Facing asymmetry

Towards a quasi-interventionist counterfactual theory of causal and non-causal explanation

Ina Jäntgen

July 07, 2021

Research Master's Thesis (30 EC)

Supervisor: Dr Conrad Heilmann
Advisor: Dr Christopher Clarke

30000 words

Erasmus Institute for Philosophy and Economics
Erasmus University Rotterdam

Abstract

We typically explain phenomena by citing their causes. Yet, we also explain phenomena without citing their causes, for instance by referring instead to mathematical theorems. To account for the explanatory power of such non-causal explanations, several philosophers aim to free the interventionist theory of causal explanation from its causal component: interventions. The resulting counterfactual theory of causal and non-causal explanation claims that X explains Y iff Y counterfactually depends on changes in X. While initially attractive, such a counterfactual theory faces a familiar problem: explanations are typically asymmetric while not all counterfactual dependence is. The challenge of asymmetry demands to qualify a counterfactual theory such as to account for the asymmetry of explanations. First, I reject two potential solutions to this challenge: amendments of a counterfactual theory using Lewisian semantics or two inference schemes that tell us when to infer that X explains Y and not vice versa. Second, I develop a quasi-interventionist version of a counterfactual theory apt to deal with explanatory asymmetry, in two steps. I identify a class of explanations that a counterfactual theory without further specification already claims to be asymmetric. And I develop the notion of a quasi-intervention to account for the asymmetry of most remaining explanations. The resulting quasi-interventionist theory claims that X explains Y iff Y counterfactually depends on changes in X brought about by a quasi-intervention. Finally, I suggest accepting the symmetry this quasi-interventionist theory entails for a particular class of non-causal explanations.

Acknowledgements

I would like to thank my supervisor Conrad Heilmann for his invaluable support and especially his encouragement to develop my ideas further. I also thank Christopher Clarke for his insightful comments on an earlier draft. Moreover, thank you to Ahmed AlJuhany, Lennart Ackermans, Maximilian Gasser, Ruth Hinz, Lizzy Miller, Marie-Aimée Salopiata, Chiara Stenico and Erica Yu for valuable discussion, comments and emotional support. Finally, thank you to the whole EIPE community for creating such an intellectually stimulating and supportive environment! It has been a great joy to be a part of it.

Contents

1 Introduction

Many philosophers agree that scientists and laypeople typically refer to causes or causal mechanisms when explaining phenomena (see e.g., Craver, 2007; Strevens, 2008; Woodward, 2003). Recently, however, some philosophers have pointed to cases in which we seem to explain without citing causes or causal mechanisms (see Reutlinger, 2017a; Reutlinger & Saatsi, 2018). The explanatory power of these non-causal explanations cannot be accounted for by prominent theories of causal explanation. This motivates the need to develop alternative theories of explanation that are capable of covering also non-causal explanations.[1]

In recent years, several authors responded to this challenge by extending the interventionist theory of causal explanation (Hitchcock & Woodward, 2003a, 2003b; Woodward, 2003) to non-causal ones (see Jansson & Saatsi, 2019; Povich, 2018; Saatsi & Pexton, 2013; Woodward, 2018). According to the interventionist theory, explaining is a matter of showing how an explained phenomenon (the explanandum) counterfactually depends on changes in the explaining phenomenon (the explanans) that are brought about by interventions. Here, an intervention is (roughly) a surgical manipulation of the explanans. Interventionists hold that an explanans causes an explanandum iff intervening on the explanans would change the explanandum. It follows that interventionist explanations are causal explanations. To extend the theory to non-causal explanations some authors propose to dispense with its causal component: interventions. Explaining, they claim, is simply a matter of showing how the explanandum counterfactually depends on changes in the explanans. I call this view a counterfactual theory of causal and non-causal explanation.

At first glance, the strategy to extend the interventionist theory to non-causal explanations is well-motivated: The interventionist theory has proven helpful in understanding many aspects of explanatory practice, for instance how explanations differ in quality. Moreover, intuitively, explaining a phenomenon has something to do with showing how it would be different under different circumstances. Furthermore, a counterfactual theory of causal and non-causal explanation

[1] Instead of accepting this need, one could deny that alleged examples of non-causal explanations cannot be covered by a theory of causal explanation (see e.g., Skow, 2014). I will not discuss any such view in this thesis.

would be a unified theory of explanation. Ceteris paribus, a unified theory seems preferable over a non-unified one (Reutlinger, 2018, p. 78).[2]

However, a counterfactual theory of causal and non-causal explanation faces a familiar challenge: to account for the asymmetry explanations exhibit. Typically, X explains Y but *not* vice versa. Many readers will be familiar with the flagpole case: The position of the sun and a flagpole's height explain its shadow's length. But one cannot appeal to the flagpole's shadow's length in explaining either its height or the position of the sun. A theory of explanation, so the traditional consensus goes, ought to account for this asymmetric feature of explanations.

A counterfactual theory struggles with accounting for explanatory asymmetry: According to this theory, an explanans X explains an explanandum Y iff Y counterfactually depends on changes in X. But counterfactual dependence is not per se asymmetric. Y can counterfactually depend on changes in X *and vice versa*. Had the flagpole's height been different (and the position of the sun unchanged), its shadow's length would have been different. And had the flagpole's shadow's length been different (and the position of the sun unchanged), the flagpole's height would have been different. This is so despite the flagpole's height and the sun explaining its shadow's length, and not the sun and its shadow's length its height. Absent further specifications a counterfactual theory fails to identify this asymmetry. Hence, proponents of a counterfactual theory are challenged to qualify their theory further such that it yields explanatory asymmetry. I call this the challenge of asymmetry.

This thesis discusses how proponents of a counterfactual theory of causal and non-causal explanation could solve the challenge of asymmetry. It, therefore, aims to contribute to the development of such a theory. The main contributions are the following:

1. I distinguish between the challenge of asymmetry and the revised challenge of (a)symmetry that demands to identify some non-causal explanations as symmetric.
2. I reject versions of a counterfactual theory of explanation relying on Lewisian semantics or two inference schemes as solutions to the challenge of asymmetry and the revised challenge of (a)symmetry.

[2] For example, a unified theory of explanation avoids a pluralist dilemma (Pincock, 2018, p. 41f): Either different kinds of explanations have nothing in common or they do. If they do, then this suggests a unified theory. If they do not, then it is hard to explain why they are all explanations.

3. I provide a quasi-interventionist theory of explanation as a promising solution to the revised challenge of (a)symmetry.

This thesis is structured as follows:

Part I introduces the challenge of asymmetry for a counterfactual theory of causal and non-causal explanation. In Chapter 2, I define a counterfactual theory of causal and non-causal explanation, a view common amongst several authors in the recent literature on non-causal explanations. In Chapter 3, I then introduce the challenge of asymmetry: it requires qualifying the counterfactual theory such as to yield explanatory asymmetry. I distinguish this challenge from a revised one that demands to also account for some symmetric non-causal explanations. This distinction allows us to specify when a theory fails to account for explanatory asymmetry and when it successfully identifies a symmetric explanation.

Part II argues against two potential solutions to the challenge of asymmetry. In Chapter 4, I argue that it is unclear how supplementing a counterfactual theory with Lewisian semantics would solve the challenge of asymmetry. For counterfactuals involving events, Lewisian semantics yields asymmetry of counterfactual dependence. In such cases, a counterfactual theory with Lewisian semantics yields explanatory asymmetry. However, I argue that it is unclear how Lewisian semantics would establish asymmetry of counterfactual dependence for counterfactuals that do not involve events. And in many non-causal explanations, the explanantia and/ or explananda are not events. Consequently, it is unclear how a counterfactual theory with Lewisian semantics would yield explanatory asymmetry in these non-causal explanations. In Chapter 5, I then discuss two principles proposed to address asymmetry in non-causal explanations, Woodward’s (2018, 2020) double explanation principle and Lange’s (2019) modal fact principle. I show that both can be understood as inference schemes, i.e. as telling us when to infer that some phenomenon explains another and not vice versa. I distinguish between two versions of a counterfactual theory, each using the inference schemes in different ways to address the challenge of asymmetry. I argue that neither of these versions solves this challenge.

Part III provides a quasi-interventionist version of a counterfactual theory, accounting for the asymmetry of many causal and non-causal explanations and the symmetry of some non-causal explanations. In Chapter 6, I narrow the scope of the challenge of asymmetry. I argue that if an explanans X and an explanandum Y relate mathematically and with a so-called many-to-one relation between their actual values then an unqualified counterfactual theory already claims that (if at all) X explains Y but not vice versa. Therefore, the challenge of asymmetry pertains only to those explanations not possessing these features. In Chapter 7, I provide a quasi-interventionist version of a counterfactual theory, accounting for the asymmetry in many of these remaining explanations. Combining a proposal by Woodward (2018) and the notion of an intervention, I develop the notion of a quasi-intervention. The resulting quasi-interventionist theory claims that X explains Y iff Y counterfactually depends on changes in X brought about by a quasi-intervention. In Chapter 8, I then propose that this quasi-interventionist theory is a promising solution to the revised challenge instead of the challenge of asymmetry. I show that it identifies a specific class of explanations (namely mathematical explanations with one-to-one relations) as symmetric. I provide initial motivation to accept this result and discuss open questions for further research on the quasi-interventionist theory as a solution to the revised challenge. I also show that amendments of a counterfactual theory using Lewisian semantics or the inference schemes do not solve the revised challenge.

In Chapter 9, I summarize the main contributions of this thesis to the literature on explanation as providing a distinction between the challenge of asymmetry and the revised challenge of (a)symmetry, a rejection of versions of a counterfactual theory relying on Lewisian semantics or two inference schemes as solutions to either challenge and a quasi-interventionist theory of explanation as a promising solution to the revised challenge. I also discuss several directions for future research emerging from this thesis.

Part I The challenge of asymmetry for a counterfactual theory of explanation

In this part, I introduce the challenge of asymmetry for a counterfactual theory of explanation. In Chapter 2, I define a counterfactual theory of causal and non-causal explanation, a view common amongst several authors in the literature on non-causal explanations. In Chapter 3, I then introduce the challenge of asymmetry and distinguish it from the revised challenge of (a)symmetry.

2 A counterfactual theory of causal and non-causal explanation

In recent years, there have been various attempts to understand the explanatory power of non-causal explanations as located in the counterfactual information they provide. In this chapter, I define one potential counterfactual theory of causal and non-causal explanation. It extends an interventionist theory of causal explanation to non-causal explanations by dispensing with the notion of an intervention. To this end, I first introduce the interventionist theory (Section 2.1). Then, I review recent work on non-causal explanations (Section 2.2) before defining the counterfactual theory these authors strive after (Section 2.3).

2.1 A primer on the interventionist theory of causal explanation

The interventionist theory of causal explanation

According to the interventionist theory, explaining is a matter of showing how the explanandum counterfactually depends on particular types of changes in the explanans (see Woodward, 2003, Chapter 5; and also Hitchcock & Woodward, 2003a). In Woodward's words, it is a matter of answering "*what-if-things-had-been-different* or *w-questions*" (Woodward, 2003, p. 191) about a to-be explained phenomenon. To illustrate this idea, let us look at a well-known example, the flagpole explanation (illustrated in Figure 1):[3]

> ***Flagpole***[4]
> *You see a flagpole with a height x_1^1. You notice that the angle of the sun above the horizon is x_2^1 and that the flagpole's shadow has a length of y^1. Why is the flagpole's shadow's length y^1?*

[3] The original version of this example is from (Bromberger, 1966).
[4] In this thesis, I use the following notation: Upper-case letters denote variables with subscripts denoting different variables. Lower-case letters denote the values of variables with x_j^i denoting the i-th value of X_j.

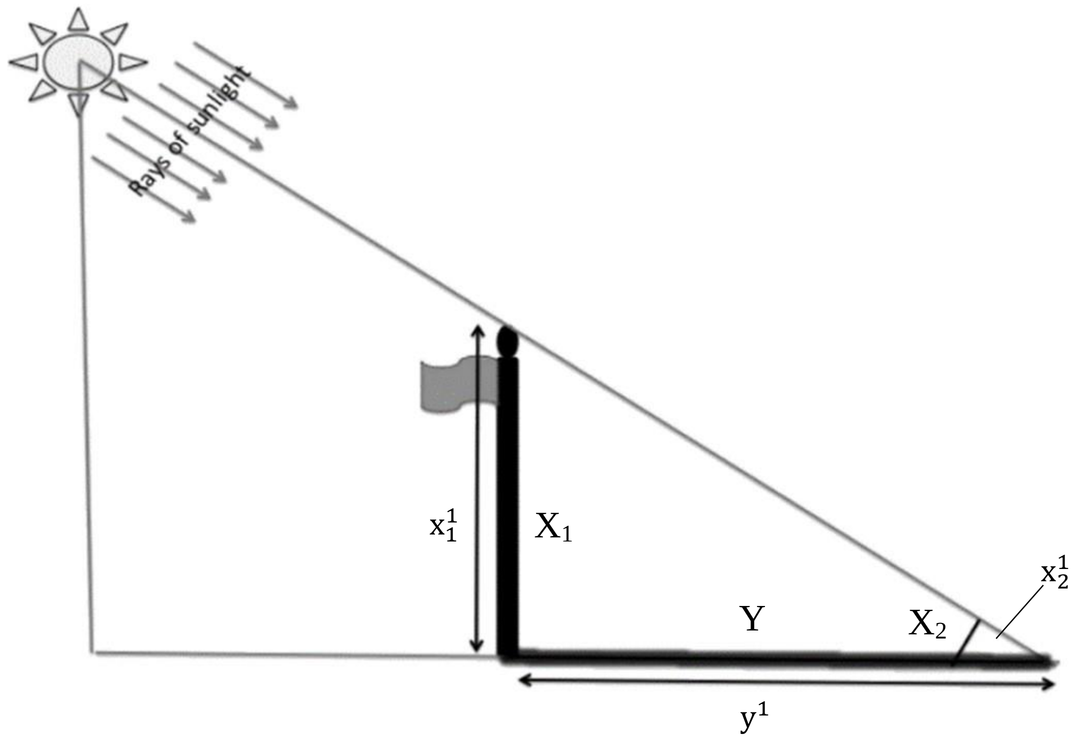


Figure 1: *The flagpole explanation* (Jansson, 2015, p. 584; notation changed).
X_1 represents the flagpole's height with x_1^1 denoting its actual height. Y represents the flagpole's shadow's length with y^1 denoting its actual length. X_2 represents the angle of sunrays with x_2^1 denoting the actual angle.

How to causally explain the flagpole's shadow's length y^1 according to the interventionist theory?

In the interventionist theory, both the explanans and the explanandum are represented with variables (Woodward, 2003, p. 39f). Variables represent properties; different values represent different instantiations of those properties. For instance, the flagpole's height can be represented by a variable X_1 which can take on different values x_1^1, $x_1^2 \ldots x_1^n$. In the present example, X_1 takes on the value x_1^1, the height of the flagpole at stake here. Similarly, Y represents the flagpole's shadow's length with y^1, $y^2 \ldots y^n$ representing different lengths.

According to interventionist theory, explaining why the flagpole's shadow's length is y^1 then means to explain why its length is y^1 rather than $y^2 \ldots y^n$ (Woodward, 2003, p. 145f). In other words, the explanandum is understood contrastively here. The explanans has two elements:

The first element of the explanans is a specification which values the explanans variables take on (Woodward, 2003, p. 203). Here, this means we specify the height of the flagpole (x_1^1) and at which angle the sun stands above the horizon (x_2^1).

The second element is the relationship between the explanandum and the explanans which must be invariant under interventions (Woodward, 2003, Chapter 5 and 6, especially p. 203). Often, scientists use generalizations to describe these relationships (Woodward, 2018, p. 120):[5]

$$\text{(G1) flagpole's shadow's length Y} = \frac{\text{height of flagpole X1}}{\tan\text{ (angle of sunrays X2)}}$$

G1 relates the flagpole's shadow's length to the flagpole's height and the angle of sunrays. To figure in an explanation, G1 must remain invariant (i.e. stable) under at least some possible interventions changing the flagpole's height and the angle of sunrays. There are two important concepts here: interventions and invariance.

First, an *intervention* is a manipulation of an explanans such that the explanandum changes only due to the change in the explanans, if at all (Woodward, 2003, p. 98). For instance, an intervention on the flagpole's height with respect to its shadow's length is a manipulation of the flagpole's height such that its shadow's length only changes due to the change in the flagpole's height, if at all. Cutting the flagpole would be such an intervention. For G1 to be explanatory, it suffices that an intervention "at least be logically possible and well-defined" (Woodward, 2003, p. 128) such that we can make sense of what would happen under such an intervention.

Second, to be explanatory, the relationship which G1 represents must be *invariant* under interventions on the flagpole's height and the angle of sunrays (Woodward, 2003, Chapter 6). It must hold (at least for some interventions) that if the angle of sunrays and/ or the flagpole's height would be changed by an intervention, then its shadow's length would change as described by G1. If so, then G1 expresses a pattern of change-relating counterfactual dependence between the flagpole's shadow's length, its height, and the angle of sunrays. Note that invariant relationships between variables are invariant as an empirical matter here (Woodward, 2018, p. 121). They are not invariant for conceptual, logical, or mathematical reasons. When turning to non-causal explanations, we will encounter relations claimed to be invariant for mathematical reasons.

