr(e_i|h, \bar{x}), Pr(e_i|\bar{h}, x) > Pr(e_i|\bar{h}, \bar{x})$ (for $i \in \{1, 2\}$). The presence of each common cause would push the likelihood of the presence of each piece of evidence further. Thus, conditioning on h would render the two pieces of evidence less informative about each other. The other part saying that $CCOU(e_1, e_2; h) < 0$ is compatible with $CEXP(\mathbf{e}; h) > 0$ is a bit more surprising. $CCOU$ being negative and $CEXP$ being positive can, for example, happen if H works like a switch for whether E_1 and E_2 probabilistically depend on each other: If H takes value $\bar{h}$, then X has no probabilistic impact on E_i whatsoever, but if H takes value h, then X's taking value x increases the probability for both e_1 and e_2. For exemplary probability distributions, see the Appendix.

What this shows is that even $CCOU$ which performed so well in the simple causal structures depicted in Figure 1 does a bad job as an indicator for explanatory relevance in general. Though we went through all the hassle to increase COU's performance by further developing it in the image of the causal variant of EXP, it turns out that neither $CCOU$ nor $CMIU$ are ordinally equivalent with $CEXP$ in all causal scenarios. This result shows that the hope of philosophers such as Kitcher (1981, 1989) and Lange (2004) that unification and explanation are intimately connected turns out as an illusion, at least if unification is understood as a single hypothesis' ability to render pieces of evidence more or less informative about each other. It is neither the case that explanation can be analyzed in terms of unification nor that unified pieces of evidence must have some deep common explanatory basis such as a common causal ancestor in causal settings. Finally, what can we say about the ongoing competition between COU and MIU? Recall that one of the main objections Lange (2004) launched against Myrvold's (2003; 2017) position was that it cannot account for explanation, while COU can. Our result takes some wind out of Lange's objection by unmasking this seemingly advantage of COU over MIU: In the end, both views are rather bad in tracking explanatory relevance. [34]

6 Conclusion

In this paper, we compared two common contemporary views of unification: mutual information unification (MIU) and common origin unification (COU). We proposed a probabilistic measure for COU and compared it with Myrvold's (2003; 2017) measure for MIU. We then explored how the two probabilistic measures perform in elementary causal structures and how well they are suited to account for explanatory power. While MIU underperforms by providing disunification in too many cases, COU turned out to be way too permissive. Both probabilistic measures also failed in indicating explanatory relevance. While MIU does not correctly indicate positive explanatory power at all, also the probabilistic measure COU is not a reliable indicator for explanatory power, because in some of the basic causal settings it indicates an explanatory relation where there is none. This shows that unification spelled out in terms of probabilities alone is a bad indicator for explanatory relevance.

As a next step, we investigated the question of whether unification can track explanation by considering modifications of both measures of unification that amend probabilistic features with structural causal constraints. For this purpose, we transformed COU into the causal measure $CCOU$ and MIU into the causal measure $CMIU$ by implementing an interventionist constraint on both measures. Though we could detect an improvement for both measures, $CCOU$ clearly took the upper hand. $CCOU$ succeeded to indicate explanatory relevance in all six of the elementary causal structures considered so far, while MIU only succeeded in three of these structures. However, this victory of COU over MIU was only short-lived: Slightly increasing the complexity of the underlying causal structure causes problems also for the measure $CCOU$. Given the current debate of unification with MIU and COU as the key approaches in this field, we conclude that unification and explanation do not go hand in hand as claimed by several authors (Kitcher 1981, 1989; Lange 2004). In particular, we have shown that upholding the thesis of a "happy marriage" comes at the cost of an increased need of modification and parametrization, and, therefore, has the characteristics of a degenerative research programme.

Appendix

In this appendix we provide proofs for the observations made throughout the paper. We assume that probability distributions are non-extreme and that $Pr(e_i|h) > Pr(e_i|\bar{h})$ for $i \in 1, 2$. Due to these assumptions the denominators in Definition 2.1 are positive and, thus, $COU(e_1, e_2; h)$ and $MIU(e_1, e_2; h)$ are always defined.

Observation 2.1 follows trivially from Definitions 2.2 and 2.3, Observation 2.2 follows trivially from Observation 2.1, and Observation 4.3 follows trivially from Definitions 4.1 and 4.2 and Observation 4.2.

Observation 3.1.

$MIU(e_1, e_2; h) < 0 < COU(e_1, e_2; h)$ for structures (a)–(e).

$MIU(e_1, e_2; h), COU(e_1, e_2; h)$ are lesser, equal, or greater 0 for structure (f).

Proof.