According to the interventionist theory, the two elements of the explanans, the invariant relation and the specification of the explanans values, then explain the explanandum iff the following conditions hold (Woodward, 2003, p. 203, 2018, p. 122):

[5] I take G1 from (Jansson, 2015, p. 584).

> Exp_1: The explanans X (where X may consist of several variables $X_1 \ldots X_n$) and the explanandum Y are (approximately) true.

It should be true that the flagpole's shadow has a length of y^1, the flagpole a height of x_1^1 and the sunrays an angle of x_2^1 above the horizon.

> Exp_2: (The generalization describing)[6] the relationship between X and Y correctly describes the actual value y of Y under an intervention setting X to its actual value x.

G1 should correctly describe that the flagpole's shadow's length is y^1 under interventions setting the flagpole's height to x_1^1 and the angle of sunrays to x_2^1.

> Exp_3: (The generalization describing) the relationship between X and Y correctly describes that Y would change from y' to y'' if X would be changed by an intervention from x' to x'', for at least some possible interventions.

G1 should correctly describe how the flagpole's shadow's length would change if one were to intervene to change the flagpole's height and/or the angle of sunrays for at least some possible interventions.

According to the interventionist theory, if and only if these three conditions hold then one explains why the flagpole's shadow's length is y^1 by citing G1, the flagpole's height x_1^1 and the angle of sunrays x_2^1. In doing so, one shows how the flagpole's shadow's length counterfactually depends on changes in the flagpole's height and the angle of sunrays. One answers what-if-things-had-been-different-questions about the flagpole's shadow's length.

Note that the interventionist theory is an ontic theory of explanation: "It is physical dependency relations, as expressed by the relevant counterfactuals about what would happen under interventions, that are primary or fundamental in causal explanation" (Woodward, 2003, p. 202). Here, explaining is a matter of providing information about ontic dependencies between explanandum and explanans (see also Salmon, 1984, 1989, p. 120f). I return to this aspect in Section 2.2.

[6] In the original formulation of the interventionist theory a generalization is a necessary component of the explanans (Woodward, 2003, p. 203), but in later work this requirement is relaxed (Woodward, 2018, p. 120). For this reason, I talk interchangeably about a generalization describing how X and Y relate and the relationship between them.

Interventions

Having described how an explanation works according to the interventionist theory, I now explain the notion of an intervention. This is important as I refer to it in Part III. Woodward defines an intervention as follows (Woodward, 2003, p. 98; notation changed):

> "**(IN)** I's assuming some value $I=z^i$ is an intervention on X with respect to Y if and only if
>
> (i) I is an intervention variable for X with respect to Y and
>
> (ii) $I=z^i$ is an actual cause of the value taken by X."

Clause (ii) says that a particular intervention on X should cause X to take on some particular value. For instance, cutting the flagpole in half should cause the flagpole to have a particular height. For our purposes, clause (i) is more important. It refers to the notion of an intervention variable, defined as follows:

> "(**IV**) I is an intervention variable for X with respect to Y if and only if I meets the following conditions:
>
> I1. I causes X.
>
> I2. I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.
>
> I3. Any directed path from I to Y goes through X. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any that are built into the I-X-Y connection itself; that is, except for (a) any causes of Y that are effects of X (i.e. variables that are causally between X and Y) and (b) any causes of Y that are between I and X and have no effect on Y independently of X.
>
> I4. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X." (Woodward, 2003, p. 98)

As illustrated in Figure 2, conditions I1 to I4 ensure that an intervention on X with respect to Y changes Y only due to the change in X, if at all. To illustrate, recall that cutting the flagpole is an intervention on the flagpole's height with respect to its shadow's length. Indeed, cutting the flagpole satisfies I1 to I4. It does not change the flagpole's shadow's length directly or via some other cause of this length (I3). It is not correlated with other causes of the flagpole's shadow's length such as the angle of sunrays (I4). And it causes the flagpole's height (I1) making it independent of its other causes such as the intentions of those building the flagpole (I2).

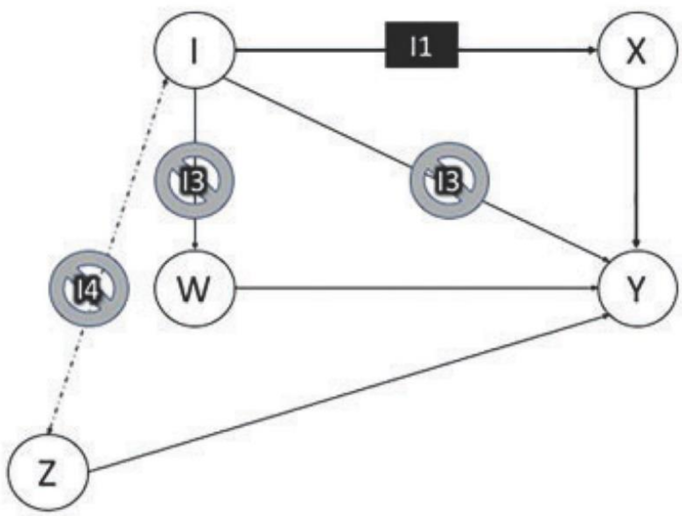


Figure 2: *The conditions for an intervention variable* (Khalifa et al., 2020, p. 1444).
Causal relationships are represented by solid arrows, correlations by dotted ones. The black box denotes I1 and the strikethroughs I3 and I4. I2 is not represented.

Importantly, interventions are designed to capture causal relationships (Woodward, 2003, p. 45ff). X causes Y iff an intervention on X would change Y. The flagpole's height causes its shadow's length iff an intervention on the flagpole's height would change its shadow's length. This means relationships between an explanandum and an explanans that are invariant *under interventions* are *causal*. In that sense, the interventionist theory demanding to cite such relationships in explanations is a theory of causal explanation.

To summarize: According to the interventionist theory, a causal explanation shows how changes in an explanandum counterfactually depend on changes in the explanans (*w-question requirement*) brought about by interventions on the explanans (*intervention requirement*). Moreover, the relationship between the explanans and the explanandum must be invariant under these interventions (*invariance requirement*).

2.2 Extending the interventionist theory to non-causal explanations

As a theory of causal explanation, the interventionist theory cannot be applied to cases in which one wants to argue that the explanans and the explanandum are not causally related. In this section, I review four attempts to still rely on elements in the interventionist theory to capture the explanatory power of non-causal explanations. All of these attempts have something in common: The authors uphold the w-question requirement while dispensing with interventions.

2.2.1 Woodward (2018) on two varieties of non-causal explanations

Woodward (2018) aims to extend his interventionist theory of causal explanation to cover two varieties of non-causal explanations.[7]

First variety: Non-causal explanations with empirical relations

Woodward argues that some explanations provide answers to w-questions although there are no possible interventions on the explanans (Woodward, 2018, pp. 122–126). Here is his example:

> ***Stability of planetary orbits***
> *"given assumptions about what the gravitational potential would be like in an n-dimensional space (...), Newton's laws of motion and a certain conception of what the stability of planetary orbits consists in, it follows that no stable planetary orbits are possible for spaces of dimensions n≥4. Obviously orbits of any sort are impossible in a space for which n=1, and it can be argued that n=2 can be ruled out on other grounds, leaving n=3 as the only remaining possibility for stable orbits."* (Woodward, 2018, p. 122f)

Woodward argues that this is a non-causal explanation of the possibility of stable planetary orbits in our three-dimensional world (Woodward, 2018, pp. 122–123):

First, Woodward claims, it is impossible to intervene on the dimensionality of space. One cannot make sense of physically manipulating the dimensionality of space. Thus, taking the interventionist

[7] Woodward also discusses a third variety (Woodward, 2018, pp. 128–136). According to him, non-causal explanations of this third variety explain by showing how the explanandum would *not* have been different under changes in the explanans, breaking the w-question requirement. This third variety will be left aside as it is not part of the common view amongst the authors that this thesis deals with.

criterion to demarcate causal from non-causal explanations, Woodward believes this explanation could not be causal.

Second, however, Woodward argues that the possibility of stable planetary orbits still counterfactually depends on changes in the dimensionality of space. The mathematical derivation of this possibility supports the following counterfactual: If the dimensionality of space had changed, so would have the possibility of stable planetary orbits. Woodward believes the change of space's dimensionality involved in this counterfactual dependence is mathematical rather than physical. One calculates the implications of changing space's dimensionality in the mathematical derivation rather than physically. Still, for Woodward, the derivation provides counterfactual information and is thus a non-causal explanation.

Finally, Woodward implies that the relationship between the explanandum (the possibility of stable planetary orbits) and the explanans (the dimensionality of space) is invariant for empirical reasons here (Woodward, 2018, p. 126). For him, this relationship involves empirical assumptions, e.g., about the form of gravitational potential and laws of motion. As a result, he views this relationship as invariant for empirical reasons. This holds even though, he thinks, the counterfactual change of space's dimensionality is mathematical rather than physical here.

Second variety: Non-causal explanations with mathematical relations

Woodward claims that some non-causal explanations provide answers to w-questions, even if there are plausibly no possible interventions on the explanans, and the explanans and the explanandum relate mathematically (Woodward, 2018, pp. 126–128). Here is one of his examples:

> ***Strawberry***
> *"That Mother has three children and twenty-three strawberries, and that twenty-three cannot be divided evenly by three, explains why Mother failed when she tried a moment ago to distribute her strawberries evenly among her children without cutting any."* (Lange, 2013, p. 488; cited in Woodward, 2018, p. 126)

In discussing this example, Woodward takes the explanandum to be that the mother's strawberries are not evenly divisible amongst her three children rather than that she failed to divide them a moment ago (Woodward, 2018, p. 126). Henceforth, I likewise stick to this interpretation of the explanandum.

In the above example, Woodward relaxes the requirement for interventions (Woodward, 2018, p. 126). According to him, we can manipulate the number of strawberries, e.g., by giving more strawberries to the mother. In his opinion, such manipulations might not be interventions in the technical sense of this notion. This is because the possibility to divide strawberries cannot be the causal effect of an intervention on the number of strawberries, says Woodward (Woodward, 2018, p. 123, footnote 4).

However, Woodward argues, the explanation of the strawberries' indivisibility still answers w-questions (Woodward, 2018, p. 126). Manipulating the number of strawberries yields a pattern of counterfactual dependence between the number of strawberries and their divisibility. This, says Woodward, makes the strawberry case a genuine explanation.

Moreover, Woodward implies that the relationship between the explanans and explanandum is invariant for mathematical reasons here (Woodward, 2018, p. 127). Manipulating the number of strawberries, he claims, brings about the indivisibility as a matter of mathematical necessity.

Woodward's view on non-causal explanations

Which elements of the interventionist theory does Woodward uphold and which ones does he dispense with when extending it to the two varieties of non-causal explanations described above? First, Woodward maintains that explanations show how changes in an explanandum counterfactually depend on changes in the explanans (w-question requirement). Second, he dispenses with interventions. Third, he relaxes the invariance requirement. The relation between the explanans and the explanandum need not be invariant under interventions, but rather under changes in the explanans. In some cases, these invariant relations are empirical, in others mathematical.

2.2.2 Jansson and Saatsi (2019) on abstract explanations

Jansson and Saatsi (2019) aim to extend the interventionist theory to cover what they call abstract explanations. Here is an example of such an abstract explanation, as also illustrated in Figure 3 (Jansson & Saatsi, 2019, p. 821):

> ***Euler's explanation of Königsberg's non-traversability***[8]
>
> *In 1736, Königsberg had four parts and seven bridges. It turned out to be impossible to pass through all parts of Königsberg while crossing each bridge exactly once. Why was Königsberg's bridge system not traversable in such a way? The mathematician Leonhard Euler provided a graph-theoretical explanation of Königsberg's non-traversability. First, Euler proved the following theorem (Euler's theorem): A path through a connected graph G is an Euler path iff it passes each edge of G exactly once. There is an Euler path through a graph G iff G is an Eulerian graph. A graph G is an Eulerian graph iff (i) any node in G is connected to an even number of edges or (ii) exactly two nodes are connected to an odd number of edges. Second, Euler showed that the bridge system in Königsberg was not isomorphic to an Eulerian graph. No part of the town (corresponding to the nodes) had an even number of bridges connected to it (corresponding to the edges) and more than two parts were connected to an odd number of bridges. Finally, Euler explained the non-traversability of Königsberg's bridge system by referring to his theorem; Königsberg's bridge system is non-traversable because it is not isomorphic to an Eulerian graph.*

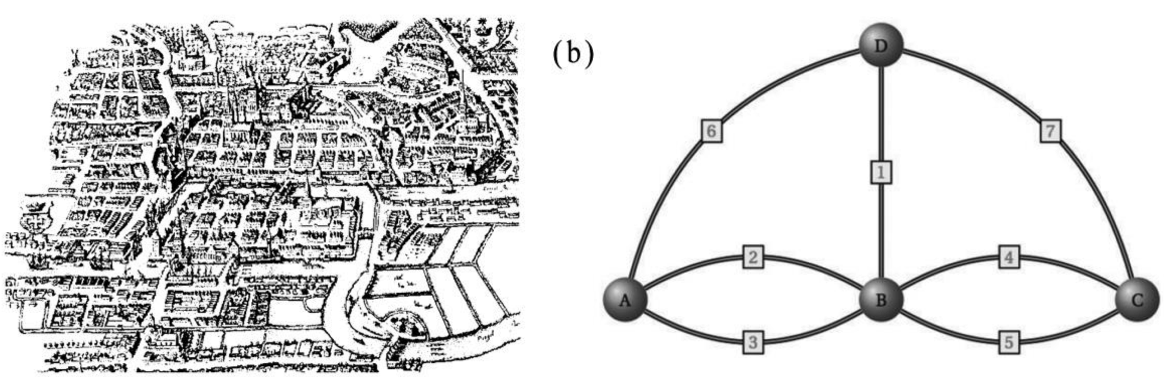
(a)


Figure 3: (a) *The bridge system of Königsberg in 1736* (Diestel, 2017, p. 22).
(b) *Graph representing Könighberg's bridge system in 1736* (Jansson & Saatsi, 2019, p. 837).
A – D denote parts of Königsberg. 1 – 7 denote the bridges connecting these parts.

[8] My description follows (Reutlinger, 2018, p. 83f; see also Diestel, 2017, p. 22ff).

Jansson and Saatsi believe Euler's explanation is plausibly a non-causal explanation (Jansson & Saatsi, 2019, p. 821). This is because, for them, it is unclear whether a manipulation of the bridge system "necessarily satisfies our intuitions about what counts as a causal intervention" (Jansson & Saatsi, 2019, p. 833). How do Jansson and Saatsi extend the interventionist theory to cover such non-causal, abstract explanations? They do so in three steps:

First, Jansson and Saatsi identify a dimension of abstraction any (causal or non-causal) explanation can exhibit to varying degrees (Jansson & Saatsi, 2019, pp. 822, 839–841): Explanations can abstract from the actual values of the explanans variables, holding for a range of values of them. The wider this range, the more abstract the explanation. For instance, Euler's explanation abstracts from the actual number of bridges and parts of Königsberg, holding for any number of bridges and any number of parts in a town.[9]

Second, Jansson and Saatsi identify the abstraction from the actual values of an explanans with the invariance of the explanation under changes in the explanans, and thus with the explanatory power of causal and non-causal explanations (Jansson & Saatsi, 2019, pp. 829ff, 839–841). The more an explanation abstracts from the actual values of the explanans, the more invariant it is to changes in these values. For example, Euler's theorem abstracts from the particular configuration of a bridge system and is thus invariant to changes in configurations of bridge systems. And in virtue of being invariant under changes in the explanans, an explanation provides counterfactual information about the explanandum, the source of its explanatory power. For instance, Euler's theorem tells us how the non-traversability of Königsberg's bridge system had been different had the bridge system been different. Moreover, this identification of explanatory power with the explanation's abstraction holds for causal and non-causal explanations alike. Both can abstract from the actual values of the explanans.

Third, Jansson and Saatsi propose to dispense with interventions, the causal component of the interventionist theory they aim to extend (Jansson & Saatsi, 2019, p. 831f). An explanation only needs to be invariant under changes in the explanans, rather than under interventions on the explanans. This, so they argue, still allows us to reason about how changes in an explanandum

[9] The authors distinguish this dimension from others (Jansson & Saatsi, 2019, p. 821 ff, 839–841).

counterfactually depend on changes in the explanans. We still understand how Königsberg's non-traversability counterfactually depends on changes in the configuration of its bridge system.

To summarize, Jansson and Saatsi uphold some and dispense with other elements of the interventionist theory: First, they maintain that explanations show how changes in an explanandum counterfactually depend on changes in the explanans (w-question requirement). Second, they dispense with interventions. Third, according to them, explanations are invariant under changes in the explanans but need not be invariant under interventions.

2.2.3 Povich (2018)'s generalized ontic conception of explanation

Povich (2018) proposes a generalized ontic conception of explanation. Here is one of his examples (Povich, 2018, pp. 124, 131; see also Widom & Mahan, Invalid Date):

> ***Liquid crystals***
> *Many substances melt directly from a solid state into a liquid one, e.g., water. However, some exhibit a so-called liquid crystal phase in which they show the properties of a solid and a liquid. Specifically, liquid crystals show fluid behaviour while the molecules are structurally ordered in the same direction (on average), unlike the molecules in normal (isotropic) liquids (see Figure 4). This so-called anisotropic behaviour is explained by the rod-like shape of the molecules in liquid crystals.*

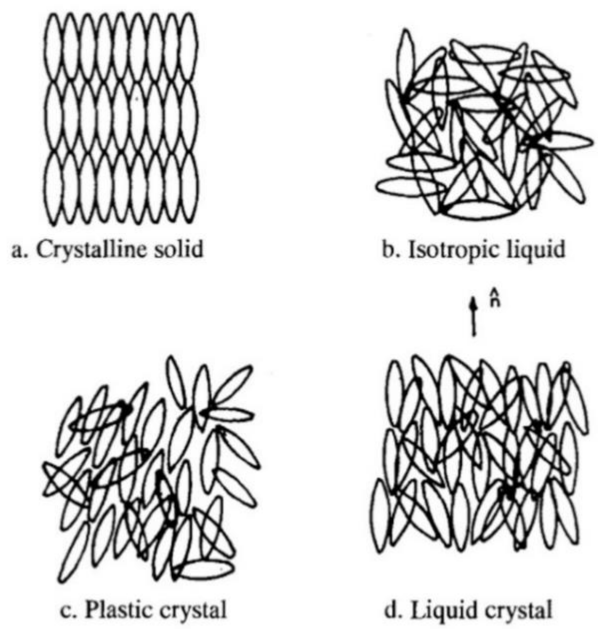


Figure 4: *Schematic representation of the arrangement of molecules in different phases* (Singh, 2002, p. 2).

Povich's generalized ontic conception of explanation accounts for the explanatory power of this explanation with two elements (Povich, 2018, pp. 131–132):

1. Explanations answer w-questions.

According to Povich, the liquid crystal explanation tells us how the fluid behaviour (the explanandum) would differ if the shape of the molecules (the explanans) would differ. Thereby, it answers w-questions about the fluid behaviour. In Povich's view such information is provided by ontic dependencies:

2. "Explanations provide information about relations of ontic dependence, causal and non-causal, which can be used to answer w-questions about the explanandum phenomenon." (Povich, 2018, p. 130)

According to Povich, the fluid behaviour of liquid crystals depends on the shape of their molecules (Povich, 2018, p. 131). He argues that this dependence is not causal as it holds between a feature of the whole liquid and features of its parts.[10] For Povich, one can still cite this ontic, non-causal dependence to provide answers to w-questions about the fluid's behaviour.

Which elements of the interventionist theory does Povich uphold in his generalized ontic conception and which ones does he dispense with? First, he upholds the w-question requirement. Second, he dispenses with interventions. In his view, one need not cite causal dependencies to explain. In other words, interventions are not required. Third, Povich does not tell us whether the ontic dependencies that ground explanatory counterfactuals ought to be invariant. For him, they need not be invariant *under interventions*. To be informative about how the explanandum would change if the explanans were to change the ontic dependence must nevertheless be invariant under such changes. Povich seems to assume such invariance. Finally, Povich emphasizes the role of ontic dependencies in explanations, just as the interventionist theory does in the case of causal explanations (see Section 2.1).

[10] Povich also claims that this dependence is not (causally) mechanistic; the molecules are not sufficiently organized to constitute a mechanism bringing about the fluid behaviour (Povich, 2018, p. 131).

2.2.4 Saatsi and Pexton (2013) on explanations of regularities

Saatsi and Pexton (2013) aim to extend the interventionist theory to cover causal and non-causal explanations of regularities. Here is their example of a non-causal explanation of a regularity, as also illustrated in Figure 5 (Saatsi & Pexton, 2013, pp. 618–620; see also West et al., 1999):

> ***Kleiber's law***
>
> *In many different types of organisms, the basal metabolic rate B of an organism and its body mass M relate proportionally with a scaling exponent $\frac{3}{4}$. This observed regularity is called Kleiber's law:*
>
> $$\text{Kleiber's law: } \mathrm{B} \propto \mathrm{M}^{\frac{3}{4}}$$
>
> *Why is the scaling exponent in Kleiber's law $\frac{3}{4}$? To answer this question, some biologists use a mathematical model relating the scaling exponent to an evolutionary optimal geometrical feature of the resource-distribution network of organisms. This geometrical feature is the fractal-like geometry of organisms' resource-distributing networks (like mammals' blood circularity system or a plants' vascular system, see Figure 5a). Fractal-like means that the network looks similar at different scales (see Figure 5b). This feature is derived from a set of assumptions about how resource-distributing networks work (e.g., that the supply of resources must serve all parts of the organism). From this geometrical feature (and other assumptions) biologists derive that the scaling exponent of Kleiber's law is $\frac{3}{4}$. This derivation can be generalized such that for d-dimensional networks the scaling exponent is $\frac{d}{d+1}$. For example, if organisms have two dimensions their basal metabolic rate relates to their body mass proportionally with a scaling exponent of $\frac{2}{3}$.*

According to Saatsi and Pexton, biologists provide a non-causal explanation of the scaling exponent $\frac{3}{4}$ in deriving it from the fractal-like geometry of resource-distributing networks (Saatsi & Pexton, 2013, p. 619f). First, this derivation, they argue, is explanatory as it provides counterfactual information about how changes in the scaling exponent depend on changes in the dimensionality of organisms (determining the fractal-like geometry of the resource-distributing networks). In virtue of providing such counterfactual information, the derivation explains the scaling exponent. Second, however, the authors argue that one cannot make sense of an intervention on the dimensionality of organisms that would change the scaling exponent (Saatsi & Pexton, 2013, p. 620). This is because this exponent is not a feature of a single (type of) organism. Therefore, they maintain, that we cannot make sense of what it would mean to change the scaling exponent as a result of physically changing the dimensionality of different organisms. Thus, according to the authors, the explanation of Kleiber's law is a non-causal explanation.

Saatsi and Pexton provide a unified account of causal and non-causal explanations of regularities by upholding some and dispensing with other elements of the interventionist theory (Saatsi & Pexton, 2013): First, they uphold that to explain means to show how the explanandum counterfactually depends on changes in the explanans (w-question requirement). Second, they dispense with interventions. For them, whether changes in the explanans are brought about by interventions is irrelevant for the counterfactual dependence to be explanatory. Third, they seem to uphold a relaxed invariance requirement. They argue that the above explanation refers to "the invariant geometrical fact" (Saatsi & Pexton, 2013, p. 620) that the scaling exponent relates to the fractal-like structure of the resource-distributing networks and thereby to the dimensionality of organisms. This suggests that Saatsi and Pexton view the explanation as invariant under changes in the dimensionality of organisms for mathematical rather than empirical reasons.

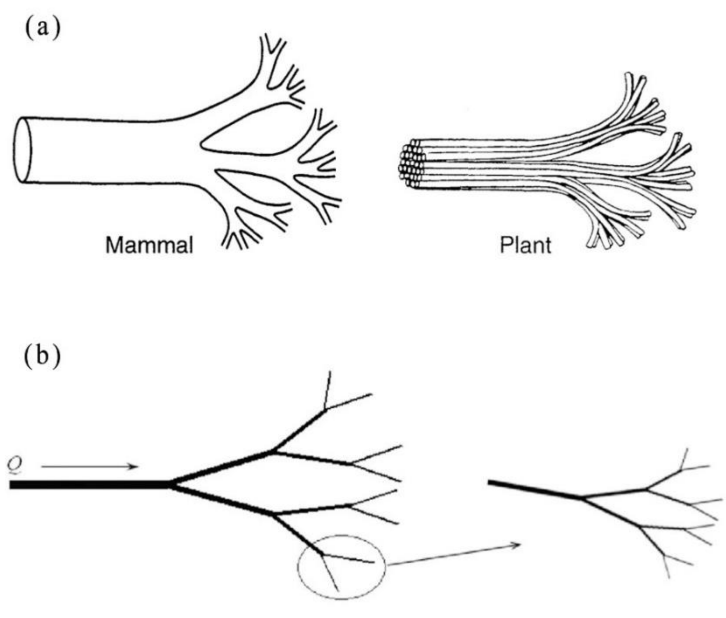


Figure 5: (a) *Schematic representation of biological resource-distribution networks* (West et al., 1997, p. 123; slightly altered).
(b) *Schematic representation of a fractal-like structure* (Kou et al., 2014, p. 528; slightly altered).

2.3 A counterfactual theory of causal and non-causal explanation

In the previous section, I reviewed four attempts to uphold some and dispense with other elements of the interventionist theory to capture the explanatory power of non-causal explanations. We can see a common view amongst the authors: They uphold the w-question requirement while dispensing with interventions. Despite this common view, none of the authors explicitly defines the counterfactual theory of explanation they aim for. We can do so by dispensing with the notion of an intervention in the three necessary and jointly sufficient conditions for an explanation in the interventionist theory, Exp_1, Exp_2 and Exp_3 (see Section 2.1, and also Woodward, 2003, p. 203). Here is the resulting view:

> (**CEXP**) Suppose that M is an explanandum consisting in the statement that some variable Y takes on y. Then an explanans E for M will consist of (a) (a generalization G describing) a relationship between Y and X (where X may consist of several variables $X_1 \ldots X_n$), and (b) a statement that X takes on x. E is explanatory with respect to M iff
> ($CExp_1$) E and M are (approximately) true.
> ($CExp_2$) (G describing) the relationship between X and Y correctly describes the actual value y of Y when X takes on its actual value x.
> ($CExp_3$) (G describing) the relationship between X and Y correctly describes that Y would change from y′ to y′′ if X would change from x′ to x′′ for at least some values of X.

CEXP captures the common view the authors reviewed in the last section strive after: Explanations answer w-questions ($CExp_3$). CEXP dispenses with interventions ($CExp_2$ and $CExp_3$). And CEXP requires that the relationship between X and Y remains invariant not under interventions but under some changes of X ($CExp_3$).[11]

Note also that $CExp_2$ refers to the dependence of the actual value y of Y on the actual value x of X, and $CExp_3$ to the change-relating counterfactual dependence of Y on X. Here, the counterfactual dependence of Y on X expressed in both conditions is necessary and sufficient for X to explain Y. I return to this feature in Chapter 3.

[11] Other authors defending ideas similar to CEXP include (Baron et al., 2017; Bokulich, 2011; French & Saatsi, 2018; Rice, 2015). (Reutlinger, 2018) also provides a counterfactual theory of causal and non-causal explanation, adding a necessary condition that the explanandum must be deductively deduced from the explanantia.

This characterization of the common view amongst the authors raises two questions:

(1) Is CEXP a theory of non-causal explanation or a theory of causal *and* non-causal explanation?

To answer question (1), it is useful to clarify when an explanation is causal for proponents of CEXP: The authors refer to the interventionist criterion for a causal relationship between explanans X and explanandum Y. X causes Y iff a possible intervention on X would change Y. Based on this demarcation, we can rephrase question (1): Does CEXP only apply to explanations in which there are no possible interventions on X that would change Y (i.e. non-causal explanations) or also to explanations in which there are possible interventions on X that would change Y (i.e. causal explanations)?

The authors differ in the clarity of their answers. Povich (2018) explicitly argues that his generalized ontic conception aims to cover causal *and* non-causal explanations. Similarly, Saatsi and Pexton (2013) explicitly argue that they aim for a unified account of causal *and* non-causal explanations of regularities. By contrast, Jansson and Saatsi (2019) remain vague on whether they propose to dispense with interventions only in the case of abstract, non-causal explanations or also in causal explanations. Likewise, Woodward (2018) leaves open whether he proposes to dispense with interventions also in the case of causal explanations. Only if these authors would deny the need for interventions in causal explanations, then CEXP would cover causal *and* non-causal explanations in their views.

What to make of this? I suggest interpreting Jansson and Saatsi (2019) and Woodward (2018) as aiming for CEXP to cover causal *and* non-causal explanations. This squares well with their respective monist aspirations (see Jansson & Saatsi, 2019, p. 818; Woodward, 2018, p. 136). Hence, I understand CEXP as a theory of causal *and* non-causal explanation.

Here is the second open question regarding CEXP:

(2) Is CEXP supposed to be a theory of *all* explanations?

Question (2) is distinct from question (1). Even if CEXP is taken to be a theory of causal and non-causal explanation, it could be taken to cover only some explanations. In response to question (2), I note two aspects of the views discussed in Section 2.2:

First, the authors only discuss explanations of physical phenomena, like Kleiber's law or Königberg's non-traversability. None of them discusses for instance explanations of mathematical objects (see also Woodward, 2018, p. 118). Second, Woodward, Povich and Jansson and Saatsi do not qualify whether there are any explanations that they do *not* aim to cover. By contrast, Saatsi and Pexton restrict their account to explanations of regularities. What to make of this divergence? Saatsi and Pexton restrict their view only due to a tension they see between a unified account of causal and non-causal explanations of regularities and a unified account of causal explanations of regularities and singular states of affairs (Saatsi & Pexton, 2013, p. 620f). Presumably, if one could overcome this tension, they would prefer a unified account of causal and non-causal explanations of regularities and singular states of affairs. Based on these two aspects, I suggest understanding CEXP as referring to explanations of physical phenomena, including regularities and singular states of affairs.

CEXP faces a problem: it does not always account for the asymmetry that explanations exhibit. In the next chapter, I describe this problem, introducing what I call the challenge of asymmetry for CEXP.

3 The challenge of asymmetry

One could raise many problems for a counterfactual theory of causal and non-causal explanation like CEXP (see e.g., Khalifa et al., 2020; Kuorikoski, 2021). In this thesis, I focus on just one, pressing and familiar challenge: how to account for the asymmetry that explanations exhibit. In this chapter, I introduce this challenge (Section 3.1) and distinguish it from a revised one, which allows for some symmetric explanations (Section 3.2).

3.1 The challenge of asymmetry

Let us return to the flagpole explanation (see Section 2.1). Traditionally, many philosophers have taken this explanation to be asymmetric (see e.g., Bromberger, 1966; Khalifa et al., 2021; Kitcher & Salmon, 1987; Woodward, 2019). The angle of sunrays and the flagpole's height explain its shadow's length, but its shadow's length can hardly be used in an explanation of its height or the angle of sunrays. In the context of CEXP, we can define explanatory asymmetry as follows:

Explanatory asymmetry

An explanation is asymmetric iff

(1) it is true that (i) the relationship between Y and X (which may consist of $X_1 \ldots X_n$) and (ii) X's taking on the value x explain that Y takes on the value y and,

(2) for at least some X_i, it is false that (i) the relationship between Y and X, (ii) Y's taking on the value y and (iii) X_j's taking on the value x_j for all $j \neq i$ explain that X_i takes on the value x_i.[12]

[12] Here is an alternative definition of explanatory asymmetry (see also Reutlinger, 2018, p. 92):

An explanation is asymmetric iff

(1) it is true that (i) the relationship between Y and X (which may consist of $X_1 \ldots X_n$) and (ii) X's taking on the value x explain that Y takes on the value y and,

(2) it is false that (i) the relationship between Y and X and (ii) Y's taking on the value y explain that X takes on the value x.

For example, G1, the angle of sunrays x_2^1 and the flagpole's height x_1^1 explain its shadow's length y^1. To be an asymmetric explanation according to this definition, it should then be false that G1 and the flagpole's shadow's length y^1 explain its height x_1^1 and the angle of sunrays x_2^1. Here is my reason to opt for the definition in the main text: It seems to better capture our intuitions on explanatory asymmetry when X consists of several variables. For instance, our intuition seems to be that we cannot use the flagpole's shadow's length in an explanation of its height or of the angle of sunrays. The definition in the main text captures this intuition. To be asymmetric, the flagpole's shadow's length should not explain its height (jointly with the angle of sunrays) and/ or the angle of sunrays (jointly with the flagpole's height). By contrast, according to the alternative definition we cannot use the flagpole's shadow's length to explain its height *and* the angle of sunrays. But our intuition was about the shadow's

Footnote continued on next page

For example, G1 (describing how the flagpole's shadow's length, its height and the angle of sunrays relate), the angle of sunrays x_2^1 and the flagpole's height x_1^1 explain its shadow's length y^1. To be an asymmetric explanation, it should be false that (a) G1, the angle of sunrays x_2^1 and the flagpole's shadow's length y^1 explain its height x_1^1 and/ or that (b) G1, the flagpole's shadow's length y^1 and its height x_1^1 explain the angle of sunrays x_2^1.

In the philosophy of explanation, many philosophers accept that all explanations are asymmetric, sometimes with the qualification of rare exceptions (e.g., Khalifa et al., 2021; Lange, 2019; Schindler, 2014). The broad consensus is that a theory of explanation ought to account for this asymmetric feature of explanations. In this thesis, I do not assume that all explanations are asymmetric. However, before turning to potentially symmetric explanations in Section 3.2, let me explain why the asymmetry of explanations poses a threat to CEXP.