Structure (a): From Equation 4 applied to structure (a) it follows that e_1 and e_2 are positively dependent, meaning that $Pr(e_1, e_2) > Pr(e_1) \cdot Pr(e_2)$. From Definition 2.1 it then follows that $I(e_1, e_2) > 0$. From Equation 4 applied to structure (a) it also follows that $Pr(e_1, e_2|h) = Pr(e_1|h) \cdot Pr(e_2|h)$. From Definition 2.1 it then follows that $I(e_1, e_2|h) = 0$. Because $I(e_1, e_2) > I(e_1, e_2|h)$, it follows from Definitions 2.2 and 2.3 that $MIU(e_1, e_2; h) < 0 < COU(e_1, e_2; h)$.

Structure (b): Since structure (b) is probabilistically indistinguishable from structure (a),[11] the purely probabilistic measures MIU and COU perform exactly as in structure (a). Thus, $MIU(e_1, e_2; h) < 0 < COU(e_1, e_2; h)$.

Structure (c): Since e_1 and e_2 are both positively depend on h and the structure is $H \longrightarrow E_1 \longrightarrow E_2$, Equation 4 rules that also e_2 depends positively on e_1 unconditionally. Because of this, $I(e_1, e_2)$ will be positive. For the same reasons, e_1 and e_2 will also be positively dependent conditional on h, except $Pr(e_1|h) = 1$, in which case e_1 and e_2 will be independent conditional on h. Hence, $I(e_1, e_2; h) \geq 0$. Now if we conditionalize on h, the probabilities for each piece of evidence as well as for their conjunction are pushed upwards such that $Pr(e_1, e_2)$ compared to $Pr(e_1) \cdot Pr(e_2)$ becomes smaller. But this means that the two pieces of evidence become less informative about each other and, thus, that $I(e_1, e_2) > I(e_1, e_2|h)$. It then follows from Definitions 2.2 and 2.3 that $MIU(e_1, e_2; h) < 0 < COU(e_1, e_2; h)$.

Structures (d) and (e): Since these structures are probabilistically indistinguishable from structure (c), the purely probabilistic measures MIU and COU perform exactly as in structure (c). Thus, $MIU(e_1, e_2; h) < 0 < COU(e_1, e_2; h)$.

Structure (f): To show that both MIU and COU each can be lesser, equal, or grater 0 we provide three exemplary probability distributions Pr^1, Pr^2, Pr^3. [35] All three distributions are positive distributions conforming to the Markov factorization. All three also satisfy the constraint that $Pr(e_i|h) > Pr(e_i|\bar{h})$ for $i \in \{1, 2\}$.

#	$Pr(e_1)$	$Pr(e_2)$	$Pr(h\|e_1, e_2)$	$Pr(h\|e_1, \bar{e}_2)$	$Pr(h\|\bar{e}_1, e_2)$	$Pr(h\|\bar{e}_1, \bar{e}_2)$
Pr^1	0.5	0.5	1	1	1	0
Pr^2	0.5	0.5	1	0	0	0
Pr^3	0.5	0.5	1	0.5	0.5	0.5

[11] This means that any distribution conforming to Equation 4 applied to structure (a) will also conform to Equation 4 applied to structure (b), and vice versa.

From these parameters we get (numbers are rounded to three digits):

$$MIU = -0.415 < 0 < 0.415 = COU \text{ for } Pr^1$$
$$MIU = 0 = COU \text{ for } Pr^2$$
$$MIU = 0.152 > 0 > 0.152 = COU \text{ for } Pr^3$$

□

Observation 4.1.

$$CEXP(\mathbf{e}; h) > 0 \textit{ for structures (a)–(c).}$$
$$CEXP(\mathbf{e}; h) = 0 \textit{ for structures (d)–(f).}$$

Proof.

Structure (a): For this case, the post intervention distribution is identical to the pre intervention distribution. Now Equation 4 implies:

$$Pr(e_1, e_2|h) = Pr(e_1|h) \cdot Pr(e_2|h)$$
$$Pr(e_1, e_2) = Pr(e_1|h) \cdot Pr(e_2|h) \cdot Pr(h) + Pr(e_1|\bar{h}) \cdot Pr(e_2|\bar{h}) \cdot Pr(\bar{h})$$

The latter is a weighted average. From $Pr(e_i|h) > Pr(e_i|\bar{h})$ (for $i \in \{1, 2\}$) we know that the upper bound of the weighted average is$Pr(e_1|h) \cdot Pr(e_2|h)$ and that its lower bound is $Pr(e_1|\bar{h}) \cdot Pr(e_2|\bar{h})$. Since we only consider positive distributions, we also know that the weight $Pr(h)$ is not 1. From this and the equations above it follows that $Pr(e_1, e_2|h) > Pr(e_1, e_2)$, meaning that h and $\mathbf{e}$ are positively dependent. Thus, $CEXP$ is positive.