As has been acknowledged (e.g., by Held, 2019; Lange, 2019; Schindler, 2014), a counterfactual theory like CEXP faces a problem in accounting for explanatory asymmetry: It can happen that Y counterfactually depends on changes in X *and* X counterfactually depends on changes in Y. In such cases, CEXP claims that X explains Y and Y explains X. But then CEXP fails to account for explanatory asymmetry!

To illustrate, consider the flagpole explanation again (see also Woodward, 2003, p. 197f). Given G1, at least for some changes in X_1 and X_2, it is true that

(CF_1) If the flagpole's height had changed from x_1' to x_1'' and/ or the angle of sunrays from x_2' to x_2'', the flagpole's shadow's length would have been y''.

According to CEXP, the angle of sunrays x_2^1 and the flagpole's height x_1^1 explain its shadow's length y^1.

length merely being part of an explanation of either, rather than doing all the explanatory work of both. For this reason, the definition in the main text seems to better capture our intuitions on asymmetry in explanations with several explanantia. In any case, though I only refer to the definition in the main text, my arguments in this thesis also hold for the alternative definition.

However, G1 also supports the truths of the following counterfactuals, for at least some changes in Y, X_1 and X_2:[13]

(CF_2) If the flagpole's shadow's length had changed from y' to y'' and/ or the angle of sunrays from x_2' to x_2'', the flagpole's height would have been x_1''.

(CF_3) If the flagpole's shadow's length had changed from y' to y'' and/ or the flagpole's height from x_1' to x_1'', the angle of sunrays would have been x_2''.

According to CEXP, the flagpole's shadow's length y^1 and the flagpole's height x_1^1 explain the angle of sunrays x_2^1. And according to CEXP the angle of sunrays x_2^1 and the flagpole's shadow's length y^1 explain its height x_1^1. But this result runs against the asymmetry this explanation is thought to exhibit![14]

Failures to account for the asymmetry of some explanations challenge proponents of CEXP to qualify their theory further:

The challenge of asymmetry
Explanations are asymmetric, but counterfactual dependence need not be. How to qualify CEXP such that a thus qualified CEXP claims all explanations it covers to be appropriately asymmetric?

[13] This is easiest to see when we transform G1 to

$$(\text{G1*})\ \text{height of flagpole X1} = \text{flagpole's shadow's length Y} \times \tan(\text{angle of sunrays X2})$$

$$(\text{G1**})\ \text{angle of sunrays X2} = \tan^{-1}\left(\frac{\text{height of flagpole X1}}{\text{flagpole's shadow's length Y}}\right)$$

[14] The same problem occurs under the alternative definition of explanatory asymmetry: Given G1, at least for some values of X_1, X_2 and Y, it is true that

(CF_1^a) If the flagpole's height had changed from x_1' to x_1'' and/ or the angle of sunrays from x_2' to x_2'', then the flagpole's shadow's length would have been y''.

(CF_2^a) If the flagpole's shadow's length had changed from y' to y'', then the angle of sunrays would have been x_2'' and the flagpole's height x_1''.

CEXP fails to yield explanatory asymmetry, also according to the alternative definition.

There are two open questions about how CEXP may be qualified so as to solve the challenge of asymmetry:

1) CEXP claims that the counterfactual dependence of Y on X referred to in $CExp_1$ – $CExp_3$ is necessary and sufficient for X to explain Y. Should one uphold this element of CEXP when meeting the challenge of asymmetry?

I propose to distinguish between two kinds of responses to this question that both qualify CEXP in different ways: A **conservative CEXP** upholds that the counterfactual dependence of Y on X referred to in $CExp_1$ – $CExp_3$ is necessary and sufficient for X to explain Y. To account for explanatory asymmetry, it qualifies $CExp_2$ and/ or $CExp_3$ by specifying which kind of counterfactual dependence they refer to, such that non-explanatory counterfactuals are false. A **radical CEXP** claims that the counterfactual dependence of Y on X referred to in $CExp_1$ – $CExp_3$ is necessary but not sufficient for X to explain Y. To account for explanatory asymmetry, it adds further necessary[15] conditions to CEXP which X and Y must fulfil for X to explain Y.[16]

2) CEXP is a theory of causal *and* non-causal explanation. Does addressing the challenge of asymmetry (i) require qualifying CEXP in the same way for causal *and* non-causal explanations or (ii) may different qualifications be used respectively?

I again suggest distinguishing between two kinds of responses that involve a qualified CEXP one might want to defend: If one uses the same qualifications to account for explanatory asymmetry in causal and non-causal explanations, then one defends a **strongly monist CEXP**. In a radical version, this amounts to adding the same condition to CEXP. In a conservative version, this amounts to qualifying $CExp_2$ and/or $CExp_3$ in the same way. If one uses different qualifications for causal and non-causal explanations respectively, then one defends a **weakly monist CEXP**. In a radical version, this amounts to adding different conditions for causal and non-causal explanations

[15] Proponents of CEXP aim to provide necessary and jointly sufficient conditions for X to explain Y. Taking $CExp_1$ – $CExp_3$ as necessary conditions and adding a condition to CEXP that is not necessary but sufficient for X to explain Y would break with this commitment. For this reason, I will not discuss this option.

[16] One could also claim that the counterfactual dependence of Y on X is neither necessary nor sufficient or only sufficient for X to explain Y. Either option would break with the core commitment of CEXP that counterfactual dependence is necessary for explanatory power. Hence, I will not discuss these options.

to CEXP. In a conservative version, this amounts to qualifying $CExp_2$ and/or $CExp_3$ in different ways for causal and non-causal explanations.[17]

In Part II and III, I discuss several conservative or radical, strongly or weakly monist versions of CEXP, proposed to solve the challenge of asymmetry. Before doing so, it is important to distinguish the challenge of asymmetry from a revised one.

3.2 The revised challenge of (a)symmetry

As already mentioned, many philosophers accept that explanations are asymmetric, perhaps with some rare exceptions. However, in recent debates, there is less consensus on the asymmetry of non-causal explanations: Some philosophers hold non-causal explanations to be asymmetric (e.g., Jansson & Saatsi, 2019; Lange, 2019; Woodward, 2018), while others allow for some symmetric ones (Reutlinger, 2017b, p. 253, 2018, p. 92; Saatsi & Pexton, 2013).

The sketched unclarity about the asymmetry of non-causal explanations risks confusing a debate about the challenge of asymmetry. Any failure to identify the asymmetry of a non-causal explanation could equally be a success to identify it as symmetric. To avoid such confusion, I propose to clearly distinguish the challenge of asymmetry from a revised challenge:

The revised challenge of (a)symmetry

Causal explanations are asymmetric. Some non-causal explanations are asymmetric and others symmetric. Counterfactual dependence need not be asymmetric. How to qualify CEXP such that a thus qualified CEXP claims all explanations it covers to be appropriately asymmetric or symmetric?

[17] A third option would be to use different qualifications for causal and also for various non-causal explanations. This pluralist version of CEXP conflicts with the monist aspirations of proponents of CEXP (see Section 2.3). Hence, I will not discuss this option.

Here is the corresponding definition of explanatory symmetry:

Explanatory symmetry

An explanation is symmetric iff

(1) it is true that (i) the relationship between Y and X (which may consist of $X_1 \ldots X_n$) and (ii) X's taking on the value x explain that Y takes on the value y and,

(2) for all X_i, it is true that (i) the relationship between Y and X, (ii) Y's taking on the value y and (iii) X_j's taking on the value x_j for all $j \neq i$ explain that X_i takes on the value x_i.

To illustrate, consider Euler's explanation of Königsberg's non-traversability (see Section 2.2.2). Euler's theorem and the configuration of Königsberg's bridge system explain its non-traversability. To be a symmetric explanation, it must also be true that Euler's theorem and Königsberg's non-traversability explain the configuration of its bridge system.

The challenge of asymmetry and the revised challenge are mutually exclusive alternatives: A response to the former must qualify CEXP such as to yield that all explanations it covers are asymmetric. By contrast, a response to the latter requires to qualify CEXP such that some non-causal explanations are symmetric while all other explanations are asymmetric. Therefore, the distinction between both allows us to specify when a theory fails to account for explanatory asymmetry and when it successfully identifies a symmetric explanation, depending on which challenge we aim to address.

In principle, which challenge a proponent of CEXP aims to solve depends on whether she views some non-causal explanations as symmetric. Within this thesis, I mainly focus on the challenge of asymmetry. This means any failure to account for the asymmetry of an explanation will pose a problem for CEXP and qualified versions of it. However, I return to the revised challenge in Chapter 8.

Having introduced the asymmetry challenges, one might wonder: Does the interventionist theory of causal explanation address explanatory asymmetry in a way that proponents of CEXP could use to solve their challenge of asymmetry? It does not. The interventionist theory accounts for asymmetry by counting only true counterfactuals whose antecedent is brought about by interventions as explanatory (see Woodward, 2003, p. 197f). I provide a detailed explanation of how interventions account for asymmetry in Appendix A. Those details do not matter for our purposes. CEXP dispenses with interventions, and thus proponents of it cannot rely on interventions to account for explanatory asymmetry. We need an alternative solution to the challenge of asymmetry. This is the task of Part II and III.

Part II Against two potential solutions to the challenge of asymmetry

In Part I, I defined a counterfactual theory of causal and non-causal explanation that several authors strive after, CEXP. Moreover, I introduced the challenge of asymmetry for CEXP: proponents of CEXP need to qualify their theory such that it identifies all explanations it is meant to cover as asymmetric. Part II argues against two potential solutions to this challenge. I reject an appeal to Lewisian semantics (Chapter 4) and to two inference schemes, Woodward's (2018, 2020) double explanation principle and Lange's (2019) modal fact principle (Chapter 5). This motivates the need for a more promising avenue to deal with explanatory asymmetry as a proponent of CEXP. The quasi-interventionist version of CEXP I develop in Part III provides such an alternative.

4 Against Lewisian semantics

To address the challenge of asymmetry, one could propose to add a semantics for counterfactuals to CEXP. In such a conservative CEXP, one would uphold that the counterfactual dependence of Y on X is necessary and sufficient for X to explain Y, i.e. upholding $CExp_1$ – $CExp_3$. A semantics adequate to account for explanatory asymmetry would then evaluate explanatory counterfactuals as true and the reversed non-explanatory ones as false. It would yield asymmetry of counterfactual dependence. As a result, a conservative CEXP with such a semantics would imply explanatory asymmetry. As is well-known, Lewisian semantics yields asymmetry of counterfactual dependence. At first glance, Lewisian semantics is hence a promising candidate for such a semantic solution to the challenge of asymmetry.

However, in this chapter, I argue that it is unclear how a conservative CEXP with Lewisian semantics would solve the challenge of asymmetry. In Section 4.1, I describe when Lewisian semantics yields asymmetry of counterfactual dependence. Readers unfamiliar with Lewisian semantics may also consult Appendix B where I describe how it evaluates counterfactuals. In Section 4.2, I then argue that it is unclear how Lewisian semantics would establish asymmetry of counterfactual dependence in non-causal explanations with explanantia and/ or explananda that are not events. Therefore, it is unclear how a conservative CEXP with Lewisian semantics would account for asymmetry in such explanations such as to solve the challenge of asymmetry.[18]

[18] Held also argues against Lewisian semantics as establishing asymmetry of counterfactual dependence in non-causal explanations (Held, 2019, pp. 463–465). More precisely, Held claims that Lewisian semantics does not establish
Footnote continued on next page

4.1 On asymmetry of counterfactual dependence in Lewisian semantics

In this section, I describe when Lewisian semantics yields asymmetry of counterfactual dependence. Throughout this section, I assume readers are familiar with Lewisian semantics. I explain in detail how Lewisian semantics evaluates counterfactuals in Appendix B. Here, I only briefly recap its key elements: Lewis' counterfactual analysis and his standard measure of similarity. Consider the following example (see e.g., Menzies & Beebee, 2020):

> ***Suzy's rock-throwing***
> At time t, Suzy throws a rock at a window. The rock flies through the air, hits the window and the window shatters at time t*.

Here are two counterfactuals involved in this example:

($\mathrm{CF_L}$) If Suzy had not thrown the rock, the window would not have shattered.

($\mathrm{CF_L^b}$) If the window had not shattered, Suzy would not have thrown the rock.

According to Lewis' counterfactual analysis, counterfactuals involving events like $\mathrm{CF_L}$ or $\mathrm{CF_L^b}$ are true iff some possible world in which both the antecedent and the consequent is true is more similar to the actual world than any possible world in which the antecedent but not the consequent is true (Lewis, 1979, p. 465, see also 1986). For example, $\mathrm{CF_L}$ is true iff some possible world in which Suzy does not throw the rock and the window does not shatter is more similar to the actual world than any possible world in which Suzy does not throw the rock but the window still shatters.

asymmetry of counterfactual dependence in an explanation with a mathematical object as an explanandum. He refers to a version of Euler's explanation in which the explanandum is that a graph representing Königsberg's bridge system does not have an Euler path (Held, 2019, p. 453f). However, as discussed in Section 2.3, proponents of CEXP are interested in explanations of physical phenomena rather than of mathematical objects. Correspondingly, proponents of CEXP understand Euler's explanation solely as one in which the explanandum is Königsberg's non-traversability rather than that a graph representing Königsberg's bridge system does not have an Euler path. For this reason, Held's criticism does not address whether a conservative CEXP with Lewisian semantics solves the challenge of asymmetry proponents of CEXP face.

To compare the similarity of possible worlds, Lewis provides a standard measure of similarity designed for “our usual sort of counterfactual reasoning” (Lewis, 1979, p. 457):[19]

Lewis’ standard measure of similarity[20]

“(1) It is of the first importance to avoid big, widespread, diverse violations of law.

(2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.

(3) It is of the third importance to avoid even small, localized, simple violations of law.

(4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.” (Lewis, 1979, p. 472)

Lewis’ counterfactual analysis and his standard measure of similarity evaluate so-called non-backtracking counterfactuals like $\mathtt{CF}_{\mathrm{L}}$ as true and the reverse so-called backtracking ones like $\mathtt{CF}_{\mathrm{L}}^{\mathrm{b}}$ as false (see Lewis, 1979). In other words, Lewisian semantics yields asymmetry of counterfactual dependence. I provide details of how this works in the case of Suzy’s rock-throwing in Appendix B.

At first glance, given that Lewisian semantics yields asymmetry of counterfactual dependence, it would seem that it could be used to address the challenge of asymmetry. In Section 4.2, I argue against this suggestion. To understand this argument, it is important to clarify when exactly does Lewisian semantics yield asymmetry of counterfactual dependence.

Notably, neither Lewis’ counterfactual analysis nor his standard measure of similarity have any asymmetry built into them (Lewis, 1979, p. 473). Rather, they yield asymmetry of counterfactual dependence if the following claim holds (Lewis, 1979, p. 473f):

Asymmetry of overdetermination

Later events are seldom overdetermined by earlier ones, while earlier events are commonly overdetermined by later ones.

---

[19] Lewis argues that there are different adequate similarity measures for different contexts (Lewis, 1979, pp. 465–467). For our purposes, the standard similarity measure is adequate, because it yields asymmetry of counterfactual dependence.

[20] In deriving this standard measure, Lewis assumes that the actual laws of nature are deterministic (Lewis, 1979, p. 460f). This implies that two possible worlds which obey these deterministic laws perfectly are either exactly alike throughout time or not exactly alike for any period of time. I throughout assume determinism in this sense.

Lewis understands overdetermination as follows (Lewis, 1979, p. 49f): In a deterministic world, any later event is predetermined by earlier ones. This means that given the laws of nature there is a minimal set of jointly sufficient conditions (a so-called determinant) for the event. For instance, there is a determinant for Suzy's rock-throwing, namely her intentions to throw. Similarly, according to Lewis, any earlier event is postdetermined by later ones. Again, this means that there is at least one determinant for it. For instance, the shattering of the window is a determinant for Suzy's rock-throwing having occurred. Moreover, at a given time an event can be overdetermined by earlier or later ones. This means there is more than one determinant, i.e. more than one set of jointly sufficient conditions for it. The more determinants there are the more overdetermined is the event. For example, the shattering of the window, Suzy's memory of throwing the rock and the movement of air are all determinants for Suzy's rock-throwing. Suzy's rock-throwing (an earlier event) is overdetermined by these later events.[21]

Asymmetry of overdetermination amounts to the claim that earlier events are typically overdetermined by later ones, but not vice versa: Suzy's rock-throwing is overdetermined by later events. By contrast, her rock-throwing is not overdetermined by earlier events. Rather, presumably, her rock-throwing is determined by one determinant, her intentions to throw the rock. Similarly, the shattering of the window (a later event) is not (or at least less) overdetermined by earlier ones. Presumably, its only determinant is Suzy's rock-throwing.

Asymmetry of overdetermination is a contingent fact for Lewis (Lewis, 1979, p. 475). It does not hold by necessity or always. Yet, it usually holds in our world, or so maintains Lewis. And if asymmetry of overdetermination holds then Lewisian semantics yields asymmetry of counterfactual dependence (Lewis, 1979, pp. 472–475). To see this, let us return to Suzy's rock-throwing.

First, consider CF_L and the following possible worlds, denoting the actual world as W_0:

> ***W_1***: W_1 is exactly like W_0 until right before t when a small violation of law occurs such that Suzy does not throw the rock. Afterwards, W_1 unfolds without any further violations

[21] Strictly speaking, Lewis talks about facts or affairs in this passage. Since he is concerned with counterfactuals involving events, we can interpret asymmetry of overdetermination still as a claim about events (as do other authors see e.g., Price, 1992, p. 503f).

of law. W_1 then differs from W_0 in spatio-temporal facts after the small miracle right before t. In particular, in W_1, the window does not shatter.

W_2: W_2 has a different past than W_0. In particular, in W_2, Suzy does not throw the rock at time t. However, in W_2 a whole range of small violations of law occur right after t such as to make W_2 exactly like W_0 after t. The air is pushed away, Suzy remembers throwing the rock and so on. In particular, in W_2, the window shatters.

According to Lewis' standard measure of similarity, W_1 is more similar to W_0 than any possible world in which Suzy does not throw the rock but the window still shatters like W_2. This is because W_1 contains only a small miracle, conformity to laws after this miracle and to spatio-temporal facts of the actual world for the longest possible period (Weatherson, 2016, p. 17). In W_1, the window does not shatter. Hence, according to Lewis' counterfactual analysis, $\mathrm{CF_L}$ is true.

Underlying the verdict that $\mathrm{CF_L}$ is true is asymmetry of overdetermination: To diverge from W_0, W_1 just needs one small miracle breaking the link between Suzy's rock-throwing and its determinant, her intentions to throw. One just needs a small miracle due to asymmetry of overdetermination; later events, like Suzy's rock-throwing, are not overdetermined by earlier ones, like her intentions to throw. By contrast, to converge to W_0, a world without Suzy's rock-throwing like W_2 needs a whole range of small miracles ensuring that all the different effects of Suzy's rock-throwing still occur. One needs many small miracles (a big miracle) due to asymmetry of overdetermination; earlier events, like Suzy's rock-throwing, are overdetermined by later ones. According to Lewis' standard measure of similarity, worlds with big miracles are less similar to the actual world than worlds with small miracles. Thus, asymmetry of overdetermination and Lewis' standard measure jointly ensure that W_1 is more similar to W_0 than any world like W_2, and thereby that $\mathrm{CF_L}$ is true.

Second, consider $\mathrm{CF_L^b}$, W_1 and the following possible world:

W_3: W_3 starts exactly like W_0. In particular, in W_3, Suzy throws the rock. However, right before t*, a small miracle occurs such that the window does not shatter. Afterwards, W_3 unfolds without any further violations of law. W_3 then differs in spatio-temporal facts from W_0 after the small miracle before t*.

According to Lewis' standard measure, W_3 is more similar to W_0 than any world in which the window does not shatter and Suzy does not throw the rock like W_1. This is again because W_3 is a world with a small miracle, conformity to laws after this miracle and to facts of the actual world for the longest possible period (Weatherson, 2016, p. 17). In particular, W_3 matches W_0 for a longer period than a world like W_1. In W_3, Suzy throws the rock. Thus, according to Lewis' counterfactual analysis, $\mathrm{CF}_{\mathrm{L}}^{\mathrm{b}}$ is false.

Underlying the verdict that $\mathrm{CF}_{\mathrm{L}}^{\mathrm{b}}$ is false is again asymmetry of overdetermination: To diverge from W_0, worlds like W_1 still need one small miracle breaking the link between Suzy's rock-throwing and her intentions to throw. However, a world like W_1 differs completely in spatio-temporal facts after this miracle. This is again due to asymmetry of overdetermination; Suzy's rock-throwing has a whole range of effects in the actual world, i.e. is overdetermined by later events. All those effects do not occur in a world like W_1. Contrast this with W_3. To diverge from W_0, W_3 also just needs one small miracle breaking the link between the shattering of the window and Suzy's rock-throwing. W_3 also differs completely from W_0 in spatio-temporal facts after the miracle, due to asymmetry of overdetermination. However, the miracle in W_3 occurs later than the miracle in a world like W_1, and indeed as late as possible. Due to this a world like W_1 perfectly matches W_0 in spatio-temporal facts for a shorter period than W_3. According to Lewis' standard measure, worlds with a less extensive period of perfect match in spatio-temporal facts are less similar. Thus, asymmetry of overdetermination and Lewis' standard measure jointly ensure that W_3 is more similar to W_0 than any world like W_1, and thereby that $\mathrm{CF}_{\mathrm{L}}^{\mathrm{b}}$ is false.

To conclude, Lewisian semantics yields asymmetry of counterfactual dependence if asymmetry of overdetermination holds. For Lewis' view on asymmetry of counterfactual dependence, this might be unproblematic (though see Menzies & Beebee, 2020, pp. 13–15). However, as I argue in the next section, relying on asymmetry of overdetermination for asymmetry of counterfactual dependence *is* problematic when addressing the challenge of asymmetry.

4.2 Against Lewisian semantics

It is unclear how a conservative CEXP with Lewisian semantics would solve the challenge of asymmetry. Here is the argument for this claim:

Lewisian semantics yields asymmetry of counterfactual dependence if asymmetry of overdetermination holds (P1). Asymmetry of overdetermination refers to earlier and later events being overdetermined or not (P2). However, some of the explanantia and/ or explananda in the non-causal explanations proponents of CEXP are interested in are not events located in time, e.g., Königsberg's non-traversability (P3). And asymmetry of overdetermination as a claim about earlier and later events does not hold for explanantia and/ or explananda that are not events located in time (P4).[22] Moreover, it is unclear how Lewisian semantics would establish asymmetry of counterfactual dependence in explanations with explanantia and/ or explananda that do not exhibit asymmetry of overdetermination (P5). Thus, it is unclear how a conservative CEXP with Lewisian semantics would solve the challenge of asymmetry.

P5 requires a defence. This is because (on a charitable reading) Lewis does not claim that asymmetry of overdetermination is necessary for his semantics to yield asymmetry of counterfactual dependence. We cannot conclude from some explanantia and/ or explananda not exhibiting asymmetry of overdetermination to Lewisian semantics not identifying them as asymmetrically counterfactually dependent. P5 maintains that it is still unclear how Lewisian semantics would establish asymmetry of counterfactual dependence in such explanations. To argue for P5, I now illustrate how Lewisian semantics yields no clear verdict on asymmetry of counterfactual dependence in a non-causal explanation without events as explanandum and explanans.

Consider the explanation of why stable planetary orbits are possible (see Section 2.2.1). Recall, physicists explain this possibility by deriving it from the dimensionality of space. The challenge of asymmetry demands that we deem this explanation to be asymmetric. The dimensionality of space

[22] These explanantia and explananda are also not facts or affairs located in time. For this reason, even interpreting asymmetry of overdetermination as referring to facts or affairs somehow distinct from events does not undermine P3 and P4.

explains the possibility of stable planetary orbits and not vice versa.[23] For a conservative CEXP to yield this verdict it must be true for at least some changes in dimensionality that

> (CF_4) If the dimensionality of space had been different, stable planetary orbits would have been impossible.

And it must be false that

> (CF_5) If stable planetary orbits had been impossible, the dimensionality of space would have been different.[24]

The (im)possibility of stable planetary orbits is not an event. Neither is the dimensionality of space. Suppose we would still use Lewisian semantics to evaluate CF_4 and CF_5. Would doing so establish that CF_4 is true for at least some changes in space's dimensionality and CF_5 false? It remains unclear.

Consider CF_5. According to Lewis' counterfactual analysis, CF_5 is true iff some possible world in which stable planetary orbits are impossible and space is not three-dimensional is more similar to the actual world than any world in which stable planetary orbits are impossible and space is nevertheless three-dimensional. Denote W_0 as the actual world, and consider the following two possible worlds:

> ***W_4***: W_4 is exactly like W_0 until at some time there is a violation of law such that space is not three-dimensional. Afterwards, W_4 unfolds without any further violations of law. W_4 then differs from W_0 in spatio-temporal facts after the violation of law. In W_4, stable planetary orbits are impossible.

> ***W_5***: W_5 is exactly like W_0. In particular, in W_5, space is three-dimensional. However, at some time in W_5, a violation of law occurs such that stable planetary orbits are impossible. Otherwise, W_5 unfolds without any violation of law.

---

[23] According to Woodward, some physicists uphold this explanatory directionality, while others deem the possibility of stable planetary orbits to explain the dimensionality of space (Woodward, 2018, p. 125). My argument holds for either directionality the explanation might exhibit.

[24] Note that $CExp_2$ would not already yield explanatory asymmetry here. It is satisfied in both directions: The mathematical derivation tells us that if space is three-dimensional, stable planetary orbits are possible and vice versa.

Does Lewis' standard measure claim W_4 or W_5 to be more similar to W_0? It is unclear. One might be tempted to say that W_5 is more similar, yielding the correct verdict that CF_5 is false. After all, one might think that the violation of law in W_5 occurs later than in W_4. This would imply that W_5 would differ in spatio-temporal facts from W_0 for a shorter period than W_4, thus be more similar. However, the violation of law in W_5 might as well occur at any earlier time, in particular before the one in W_4. After all, the possibility of stable planetary orbits does not occur at any particular time.[25] W_4 would then be more similar to W_0 than W_5, as it would differ in spatio-temporal facts for a shorter period. If so, then CF_5 would be true. One could propose that the violation of law in W_5 ought to occur as late as possible, or at least later than the violation in W_4. However, there is no latest or later time. There is no particular time at which the possibility of stable planetary orbits and the dimensionality of space occur at all. Thus, there is also no latest or later time at which to make stable planetary orbits impossible or space not three-dimensional.

What ought we to conclude then about the truth value of CF_5? We do not know what to conclude. This is because we cannot make sense of the possibility of stable planetary orbits and the dimensionality of space as events occurring at a particular time. As a result, we fail to establish that CF_5 is false. Hence, it is unclear how Lewisian semantics would establish asymmetry of counterfactual dependence in this case.[26]

The explanation of the possibility of stable planetary orbits illustrates P5; the claim that it is unclear how Lewisian semantics would establish asymmetry of counterfactual dependence in explanations with explanantia and/ or explananda that do not exhibit asymmetry of overdetermination. Similar problems would arise for other non-causal explanations with explanantia and/ or explananda that are not events, and hence do not exhibit asymmetry of overdetermination. The problem in the above

[25] Similarly, the violation of law in W_4 might occur later since the dimensionality of space also does not occur at any particular time.

[26] Similar problems arise for establishing that CF_4 is true. One might object to my argument: Would W_5 not be an impossible world and thus excluded as Lewisian semantics not typically considers impossible worlds (see Lewis, 1986)? After all, one would argue, in a possible world with a miracle making stable planetary orbits impossible space is by necessity not three-dimensional. But in W_5 space is three-dimensional despite such a miracle. Hence, one might think, W_5 is impossible. In response, note that W_5 is not obviously mathematically, logically, or metaphysically impossible. More importantly, even if we were to exclude W_5 Lewisian semantics would not establish asymmetry of counterfactual dependence in our example. If all possible worlds were such that either stable planetary orbits are possible and space three-dimensional or stable planetary orbits are impossible and space not three-dimensional, then CF_5 and CF_4 are both true according to Lewisian semantics. Lewisian semantics would not yield asymmetry of counterfactual dependence.

example lies in the possibility of stable planetary orbits and the dimensionality of space not being events located in time. Due to this, they are not subject to the asymmetry of overdetermination that would make W_5 more similar than W_4. The same problem would arise for other explanations involving explanantia and/ or explananda that are not events.

I conclude that it is unclear how a conservative CEXP with Lewisian semantics would solve the challenge of asymmetry. This holds for strongly and weakly monist proponents of CEXP. Both need some qualification of CEXP such as to account for explanatory asymmetry in non-causal explanations. As just argued, it is unclear how Lewisian semantics would provide this qualification in the case of non-causal explanations that do not involve events.[27]

My argument is not intended to conclusively undermine the prospects of adding some version of Lewisian semantics to CEXP that establishes asymmetry of counterfactual dependence in non-causal explanations. One could try to develop a view according to which the possibility of stable planetary orbits, Königsberg's non-traversability and so on are events located in time (though I doubt this is tenable). One could also propose a similarity measure yielding asymmetry of counterfactual dependence in non-causal explanations not involving events. However, in the absence of such alternatives, I suggest looking elsewhere for a solution to the challenge of asymmetry. The next chapter investigates the prospects of two principles for asymmetry in non-causal explanations.

[27] Within a weakly monist solution, one could see Lewisian semantics as qualifying CEXP such as to account for asymmetry in causal explanations, which involve events (one would argue). In this case, for a full weakly monist solution, one would need to add some other qualification apt to yield asymmetry in non-causal explanations. However, this strategy would leave open what this other qualification would be. At best, Lewisian semantics would provide a partly weakly monist solution, leaving the difficult part (how to account for asymmetry in non-causal explanations) open.

5 Against two inference schemes

In the last chapter, I argued that it is unclear how a conservative CEXP with Lewisian semantics would solve the challenge of asymmetry. In this chapter, I discuss two principles to identify explanatory asymmetry, Woodward's (2018, 2020) double explanation principle and Lange's (2019) modal fact principle. In Section 5.1, I introduce both suggestions. I argue that they can be understood as inference schemes, i.e. as telling us when to infer that X explains Y and not vice versa. In the literature, it is left open how one would use either inference scheme to amend CEXP to solve the challenge of asymmetry. In Section 5.2, I therefore first distinguish between a conservative and a radical version of CEXP resulting from distinct ways of using the inference schemes to qualify CEXP. I then argue that neither of these versions solves the challenge of asymmetry.

5.1 Two inference schemes

5.1.1 The double explanation principle

Woodward aims to identify asymmetry in (some) non-causal explanations using what I call the double explanation principle, as also illustrated in Figure 6 (Woodward, 2018, p. 128, 2020, p. 45ff):

The double explanation principle

If (i) Z causally explains X, (ii) X and Y are statistically dependent, (iii) Y and Z are statistically dependent and (iv) "*in the absence of some further explanation of these dependencies*" (Woodward, 2020, p. 46), X explains Y and not vice versa.

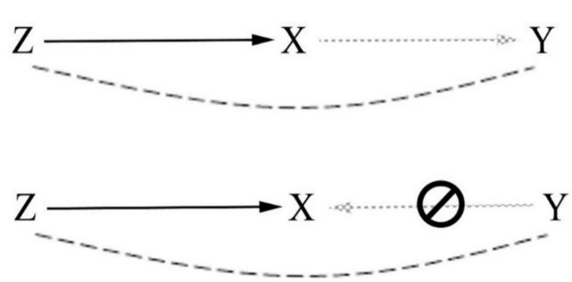


Figure 6: *Illustration of Woodward's double explanation principle*.
X, Y, Z denote variables. Solid arrows denote causal relations, dashed arrows non-causal relations and dashed lines to be explained correlations.

To illustrate the double explanation principle, Woodward uses Euler's explanation (Woodward, 2020, p. 43ff). He understands the explanandum Y to represent Königsberg's (non-)traversability, and the explanans X to represent that Königberg's bridge system is or is not isomorphic to an Eulerian graph. For him, X and Y are statistically dependent. Furthermore, Woodward thinks that X is causally explained by the intentions of those building the bridge system (Z). Moreover, according to him, the builders' intentions (Z) and the bridge system's non-traversability (Y) are statistically dependent. Given this situation and in the absence of other explanations for these dependencies, we should, says Woodward, infer that the non-isomorphic configuration of the bridge system (X) explains its non-traversability (Y). By contrast, Woodward argues, we should *not* infer that the non-traversability (Y) explains the non-isomorphic configuration (X). This would leave the non-isomorphic configuration with two explanations (a causal and a non-causal one) while not explaining the dependence of the bridge builders' intentions and Königsberg's non-traversability.

The qualification that other explanations for the correlation between Y and Z should be absent is intended to allow for some explananda to have a causal *and* a non-causal explanation (Woodward, 2020, p. 46f). An explanandum is allowed to be explained causally and non-causally if there is an alternative explanation for the correlation of both explanantia. For instance, if one explanans describes a system at a fine-grained level and the other at a more coarse-grained one this explains their correlation. In such a situation, the double explanation principle does not demand to infer that the explanandum is not explained causally *and* non-causally.

Woodward does not explain further how the double explanation principle aims to account for explanatory asymmetry. I argue that it can be understood as an inference scheme, i.e. as telling us when to infer that X explains Y and not vice versa. To see this, consider the following counterfactuals in Euler's explanation:

(CF_6) If the configuration of Königsberg's bridge system had changed to be isomorphic to an Eulerian graph (X), Königsberg's bridge system would have become traversable (Y).

(CF_7) If Königsberg's bridge system had changed to be traversable (Y), its configuration would have become isomorphic to an Eulerian graph (X).

(CF_8) If the bridge builders' intentions (Z) had been different, the configuration of Königsberg's bridge system (X) would have been different.

(CF_9) If the bridge builders' intentions (Z) had been different, Königsberg's bridge system would have been traversable (Y).