Structure (b): Equation 4 applied to the post intervention graph gives us:

$$Pr(e_1, e_2|h) = Pr(e_1) \cdot Pr(e_2|h)$$
$$Pr(e_1, e_2) = Pr(e_1) \cdot Pr(e_2|h) \cdot Pr(h) + Pr(e_1) \cdot Pr(e_2|\bar{h}) \cdot Pr(\bar{h})$$

Again, the latter is a weighted average and from $Pr(e_2|h) > Pr(e_2|\bar{h})$ we know that $Pr(e_1) \cdot Pr(e_2|h)$ is its upper bound and $Pr(e_1) \cdot Pr(e_2|\bar{h})$ its lower bound. Again, by assumption the weight $Pr(h)$ cannot be 1 and, hence, $Pr(e_1, e_2|h) > Pr(e_1, e_2)$ follows. Thus, h and $\mathbf{e}$ are positively dependent and $CEXP$ is positive.

Structure (c): For this case, the post intervention distribution is again identical to the pre intervention distribution. Equation 4 implies:

$$Pr(e_1, e_2|h) = Pr(e_1|h) \cdot Pr(e_2|e_1)$$
$$Pr(e_1, e_2) = Pr(e_1|h) \cdot Pr(e_2|e_1) \cdot Pr(h) + Pr(e_1|\bar{h}) \cdot Pr(e_2|e_1) \cdot Pr(\bar{h})$$

Again, the latter is a weighted average and from $Pr(e_1|h) > Pr(e_1|\bar{h})$ we know that $Pr(e_1|h) \cdot Pr(e_2|e_1)$ is its upper bound and $Pr(e_1|\bar{h}) \cdot Pr(e_2|e_1)$ its lower

bound. Since the weight $Pr(h)$ cannot be 1 by assumption, it follows that $Pr(e_1, e_2|h) > Pr(e_1, e_2)$. Thus, h and $\mathbf{e}$ are positively dependent and $CEXP$ is positive.

Structures (d)–(f): For all three post intervention structures, Equation 4 rules that h is probabilistically independent of $\mathbf{e}$. Because of this and since we consider only positive distributions, the numerator in Equation 7 determining $CEXP$ will be 0 and the denominator will be greater than 0. Hence, $CEXP$ will be 0. □

Observation 4.2.

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

Proof.

Structures (a) and (c): In both cases the intervention does not break any arrows. Hence, the post intervention distribution will be identical to the pre intervention distributions and $CCOU$ and COU coincide. And since we already know that $COU > 0$, we also know that $CCOU > 0$.

Structure (b): We already know that e_1 and e_2 are positively dependent in the pre intervention distribution. Thus, $I(e_1, e_2) > 0$. For the post intervention, however, Equation 4 rules that e_1 and e_2 are independent conditional on h. Hence, $I(e_1, e_2|\hat{h}) = 0$. From Definition 4.1 it then follows that $CCOU(e_1, e_2; h) > 0$.

Structures (d) and (e): Equation 4 applied to the two structures' pre intervention graphs gives us:

$$\begin{aligned} Pr(e_1) &= Pr(e_1) \\ Pr(e_2) &= Pr(e_2|e_1) \cdot Pr(e_1) + Pr(e_2|\bar{e}_1) \cdot Pr(\bar{e}_1) \\ Pr(e_1, e_2) &= Pr(e_2|e_1) \cdot Pr(e_1) \end{aligned}$$

[36] These are the probabilities that determine $I(e_1, e_2)$. Applied to the post intervention graphs, Equation 4 gives us:

$$\begin{aligned} Pr(e_1|h) &= Pr(e_1) \\ Pr(e_2|h) &= Pr(e_2|e_1) \cdot Pr(e_1) + Pr(e_2|\bar{e}_1) \cdot Pr(\bar{e}_1) \\ Pr(e_1, e_2|h) &= Pr(e_2|e_1) \cdot Pr(e_1) \end{aligned}$$

These are the probabilities that determine $I(e_1, e_2|\hat{h})$. It then follows from Definition 2.1 that $I(e_1, e_2) = I(e_1, e_2|\hat{h})$ and, thus from Definition 4.1 that $CCOU(e_1, e_2; h) = 0$.

Structure (f): Equation 4 applied to the pre intervention graph tells us that e_1 and e_2 are independent. Hence, $I(e_1, e_2) = 0$. Applied to the post intervention graph the same equation implies that e_1 and e_2 are independent conditional on h and, thus, that $I(e_1, e_2|\hat{h}) = 0$ holds. It follows from Definition 4.1 that $CCOU(e_1, e_2; h) = 0$. □

Observation 5.1.

> $CCOU(e_1, e_2; h) > 0$ *and* $CCOU(e_1, e_2; h) < 0$ *are both compatible with* $CEXP(\mathbf{e}; h) > 0$ *for structure (a*).*

Proof.