(CF_{10}) If Königsberg's bridge system had changed to be traversable (Y), the bridge builders' intentions (Z) would have been different.

As Woodward understands Euler's explanation, these counterfactuals are all true. This implies that the antecedent of the double explanation principle is satisfied: One explanation for the truth of CF_8 is that the bridge builders' intentions causally explain the bridge system's configuration, as (i) requires. And Y depends on X (CF_6) and vice versa (CF_7), as (ii) requires. Moreover, Z depends on Y (CF_{10}) and vice versa (CF_9), as (iii) requires. And Woodward assumes that there are no other explanations for the truth of CF_9 and CF_{10}, as (iv) requires. Given (i) – (iv), CF_6 and CF_8 jointly explain why CF_9 and CF_{10} are true, but CF_7 and CF_8 do not. Based on this observation, the double explanation principle tells us that CF_6 and not CF_7 is explanatory. The non-isomorphic configuration of the bridge system explains its non-traversability and not vice versa.

This reconstruction of how Woodward's double explanation principle picks out explanatory counterfactuals illustrates that it can be understood as an inference scheme, i.e. as telling us when to infer that X explains Y and not vice versa. This is because it refers to a set of counterfactuals based on which it tells us to infer that CF_6 and not CF_7 is explanatory. In other words, it tells us to infer that the non-isomorphic configuration (X) explains its non-traversability (Y) and not vice versa. Note also that the double explanation principle does not claim CF_7 to be false. I return to this aspect in Section 5.2. Now, I turn to the second principle proposed to identify asymmetry in non-causal explanations.[28]

[28] As proposed by Woodward, the double explanation principle fails to apply to non-causal explanations in which the explanans X cannot be a causal effect, e.g., the dimensionality of space in the explanation of the possibility of stable planetary orbits. To circumvent this restriction, one could allow for Z to also non-causally explain X. This seems to uphold the idea of the double explanation principle, that we ought to infer that X explains Y and not vice versa if a correlation of Z and Y would otherwise be unexplained. My arguments against using the double explanation principle to solve the challenge of asymmetry in Section 5.2 also hold for such an amended version of the principle.

5.1.2 The modal fact principle

Woodward argues that the (im)possibility of some phenomenon cannot be the target of an intervention, e.g., Königsberg's non-travers*ability* (Woodward, 2018, p. 123, footnote 4). Lange (2019) suggests that proponents of CEXP could propose the following principle to identify asymmetry in non-causal explanations based on Woodward's assertion:[29]

The modal fact principle
"[T]he explanans is eligible to be the target of an intervention whereas the explanandum is not" (Lange, 2019, p. 3899). Moreover, modal facts cannot be the target of an intervention.

According to Lange, a proponent of CEXP could aim to use the modal fact principle to identify Euler's explanation as asymmetric (Lange, 2019, p. 3899). Here, the explanandum is a modal fact (the non-travers*ability*) while the explanans (the bridge system's configuration) is not.

Just like the double explanation principle, we can understand the modal fact principle as an inference scheme. To illustrate this, consider again the following counterfactuals in Euler's explanation:

(CF_6) If the configuration of Königsberg's bridge system had changed to be isomorphic to an Eulerian graph, Königsberg's bridge system would have become traversable.

(CF_7) If Königsberg's bridge system had changed to be traversable, its configuration would have become isomorphic to an Eulerian graph.

The modal fact principle tells us to hold CF_6 and not CF_7 explanatory based on which antecedent can be intervened on: One can intervene on the bridge system's configuration but not on its non-traversability, or so one would argue. Hence, CF_6 is explanatory while CF_7 is not. This illustrates that the modal fact principle can be seen as an inference scheme. Based on features of several counterfactuals it tells us to infer that the bridge system's non-isomorphic configuration explains its non-traversability but not vice versa. Note also that just like the double explanation principle, the modal fact principle does not claim CF_7 to be false.

[29] Lange (2019) argues against the modal fact principle. I return to his argument in Section 5.2.3.

5.2 Against the inference schemes

In the last section, I introduced two principles for explanatory asymmetry. I argued that they can be understood as inference schemes telling us when to hold a counterfactual dependence as explanatory. Notably, Woodward and Lange do not discuss how one would amend CEXP using either inference scheme to solve the challenge of asymmetry. In this section, I first distinguish between a conservative and a radical version of CEXP resulting from distinct ways of using the inference schemes to qualify CEXP. I then argue that neither of these versions solves the challenge of asymmetry.

5.2.1 Two versions of CEXP with the inference schemes

To address the challenge of asymmetry, one might propose the following versions of CEXP, each of which using the double explanation principle and/ or the modal fact principle in different ways to qualify CEXP:

1) **Conservative CEXP**: One upholds that the counterfactual dependence of Y on changes in X as referred to in $CExp_1$ – $CExp_3$ is necessary and sufficient for X to explain Y. One then uses either inference scheme, both, or a combination of them to qualify which counterfactual dependence $CExp_2$ and/ or $CExp_3$ should refer to. More precisely, in this view, $CExp_2$ and/ or $CExp_3$ should refer to those counterfactuals which either inference scheme or a combination of both identifies as explanatory.
2) **Radical CEXP**: One adds either inference scheme, both or a combination of them as further necessary conditions X and Y must satisfy for X to explain Y to CEXP.

In the following, I argue that neither version of CEXP solves the challenge of asymmetry.

5.2.2 Against a conservative CEXP with the inference schemes

A conservative CEXP which refers to those counterfactuals either inference scheme or a combination of both hold to be explanatory does not solve the challenge of asymmetry. Here is my argument for this claim:

A conservative CEXP claims that X explains Y iff Y counterfactually depends on changes in X. Moreover, some counterfactual dependence as expressed in counterfactuals should be non-explanatory. For instance, the counterfactual dependence of the bridge system’s configuration on its non-traversability as expressed in CF_7 should be non-explanatory. To be non-explanatory according to a conservative CEXP, a counterfactual must be false. For instance, to be non-explanatory CF_7 must be false.

However, neither the double explanation principle nor the modal fact principle entails that a counterfactual is false. For instance, they do not entail that CF_7 is false (see Section 5.1). As a result, even if a conservative CEXP only refers to counterfactuals either inference scheme or some combination of both claims to be explanatory it fails to imply that other counterfactuals are false. Hence, it fails to imply that they are non-explanatory. For example, even if a conservative CEXP only refers to CF_6 it fails to imply that CF_7 is false. Hence, it fails to imply that CF_7 is non-explanatory.

Crucially, a conservative CEXP with the inference schemes then fails to account for asymmetry in some explanations: it identifies a counterfactual dependence of Y on X as explanatory but fails to identify a counterfactual dependence of X on Y as non-explanatory. For instance, such a conservative CEXP fails to identify Euler's explanation as asymmetric. It claims that the bridge system's configuration explains its non-traversability. But it does not claim that its non-traversability does not explain its configuration. Thus, it does not identify Euler's explanation as asymmetric. I conclude that a conservative CEXP with either inference scheme or a combination of both does not solve the challenge of asymmetry.[30]

5.2.3 Against a radical CEXP with the inference schemes

Instead of using the inference schemes in a conservative CEXP, one could use either, both or a combination of them as further necessary conditions X and Y must satisfy for X to explain Y. Such a radical CEXP implies that, even if X counterfactually depends on Y, Y does not explain X in some cases. Thereby, it avoids the argument against a conservative CEXP with the inference schemes. Even if neither inference scheme establishes that X does not counterfactually depend on

[30] To encounter my argument, one might want to add the following condition to CEXP:

($CExp_4$) If $CExp_1$ – $CExp_3$ do not entail that X explains Y, then X does not explain Y.

A conservative CEXP with the inference schemes does *not* claim that CF_7 is explanatory. Hence, it does not claim that the bridge system's non-traversability explains its configuration. Adding $CExp_4$ would lead to a new version of CEXP. The resulting theory would claim that the bridge system's non-traversability does *not* explain its configuration, yielding asymmetry in Euler's explanation.
However, the sketched version of CEXP with $CExp_4$ would be an untenable position: Either inference scheme fails to identify some explanatory counterfactual dependence of Y on X, as I argue in Section 5.2.3. In such cases, this version of CEXP would not only fail to claim that X explains Y but also (due to $CExp_4$) falsely claim that X does not explain Y.

Y a radical CEXP with the inference schemes might still correctly claim that Y does not explain X but only vice versa. However, neither a radical CEXP with either inference scheme nor one with both or a combination of them solves the challenge of asymmetry, as I argue below.

Against a radical CEXP with the double explanation principle

The double explanation principle would add the following necessary condition to CEXP:

Double explanation condition

X explains Y only if (i) Z causally explains X, (ii) X and Y are statistically dependent, (iii) Y and Z are statistically dependent and (iv) there are no other explanations of these dependencies.

However, a radical CEXP with the double explanation condition fails to provide a strongly or weakly monist solution to the challenge of asymmetry. To provide such a solution, a radical CEXP must claim that X explains Y but not vice versa either in causal and non-causal explanations (strongly monist) or at least in non-causal ones (weakly monist).[31] However, a radical CEXP with the double explanation condition does not do so. Instead, in some non-causal explanations, such a radical CEXP wrongly claims that X does not explain Y. This happens if (i), (iii) or (iv) are not satisfied despite X explaining Y.

As an example, consider Euler's explanation, the case Woodward claims his double explanation principle could account for. He takes the bridge system's non-isomorphic configuration (X) to explain its non-traversability (Y). However, (iv) is not satisfied in this explanation (see also Elliott & Lange, 2021, p. 9ff): Here, the bridge builders' intentions (Z) explain the bridge system's non-traversability (Y), beyond the bridge system's non-isomorphic configuration (X) explaining its non-traversability (Y). This is because the bridge builders' intentions explain the bridge system's non-isomorphic configuration *and thereby* also its non-traversability. The reason is that the bridge system's non-traversability and whether it is isomorphic to an Eulerian graph are necessarily related via Euler's theorem. Due to this, the bridge builders' intentions cannot bring about the latter without bringing about the former. In other words, their intentions explain both the bridge system's

[31] In the latter case, to provide a full weakly monist solution, one would need to add some other condition to CEXP apt to account for asymmetry in causal explanations. Note that a radical CEXP with the double explanation condition could also be seen as part of a weakly monist solution to the challenge of asymmetry if accounting for the asymmetry of causal explanations. In this case, for a full weakly monist solution, one would need to add some other condition apt to yield asymmetry in non-causal explanations. I will not discuss this option as it would leave the difficult part (how to account for asymmetry in non-causal explanations) open.

configuration and its non-traversability. Thus, there is another explanation for the dependence of the bridge builders' intentions and non-traversability, beyond X explaining Y. Hence, (iv) is not satisfied here. Therefore, X and Y do not satisfy the necessary double explanation condition. Therefore, a radical CEXP with the double explanation condition falsely claims that the bridge system's non-isomorphic configuration does not explain its non-traversability.

I conclude that a radical CEXP with the double explanation condition wrongly claims X to not explain Y in some non-causal explanations. Hence, it fails to provide a strongly or weakly monist solution to the challenge of asymmetry.

Against a radical CEXP with the modal fact principle

The modal fact principle would add the following necessary condition to CEXP:

Modal fact condition

X explains Y only if Y is a modal fact (and thus cannot be the target of an intervention) but X is not (and thus can be the target of an intervention).

However, a radical CEXP with the modal fact condition does not provide a strongly or weakly monist solution to the challenge of asymmetry. To provide such a solution, again, such a radical CEXP must claim that X explains Y but not vice versa either in causal and non-causal explanations or at least in non-causal ones. However, such a radical CEXP wrongly claims that X does not explain Y in some causal and non-causal explanations. This is because in these cases either X and Y are both modal facts or both are not modal facts. If so, then X and Y fail to satisfy the modal fact condition.[32]

[32] Lange also notes that the modal fact principle fails to yield explanatory asymmetry if X and Y are both modal facts (Lange, 2019, p. 3899f). However, he only criticizes the principle as a sufficient condition for asymmetry in non-causal explanations. And proponents of CEXP aim to provide necessary and jointly sufficient conditions for X to explain Y (see Section 3.1). Adding the modal fact principle as a sufficient condition does not uphold this commitment. For this reason, Lange's criticism does not suffice to undermine the modal fact principle as a solution to the challenge of asymmetry.

Here are two examples: First, recall, scientists explain the anisotropic fluid behaviour of liquid crystals by referring to the shape of their molecules (see Section 2.2.3). Neither the shape of the molecules nor the fluid behaviour is a modal fact. A radical CEXP with the modal fact condition wrongly claims that the shape of the molecules does not explain the fluid behaviour. Similarly, the flagpole's height, its shadow's length and the angle of sunrays are all not modal facts. A radical CEXP with the modal fact condition wrongly claims that the angle of sunrays and the flagpole's height do not explain its shadow's length. I conclude that a radical CEXP with the modal fact condition wrongly claims that X does not explain Y in some causal and non-causal explanations. Hence, it does not provide a strongly or weakly monist solution to the challenge of asymmetry.

In response, one might want to generalize the modal fact condition:

Generalized modal fact condition

X explains Y only if Y cannot be the target of an intervention, but X can be the target of an intervention.

However, a radical CEXP with the generalized modal fact condition likewise fails as a strongly or weakly monist solution to the challenge of asymmetry. Again, some causal and non-causal explanations do not satisfy the generalized modal fact condition. As a result, a radical CEXP with this condition wrongly claims that X does not explain Y in these cases.

Here are two examples: First, if at all, both the fluid behaviour of liquid crystals and the shape of their molecules can be intervened on. A radical CEXP with the generalized modal fact condition claims that the shape of the molecules does not explain the fluid behaviour. Second, the angle of sunrays, the flagpole's height and its shadow's length can all be targets of an intervention. Again, a radical CEXP with the generalized modal fact condition claims that the angle of sunrays and the flagpole's height do not explain its shadow's length. I conclude that such a radical CEXP wrongly claims that X does not explain Y in some causal and non-causal explanations. Thus, it fails to solve the challenge of asymmetry in a strongly or weakly monist way.

Against a radical CEXP with both inference schemes

Instead of adding either inference scheme as a necessary condition to CEXP one might propose to add both or a combination of both as necessary conditions to CEXP. However, such a version of a radical CEXP also does not provide a strongly or weakly monist solution to the challenge of asymmetry:

First, one could add both the double explanation condition and the generalized modal fact condition individually. However, the resulting radical CEXP does not avoid the problems posed against a radical CEXP with either condition. In some causal and non-causal explanations, X and Y do not satisfy the generalized modal fact condition and/ or the double explanation condition even if X explains Y. As a result, a radical CEXP with both conditions wrongly claims that X does not explain Y in some causal and non-causal explanations. Hence, it does not solve the challenge of asymmetry in a strongly or weakly monist way.

Second, one could propose to combine both conditions:

Combined condition

X explains Y only if (i) Z causally explains X, (ii) X and Y are statistically dependent, (iii) Y and Z are statistically dependent, (iv) there are no other explanations of these dependencies and (v) Y cannot be the target of an intervention but X can be the target of an intervention.

However, as argued above, in some causal and non-causal explanations, X explains Y despite X and Y not satisfying (i), (iii), (iv) or (v). A radical CEXP with this combined condition wrongly claims that X does not explain Y in these cases. Hence, it does not provide a strongly or weakly monist solution to the challenge of asymmetry. I conclude that neither a radical CEXP with both inference schemes nor one with a combination of both solves the challenge of asymmetry.[33]

Before ending my discussion of the two inference schemes, I want to mention another objection against using them to solve the challenge of asymmetry: Rather than providing a condition for X to explain Y and not vice versa, these inference schemes seem to merely "justify our belief about the [explanation's, IJ] directionality" (Elliott & Lange, 2021, p. 9f). At the very least, there is no

[33] Could one solve the challenge of asymmetry by using either inference scheme or both as indicating when further necessary conditions for X to explain Y are satisfied? One could not. Even if the inference schemes would correctly indicate when such further conditions are satisfied, they would leave open what these further conditions are. Without specifying these further conditions, such a version of CEXP would remain incomplete.

argument for these inference schemes to do any other work than telling us when to believe that X explains Y rather than vice versa. Woodward (2018, 2020) does not provide an argument why it would be a plausible condition for X to explain Y that X is also explained by Z. Similarly, we lack an argument why it would be a plausible condition for X to explain Y that X but not Y can be the target of an intervention, within a theory that per se dispenses with interventions. Both inference schemes could equally well play a purely epistemic role in our explanatory reasoning rather than constituting conditions for explanatory power.

To summarize, I argued that a conservative CEXP using either or both inference schemes fails to solve the challenge of asymmetry. Moreover, I argued that neither a radical CEXP using each inference scheme individually nor one using both or a combination of them solve this challenge.

Let me take stock: In this part, I rejected two potential solutions to the challenge of asymmetry, an appeal to Lewisian semantics and to two inference schemes, Woodward's (2018, 2020) double explanation and Lange's (2019) modal fact principle. Beyond these two strategies, some other avenues one might suggest as solutions to the challenge of asymmetry are not promising for proponents of CEXP. These either conflict with core commitments of the interventionist theory (and thus presumably of CEXP) or struggle with being generalizable to all explanations proponents of CEXP are interested in. As they are not promising, I do not discuss these strategies here, but elaborate on their shortcomings in Appendix C. The upshot of Part II is that proponents of CEXP need a new avenue to deal with explanatory asymmetry. In Part III, I propose one.

Part III Towards a quasi-interventionist theory of explanation

In this part, I develop a quasi-interventionist version of CEXP which accounts for the asymmetry of causal and many non-causal explanations and the symmetry of some non-causal explanations. I develop my view in three chapters:

In Chapter 6, I argue that we can narrow the scope of the challenge of asymmetry. In short, if an explanans X and an explanandum Y relate mathematically and with a so-called many-to-one relation between their actual values, then CEXP already entails that (if at all) X explains Y and not vice versa. However, CEXP can still fail to yield explanatory asymmetry if X and Y do not relate mathematically and with a many-to-one relation between their actual values.

In Chapter 7, I develop a quasi-interventionist version of CEXP, accounting for the asymmetry in most of these remaining explanations. Combining a proposal by Woodward (2018) and the notion of an intervention, I develop the notion of a quasi-intervention. In my quasi-interventionist CEXP, $CExp_2$ and $CExp_3$ are qualified such that X explains Y iff Y counterfactually depends on changes in X brought about by a quasi-intervention.

In Chapter 8, I then propose that the quasi-interventionist CEXP is a promising solution to the revised challenge of (a)symmetry rather than the challenge of asymmetry. To this end, I show that mathematical explanations with one-to-one relations between their actual values are symmetric according to this theory. I propose to accept this result and outline open questions on the quasi-interventionist CEXP as solving the revised challenge. Finally, I argue that both approaches discussed in Part II, amendments of CEXP using Lewisian semantics or the inference schemes, do not solve the revised challenge.

6 Narrowing the scope of the challenge

In this chapter, I argue that we can narrow the scope of the challenge of asymmetry. In short, if an explanans X and an explanandum Y relate mathematically and with a many-to-one relation between their actual values, then CEXP already entails that (if at all) X explains Y but not vice versa. To argue for this claim, I first introduce Jansson and Saatsi's (2019) proposal to account for asymmetry in non-causal explanations, the fixing principle (Section 6.1). I show, pace Jansson and Saatsi, that the asymmetry the fixing principle alerts to can be understood as an asymmetric dependence of the actual values of X and Y. Then, I argue that this so-called actual value asymmetry holds if X and Y relate mathematically and with a many-to-one relation between their actual values (Section 6.2). Finally, I identify the implication of this observation for the challenge of asymmetry (Section 6.3): We do not need to account for the asymmetry of mathematical explanations with a many-to-one relation between their actual values. As a preliminary remark, my argument in this chapter is restricted to explanations in which the explanans X is a single variable. I leave an extension to explanations with several explanantia to further research.

6.1 The fixing principle and actual value asymmetry

To account for the asymmetry of (some) non-causal explanations, Jansson and Saatsi propose the following principle (Jansson & Saatsi, 2019, p. 832ff):

The fixing principle

"Fixing the explanans variable to its actual value should fix the explanandum variable to its actual value" (Jansson & Saatsi, 2019, p. 833). But fixing the explanandum variable to its actual value does not fix the explanans variable to its actual value (Jansson & Saatsi, 2019, p. 833).

To illustrate, recall Euler's explanation. For Jansson and Saatsi, the explanandum is the bridge system's non-traversability (Y) and the explanans the bridge system's configuration (X) (Jansson & Saatsi, 2019, p. 833). According to them, fixing Königsberg's bridge system to have four parts oddly connected by seven bridges (its actual configuration) fixes the bridge system's non-traversability. But, they argue, fixing the bridge system's non-traversability does *not* fix the bridge system to have its actual configuration. This is because the bridge system could be non-traversable even if it would have a different configuration (e.g., if it would have five parts oddly connected by ten bridges). Thus, Jansson and Saatsi conclude, the fixing principle correctly entails that the bridge system's configuration explains its non-traversability and not vice versa.

Jansson and Saatsi do not specify what is supposed to count as a 'fix' in their principle. Instead of specifying what 'fixing' means, we can dispense with any mention of fixing and still identify the same asymmetry between the actual values of an explanans and an explanandum. To see this, consider the following asymmetry involved in Euler's explanation:

According to Euler's theorem, it is true that

(1) If the bridge system has its actual configuration, then it is non-traversable.

But according to Euler's theorem, it is false that

(2) If the bridge system is non-traversable, then it has its actual configuration.

(2) is false because many different configurations are non-traversable, according to Euler's theorem. The asymmetry between (1) and (2) then seems to be exactly the asymmetry the fixing principle alerts to. And it persists without any mentioning of fixing.

We can define the asymmetry identified in Euler's explanation in general terms:

Actual value asymmetry

It is true that if X takes on its actual value x, Y takes on its actual value y.

But it is false that if Y takes on its actual value y, X takes on its actual value x.

The actual value asymmetry in Euler's explanation is not a spurious feature of this explanation. In the next section, I argue that actual value asymmetry holds if an explanans and an explanandum relate mathematically and with a many-to-one relation between their actual values.

6.2 When actual value asymmetry holds

To see when actual value asymmetry holds, let us clarify how the actual values of an explanans X and an explanandum Y can relate within CEXP (see also Figure 7):

Four possible relations between the actual values of X and Y

Many-to-many: The actual value of X relates to more than only the actual value of Y. And the actual value of Y relates to more than only the actual value of X.

Many-to-one: The actual value of X only relates to the actual value of Y. And the actual value of Y relates to more than only the actual value of X.

One-to-many: The actual value of Y only relates to the actual value of X. And the actual value of X relates to more than only the actual value of Y.

One-to-one: The actual value of X only relates to the actual value of Y. And the actual value of Y only relates to the actual value of X.

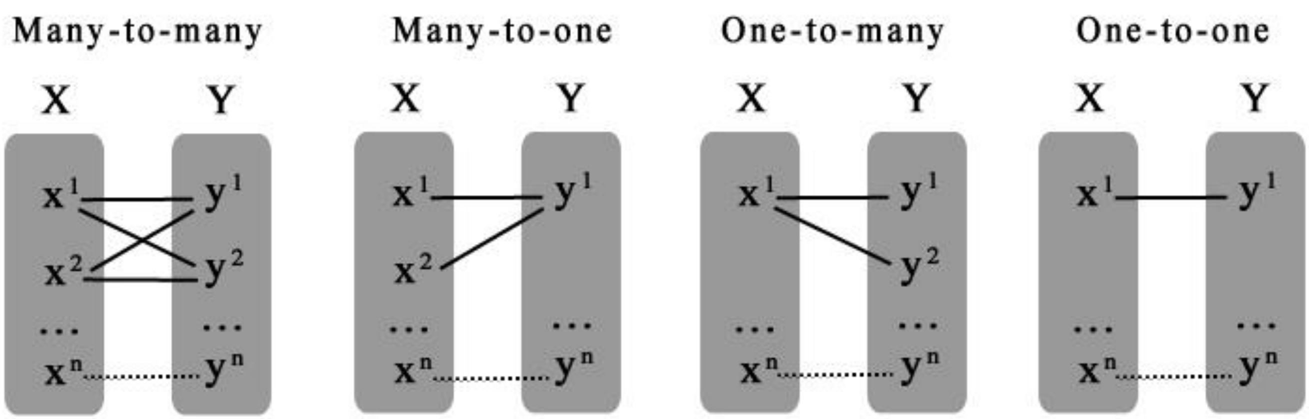


Figure 7: *Four possible relations between the actual values of X and Y.* Boxes denote variables with x^idenoting the values of X and y^i denoting the values of Y. x^1 and y^1 denote the actual values of X and Y. Lines denote (causal or non-causal) relations. Dashed lines denote (causal or non-causal) relations not further specified.

To illustrate, consider Euler's explanation, as understood by Jansson and Saatsi. Here, the explanans X, the bridge system's configuration, can take on many values x^i representing different configurations. x^1 represents the actual configuration. The explanandum Y can take on its actual value y^1 (non-traversable) or y^2 (traversable). Here, the actual values of X and Y relate many-to-one. The actual configuration of the bridge system is non-traversable, and there is more than one configuration that is non-traversable.

Now, let us return to the actual value asymmetry in Euler's explanation. It holds because the actual values in Euler's explanation relate many-to-one: Actual value asymmetry says that if X takes on its actual value x, Y takes on its actual value y, and that it is false that if Y takes on its actual value y, X takes on its actual value x. If Königsberg's bridge system has its actual configuration, then it is non-traversable. But it is false that if it is non-traversable, it has its actual configuration. This holds because the actual values of X and Y relate many-to-one here: There is more than one value X could take on if Y takes on y, but only one value Y takes on if X takes on x. There is more than one configuration a non-traversable bridge system could have, but the actual configuration is only non-traversable. Thus, the actual value asymmetry in Euler's explanation arises because the actual values of X and Y relate many-to-one.

Can we generalize the insight described above such that causal and non-causal relations alike can exhibit a many-to-one relation between their actual values that implies actual value asymmetry? I will not argue for such a generalization here. Instead, I only claim that actual value asymmetry holds if the actual values of X and Y relate many-to-one *and* if X and Y relate mathematically.[34] Below, I explain the requirement for a mathematical relation.

By X and Y relating mathematically, I mean that X's taking on some value necessitates for mathematical reasons that Y takes on some value and vice versa. For example, in Euler's explanation, the relationship between the bridge system's configuration (X) and its (non-) traversability (Y) is mathematical. As Euler's theorem tells us, the bridge system's configuration necessitates it being either traversable or non-traversable. And its traversability necessitates it having some configuration isomorphic to an Eulerian graph, and its non-traversability a non-isomorphic configuration. Similarly, in the strawberry explanation (see Section 2.2.1), the number of strawberries a mother with three children possesses (X) relates mathematically to their (in)divisibility amongst the mother's three children (Y). The number of strawberries the mother possesses necessitates either their divisibility or indivisibility amongst her three children. And their divisibility necessitates her having some number of strawberries that is a multiple of three, and their indivisibility some number that is not a multiple of three.

If X and Y relate mathematically in the described way and with a many-to-one relation between their actual values, then actual value asymmetry holds. In such cases, that X takes on its actual

[34] Any extension beyond these conditions will be left for further research.

value x necessitates that Y takes on its actual value y. And that Y takes on its actual value y necessitates that X takes on some value but not necessarily its actual value x. If Königsberg's bridge system has its actual configuration, it is necessarily non-traversable. This holds because the actual configuration is, by mathematical necessity, non-traversable. Moreover, it is false that if Königsberg's bridge system is non-traversable, it has its actual configuration. This holds because its non-traversability necessitates that it has *some* configuration not isomorphic to an Eulerian graph, but not which one. If X and Y relate mathematically and with a many-to-one relation between their actual values, then actual value asymmetry holds. In the next section, I argue that this observation limits the challenge of asymmetry to explanations lacking these features.[35]

6.3 Implication for the challenge of asymmetry

In the last section, I argued that if X and Y relate mathematically and with a many-to-one relation between their actual values, then actual value asymmetry holds. Based on this observation, we can narrow the scope of the challenge of asymmetry:

If X and Y relate mathematically and with a many-to-one relation between their actual values, even an unqualified CEXP entails that (if at all) X explains Y but not vice versa. Thus, for these cases, there is no challenge of asymmetry. To see this, recall the second necessary condition for X to explain Y in CEXP (see Section 2.3):

> ($CExp_2$) (G describing) the relationship between X and Y correctly describes the actual value y of Y when X takes on its actual value x.

Actual value asymmetry between X and Y implies that $CExp_2$ is satisfied in the X to Y direction. If X takes on its actual value x, Y takes on its actual value y. And actual value asymmetry also implies that $CExp_2$ is not satisfied in the Y to X direction. It is false that if Y takes on its actual value y, X takes on its actual value x. Thus, if actual value asymmetry holds, then CEXP already

[35] Does actual value asymmetry also hold if X and Y relate mathematically and with some other relation between their actual values? First, if the actual values of X and Y relate one-to-many actual value asymmetry holds but in the reverse direction. It is false that if X takes on its actual value x, Y takes on its actual value y, but true that if Y takes on its actual value y, X takes on its actual value x. Second, if the actual values of X and Y relate many-to-many, then actual value asymmetry does not hold. It is false that if X takes on its actual value x, Y takes on its actual value y. And it is false that if Y takes on its actual value y, X takes on its actual value x. Third, if the actual values of X and Y relate one-to-one, then actual value asymmetry does not hold either. It is true that if X takes on its actual value x, Y takes on its actual value y. And it is true that if Y takes on its actual value y, X takes on its actual value x.

entails that (if at all) X explains Y but not vice versa. If at all, the bridge system's configuration explains its non-traversability but not vice versa.[36]

Narrowing the scope of the challenge of asymmetry does not solve it completely.[37] For non-mathematical explanations and mathematical ones without a many-to-one relation between the actual values of X and Y, CEXP can still fail to yield explanatory asymmetry.[38] We still need a solution to the challenge of asymmetry in these cases. Moreover, this solution must uphold that $CExp_2$ is necessary for X to explain Y. Otherwise, actual value asymmetry does not ensure explanatory asymmetry if X and Y relate mathematically and many-to-one between their actual values. In the next chapter, I develop such a qualified CEXP.[39]

[36] Jansson and Saatsi also note that their fixing principle has this implication (Jansson & Saatsi, 2019, p. 833)..

[37] Can we narrow the scope of the challenge further including also other non-causal or causal relations? As mentioned, I leave this open for further research. Note that even if we could extend the insight from this chapter to causal and other non-causal relations, we would still have to solve the challenge of asymmetry for cases in which the actual values of X and Y do not relate many-to-one. Chapter 7 provides a way to do so.

[38] Can we not also narrow the scope of the challenge in the case of other possible relations between the actual values of X and Y if they relate mathematically? No. First, if the actual values of X and Y relate one-to-many then actual value asymmetry holds in the reverse direction (see footnote 35). Here, CEXP entails that (if at all) Y explains X and not vice versa, yielding asymmetry but in the wrong direction. Second, if the actual values of X and Y relate many-to-many then CExp2 is not satisfied in either direction. CEXP entails that X does not explain Y and vice versa. Third, if the actual values of X and Y relate one-to-one, then CExp2 is satisfied in both directions, and CEXP does not yield asymmetry. I discuss this last case in Chapter 8.

[39] Lange poses counterexamples against Jansson and Saatsi's fixing principle (Lange, 2019, p. 3902f). However, his counterexamples do not pose a threat to my argumentation. Some do not respect the requirement that X and Y be fixed *to their actual values*. Others pertain to explanations which do not relate many-to-one between their actual values. Thus, these examples do not object to many-to-one relations between actual values implying actual value asymmetry and thereby explanatory asymmetry.

7 A quasi-interventionist theory of explanation

In this chapter, I propose a conservative, strongly monist version of CEXP, accounting for asymmetry in many non-causal and causal explanations. My quasi-interventionist CEXP qualifies $CExp_2$ and $CExp_3$ such that X explains Y iff Y counterfactually depends on changes in X brought about by a quasi-intervention. I develop this theory in five steps:

In Section 7.1, I introduce Woodward's (2018) proposal to identify asymmetry in non-causal explanations, the independence notion. I argue that one could aim to use the independence notion in a conservative CEXP, claiming that X explains Y iff Y counterfactually depends on changes in X that are independent in the sense of this notion. In Section 7.2, I then argue that Woodward's independence notion fails to yield asymmetry of counterfactual dependence in the flagpole explanation. As a result, a conservative CEXP relying on it fails to identify the flagpole explanation as asymmetric. In Section 7.3, I suggest overcoming this limitation by combining the independence notion with the notion of an intervention to define an initial notion of a quasi-intervention. The resulting conservative CEXP requires Y to counterfactually depend on changes in X brought about by a quasi-intervention as initially defined. In Section 7.4, I show that such a CEXP wrongly claims X to not explain Y if X and Y relate mathematically. In Section 7.5, I avoid this undesirable consequence by defining a second, and final, notion of a quasi-intervention and a corresponding quasi-interventionist CEXP.

7.1 The independence notion

Woodward proposes the following strategy to account for asymmetry in non-causal explanations (see also Figure 8):

The independence notion

"if X and Z are independent and Y is dependent on X and Z, then the direction of explanation runs from X and Z to Y" (Woodward, 2018, p. 125, footnote 7; notation changed)

The independence notion is best illustrated using the case Woodward derives it from, the explanation of the possibility of stable planetary orbits (Woodward, 2018, p. 124f, 2020, p. 47f).