To show that $CCOU(e_1, e_2; h) > 0$ and $CCOU(e_1, e_2; h) < 0$ are both compatible with $CEXP(\mathbf{e}; h) > 0$ for structure (a*), we provide two exemplary probability distributions Pr^1, Pr^2. Both distributions are positive distributions conforming to the Markov factorization. Both also satisfy the constraints that $Pr(e_i|h) > Pr(e_i|\bar{h})$ and $Pr(e_i|x) > Pr(e_i|\bar{x})$ for $i \in \{1, 2\}$.

#	$Pr(h)$	$Pr(x)$	$Pr(e_i\|h, x)$	$Pr(e_i\|h, \bar{x})$	$Pr(e_i\|\bar{h}, x)$	$Pr(e_i\|\bar{h}, \bar{x})$
Pr^1	0.5	0.5	0.8	0.5	0.5	0.2
Pr^2	0.5	0.5	0.8	0.5	0.5	0.5

From these parameters we get (numbers are rounded to three digits):

$$CCOU(e_1, e_2; h) = 0.164 \text{ and } CEXP(\mathbf{e}; h) = 0.314 \text{ for } Pr^1$$
$$CCOU(e_1, e_2; h) = -0.003 \text{ and } CEXP(\mathbf{e}; h) = 0.202 \text{ for } Pr^2$$

□

Funding: Research on this paper was funded by DFG, Research Group (FOR 2495), Project A2.1.

Acknowledgements: We would like to thank Jean Baccelli, Andrew Buskell, Samuel Fletcher, Marie Gueguen, Leah Henderson, Paola Hernández-Chávez, Edouard Machery, Adina Roskies, Sander Verhaegh, and an anonymous reviewer for helpful comments on an earlier draft.

References

Bovens, Luc and Hartmann, Stephan (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Eva, Benjamin and Stern, Reuben (2019). "Causal Explanatory Power". In: *The British Journal for the Philosophy of Science* 70.4, pp. 1029–1050. DOI: 10.1093/bjps/axy012.

Friedman, Michael (1974). "Explanation and Scientific Understanding". In: *The Journal of Philosophy* 71.1, pp. 5–19. DOI: 10.2307/2024924.

Gebharter, Alexander (2017a). "Causal Exclusion and Causal Bayes Nets". In: *Philosophy and Phenomenological Research* 95.2, pp. 353–375. DOI: 10.1111/phpr.12247.

— (2017b-11). "Uncovering Constitutive Relevance Relations in Mechanisms". In: *Philosophical Studies* 174.11, pp. 2645–2666. DOI: 10.1007/s11098-016-0803-3.

— (2022). "A Causal Bayes Net Analysis of Glennan's Mechanistic Account of Higher-Level Causation (and Some Consequences)". In: *The British Journal for the Philosophy of Science* 73.1, pp. 185–210. DOI: 10.1093/bjps/axz034.

Gebharter, Alexander and Eronen, Markus I. (forthcoming). "Quantifying Proportionality and the Limits of Higher-Level Causation and Explanation". In: *British Journal for the Philosophy of Science*. DOI: doi:10.1086/714818.

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

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.

Kneale, Matthew (1949). *Probability and Induction*. Oxford: Oxford University Press.

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

McGrew, Timothy (2003-12). "Confirmation, Heuristics, and Explanatory Reasoning". In: *The British Journal for the Philosophy of Science* 54.4, pp. 553–567. DOI: 10.1093/bjps/54.4.553.

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.

Niiniluoto, Ilkka (2016). "Unification and Confirmation". In: *Theoria. An International Journal for Theory, History and Foundations of Science* 31.1, pp. 107–123. DOI: 10.1387/theoria.13084.

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

Reichenbach, Hans (1956). *The Direction of Time*. Berkeley: University of California Press.

Schaffer, Jonathan (2016-01). "Grounding in the Image of Causation". In: *Philosophical Studies* 173.1, pp. 49–100. DOI: 10.1007/s11098-014-0438-1.

Schupbach, Jonah N. (2005). "On a Bayesian Analysis of the Virtue of Unification". In: *Philosophy of Science* 72.4, pp. 594–607. DOI: 10.1086/505186.

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.

Sober, Elliott (1988). *Reconstructing the Past. Parsimony, Evolution, and Inference*. Cambridge, Massachusetts: Bradford.

Spirtes, Peter, Glymour, Clark, and Scheines, Richard (1993). *Causation, Prediction, and Search*. Dordrecht: Springer.

Sprenger, Jan (2018). "Foundations of a Probabilistic Theory of Causal Strength". In: *The Philosophical Review* 127.3, pp. 371–398. DOI: 10.1215/00318108-6718797.

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.

Whewell, William (1840/2014). *The Philosophy of the Inductive Sciences: Founded upon their History*. Cambridge Library Collection - Philosophy. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139644662.

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.