According to Woodward, physicists use the dimensionality of space (X) jointly with assumptions about the form of gravitational potential and laws of motion (Z) to derive the possibility of stable planetary orbits (Y). However, physicists also derive the dimensionality of space from the possibility of stable planetary orbits jointly with assumptions about the form of gravitational potential and laws of motion. Woodward argues that some physicists deem the former rather than the latter derivation to be explanatory based on several (in)dependence assumptions. These physicists assume that the form of gravitational potential and laws of motion are independent of the dimensionality of space. And they assume that the form of gravitational potential and laws of motion depend on the possibility of stable planetary orbits. Based on these assumptions, so Woodward argues, these physicists see the dimensionality of space as explaining the possibility of stable planetary orbits and not vice versa. They think that a change in the explanans X (dimensionality of space) is independent of changes in the explanandum's other explainers Z (the form of gravitational potential and laws of motion). And they think that a change in the explanandum Y (possibility of stable planetary orbits) depends on changes in Z (the form of gravitational potential and laws of motion). Thus, these physicists conclude, the dimensionality of space (X) explains the possibility of stable planetary orbits (Y) and not vice versa.[40]

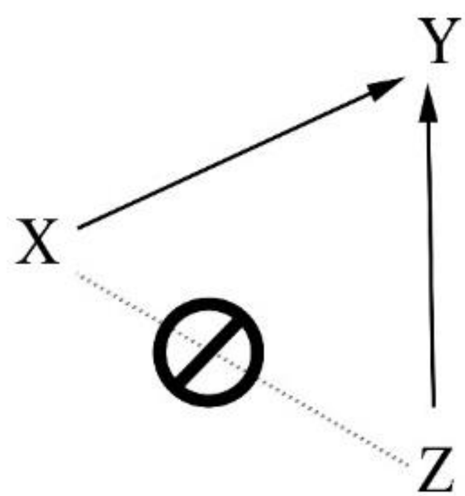


Figure 8*: Illustration of the independence notion*.
X, Y, Z denote variables. Solid arrows denote explanatory counterfactual dependence. The crossed dashed line denotes the independence of X and Z.

[40] According to Woodward, other physicists deem the possibility of stable planetary orbits to explain the dimensionality of space (Woodward, 2018, p. 125). These assume that the form of gravitational potential and laws of motions are independent of the possibility of stable planetary orbits but not of the dimensionality of space, contrary to those mentioned in the main text. Woodward suggests that there is no empirical basis of deciding which of these (in)dependence assumptions are true.

Woodward leaves open how one could use the independence notion to solve the challenge of asymmetry. We could aim to use it in a conservative CEXP: We uphold that a counterfactual dependence of Y on X is necessary and sufficient for X to explain Y. We then use the independence notion to qualify the antecedent of the relevant counterfactuals such that explanatory ones are true and non-explanatory ones false, yielding asymmetry of counterfactual dependence. As a result, this conservative CEXP implies explanatory asymmetry.

To see this conservative solution to the challenge of asymmetry at work, consider two counterfactuals involved in the above explanation:

(CF_4) If the dimensionality of space had been different, stable planetary orbits would have been impossible.

(CF_5) If stable planetary orbits had been impossible, the dimensionality of space would have been different.

The mathematical derivation physicists use supports the truth of both counterfactuals (Woodward, 2018, p. 124f). Therefore, an unqualified CEXP claims the dimensionality of space to explain the possibility of stable planetary orbits and vice versa, failing to yield asymmetry. By contrast, a conservative CEXP with the independence notion implies explanatory asymmetry in this example. To see this, consider CF_4 and CF_5 qualified with the independence notion:

(CF_4^{ip}) If the dimensionality of space had changed^{ip}, stable planetary orbits would have been impossible.

(CF_5^{ip}) If the possibility of stable planetary orbits had changed^{ip}, the dimensionality of space would have been different.

We can define change^{ip} as follows, expressing the independence notion (for a similar interpretation see Lange, 2019, p. 3900f):

(**Change^{ip}**) IP is a way to change^{ip} X with respect to Y iff
IP changes X such that any such change in X is independent of any change in Z which explains Y not via explaining X.[41]

Suppose we take the dimensionality of space as explaining the possibility of stable planetary orbits but not vice versa.[42] To establish this verdict, the proposed conservative CEXP requires that CF_4^{ip} is true for some changes in the dimensionality of space and CF_5^{ip} false for all changes in the possibility of stable planetary orbits. Under certain assumptions, this is the case. In particular, we assume that the laws of motion and the form of gravitational potential are independent of the dimensionality of space and dependent on the possibility of stable planetary orbits. And we assume that the possibility of stable planetary orbits depends on the dimensionality of space.

First, consider CF_4^{ip}. A way to change^{ip} the dimensionality of space is to change n (dimensionality of space) to 5 in the mathematical derivation. This is because changing n leaves the other explainers of the possibility of stable planetary orbits, namely laws of motion and form of gravitational potential, unchanged, according to our assumptions. The mathematical derivation tells us that stable planetary orbits are impossible if n=5. Therefore, CF_4^{ip} is true for at least some independent change in space's dimensionality. Hence, a conservative CEXP with the independence notion claims that the dimensionality of space explains the possibility of stable planetary orbits.

Second, consider CF_5^{ip}. Under the (in)dependence assumptions, one cannot change^{ip} the possibility of stable planetary orbits. To see this, we have to think about what Z, the other explainer of space's dimensionality, is here. Woodward's reconstruction of how some physicists judge explanatory directionality in this case suggests that Z would also be the laws of motion and form of gravitational potential (see also Lange, 2019, p. 3900f). After all, we can use the laws of motion and form of

[41] IP is a variable that represents whatever changes X. One need not reconstruct the independence notion as involving such a variable. However, there is no harm in spelling out what changes X, so long as one keeps the independence requirement that changes in X are independent of changes in Z. The motivation for introducing a variable IP will become clearer when developing the notion of a quasi-intervention (see Section 7.3).
[42] The below reconstruction of how a conservative CEXP with the independence notion yields explanatory asymmetry also holds for the reversed directionality some physicists defend, given their respective (in)dependence assumptions.

gravitational potential jointly with the possibility of stable planetary orbits to derive the dimensionality of space. If we accept that Z denotes the laws of motion and form of gravitational potential here, then any change to the possibility of stable planetary orbits fails the independence requirement. Given our assumptions, it is impossible to change the possibility of stable planetary orbits without changing the laws of motion and form of gravitational potential. Thus, $\mathrm{CF}_5^{\mathrm{ip}}$ has an impossible antecedent.

What does the impossible antecedent of $\mathrm{CF}_5^{\mathrm{ip}}$ imply about its truth value? The truth value of counterfactuals with impossible antecedents is disputed (see Berto & Jago, 2018). Instead of settling this debate here, I suggest that $\mathrm{CF}_5^{\mathrm{ip}}$ is at least not true in a way that would be needed for $\mathrm{CF}_5^{\mathrm{ip}}$ to be explanatory. For instance, we may say $\mathrm{CF}_5^{\mathrm{ip}}$ is not non-vacuously true, and that non-vacuous truth would be needed for $\mathrm{CF}_5^{\mathrm{ip}}$ to be explanatory. Hence, the independence notion implies an asymmetry of counterfactual dependence in the above example. And a conservative CEXP with the independence notion claims that the possibility of stable planetary orbits does not explain the dimensionality of space. It yields explanatory asymmetry in this explanation.

Before moving to my criticism of a conservative CEXP with the independence notion, one more clarification is important: How should we understand what Z, the other explainer of Y, may be? Notably, Woodward (2018, 2020) does not tell us what Z may be. Here are two suggestions:

(1) Lange suggests that Z are "auxiliary hypotheses" (Lange, 2019, p. 3917) used to derive Y from X, as in the above example.
(2) Z consists of other explainers of Y, that may (but need not) be used to derive Y from X.

I suggest opting for (2). (2) implies better prospects for the independence notion to apply to many explanations than (1) does. For instance, in Euler's explanation, there are no auxiliary hypotheses that are somehow used in deriving Königsberg's non-traversability from its bridge system's configuration (Lange, 2019, p. 3901). However, there are other explainers of Y, e.g., the oddness/ evenness of the bridge system.[43] (2) but not (1) allows for the independence notion to apply to Euler's explanation. Hence, I adopt (2).

[43] One can also explain Königsberg's non-traversability by referring to it having an odd number of bridges leaving each part of the town rather than its bridge system not being isomorphic to an Eulerian graph (Jansson & Saatsi, 2019, p. 836ff). The evenness/ oddness of the bridge system is another explainer of Y.

7.2 A counterexample to the independence notion

In the last section, I introduced Woodward's independence notion. I argued that one could use it in a conservative CEXP, claiming that X explains Y iff Y counterfactually depends on changes in X that are independent in the sense of this notion. This view correctly identifies the asymmetry in the explanation of the possibility of stable planetary orbits. In this section, I argue that the independence notion fails to yield asymmetry of counterfactual dependence in the flagpole explanation. As a result, a conservative CEXP relying on it fails to identify the flagpole explanation as asymmetric.

Here are the respective counterfactuals in the flagpole explanation, qualified with the independence notion:

> (CF_1^{ip}) If the flagpole's height had changed^{ip} from x_1' to x_1'' and/ or the angle of sunrays from x_2' to x_2'', the flagpole's shadow's length would have been y''.

> (CF_2^{ip}) If the flagpole's shadow's length had changed^{ip} from y' to y'' and/or the angle of sunrays from x_2' to x_2'', the flagpole's height would have been x_1''.

To account for the asymmetry of the flagpole explanation within a conservative CEXP, CF_1^{ip} should be true for some values of X_1 and X_2 and CF_2^{ip} false for all values of Y and X_2.[44] I claim that this is not the case.

First, consider CF_1^{ip}. Let us focus on independently changing the flagpole's height. Cutting the flagpole is a way of changing the flagpole's height. And any such change of the flagpole's height is plausibly independent of other explainers of its shadow's length such as the angle of sunrays.

[44] According to the definition of explanatory asymmetry in Section 3.1, the flagpole explanation would also be asymmetric if CF_1^{ip} would be true for some values of X_1 and X_2 and it would be false for all values of Y and X_1 that

> (CF_3^{ip}) If the flagpole's shadow's length had changed^{ip} from y' to y'' and/ or the flagpole's height from x_1' to x_1'', the angle of sunrays would have been x_2''.

However, CF_3^{ip} is true for some values of Y and X_1:
Suppose we focus on an independent change of the flagpole's shadow's length. This could be to observe the flagpole a few hours later when the sun is lower above the horizon. This change in the shadow's length is plausibly independent of other explainers of the angle of sunrays, e.g., the earth-sun gravitation. And under similar assumptions as in the main text, for at least some such changes in the shadow's length the angle of sunrays would be different. Thus, CF_3^{ip} is true for some values of Y and X_1.

Moreover, let us assume that we otherwise observe the flagpole as in the original case. There is no wall between the sun and the flagpole and so on. We also assume that the sun is still somewhere above the horizon. In such a situation, at least some ways of cutting the flagpole would change its shadow's length. Thus, CF_1^{ip} is true for some values of X_1 and X_2.

Second, consider CF_2^{ip}. Let us focus on independently changing the flagpole's shadow's length. Cutting the flagpole is a way to change the shadow's length. And any such change in the flagpole's shadow's length is plausibly independent of other explainers of the flagpole's height, e.g., the intentions of those building the flagpole. Let us also again assume that we still observe the flagpole as in the original case. There is no wall between the sun and the flagpole and so on. We still also assume that the sun is somewhere above the horizon. In such a situation, at least some ways of changing the flagpole's shadow's length by cutting the flagpole change its height. Thus, CF_2^{ip} is true for some values of Y and X_2. But this is the wrong result! The independence notion fails to yield asymmetry of counterfactual dependence in this case. As a result, a conservative CEXP with the independence notion fails to identify the flagpole explanation as asymmetric.[45]

Why does the independence notion not yield asymmetry of counterfactual dependence in the flagpole explanation? This is because the independence notion allows for a change in the explanandum Y to create a spurious counterfactual dependence of the explanans X on Y. Consider that this notion does not rule out that a change in Y is brought about by changing X. In CF_2^{ip}, changing the flagpole's shadow's length (Y) by changing the flagpole's height (X) is an independent change of the shadow's length according to the independence notion. However, changing Y by changing X creates a spurious counterfactual dependence of X on Y. In CF_2^{ip}, changing the flagpole's shadow's length by changing its height makes it look as if the flagpole's height changes as a result of the change in its shadow's length. But it does not. The flagpole's height merely changes because we change it to change its shadow's length. CF_2^{ip} is true for merely spurious reasons. The independence notion allows for a change in Y to create a spurious

[45] According to the definition of explanatory asymmetry in Section 3.1, a conservative CEXP with the independence notion fails to identify explanatory asymmetry in the flagpole explanation because CF_3^{ip} is also true for at least some values of Y and X_1 (see footnote 44).

counterfactual dependence of X on Y. For this reason, it fails to yield asymmetry of counterfactual dependence in the flagpole explanation.

This diagnosis of the problem suggests the following solution: To ensure asymmetry of counterfactual dependence, we should qualify counterfactuals such that the antecedent is changed in a way that changes the consequent only due to the change in the antecedent, if at all. In $\mathrm{CF}_2^{\mathrm{ip}}$, we should change the flagpole's shadow's length such that the flagpole's height only changes due to the change in its shadow's length, if at all. This can be done by imposing two further requirements on the change in the antecedent:

(i) The antecedent may not be changed in a way that changes the consequent directly.

In $\mathrm{CF}_2^{\mathrm{ip}}$, we may not change the flagpole's shadow's length by changing the flagpole's height.

(ii) The antecedent may not be changed in a way that changes a variable V which changes its consequent.

In $\mathrm{CF}_2^{\mathrm{ip}}$, we may not change the flagpole's shadow's length in a way that changes the intentions of those building the flagpole, and thereby also the flagpole's height.

In the literature, we already have a condition for a change of the antecedent that satisfies requirements (i) and (ii): The third condition of an intervention variable, I3, ensures that an intervention on X with respect to Y changes Y only via changing X if it does so at all (see Section 2.1). I3 implies that an intervention on X does not change Y directly, satisfying (i). And I3 implies that an intervention on X does not change V which changes Y, satisfying (ii).

I3 satisfies requirements (i) and (ii). Therefore, qualifying counterfactuals such that the change in their antecedent must satisfy I3 should yield asymmetry of counterfactual dependence in the flagpole explanation, avoiding the failure of the independence notion to do so. In the next section, I rely on this observation to propose an initial notion of a quasi-intervention. As I argue, this notion qualifies the antecedent of counterfactuals such as to yield asymmetry of counterfactual dependence in the flagpole explanation.

7.3 An initial notion of a quasi-intervention

In the last section, I argued that the independence notion fails to yield asymmetry of counterfactual dependence in the flagpole explanation. A conservative CEXP relying on it fails to identify this explanation as asymmetric. As mentioned, to solve this problem, we should require that the antecedent in a counterfactual is changed in a way that changes the consequent only due to the change in the antecedent, if at all. This can be done by demanding that the change in the antecedent must satisfy I3, in addition to the independence notion. Here is the resulting notion of a quasi-intervention, where QI3 is equivalent to I3 and QI4 to the independence notion (see also Figure 9):

(**QI**) QI is a quasi-intervention variable for X with respect to Y iff:

QI1. X counterfactually depends on QI.

QI2. Certain values of QI are such that when QI attains those values, X ceases to depend on the values of other variables that X previously depended on and instead depends only on the value taken by QI.[46]

QI3. QI changes Y if at all via changing X.

QI4. QI changes X such that any such change in X is independent of any changes in Z which explains Y not via explaining X.

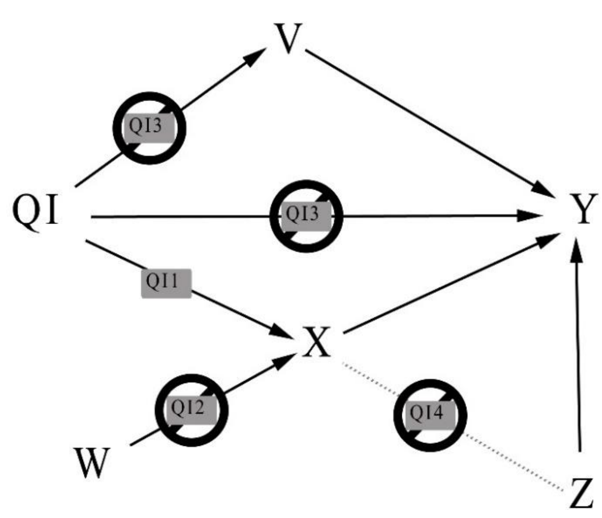


Figure 9: *Illustration of QI.*
W, V, X, Y, Z and QI denote variables. Solid arrows denote counterfactual dependence, dashed arrows unexplanatory correlations. The gray box denotes QI1 and the strikethroughs QI2, QI3 and QI4.

[46] I add QI2 in analogy to I2 in an intervention variable (see Section 2.1). It demands, for instance, that a quasi-intervention on the flagpole's shadow's length makes this length independent of any factor it previously depended on, e.g., the angle of sunrays. I2 is sometimes relaxed (see e.g., Woodward, 2015, p. 3584). I leave it to further research to defend, relax or give up on QI2.

Qualifying counterfactuals such that the change in the antecedent is brought about by a quasi-intervention yields asymmetry of counterfactual dependence in the flagpole explanation. To see this, consider the corresponding quasi-interventionist counterfactuals:

(CF_1^{qi}) If the flagpole's height had been changed from x_1' to x_1'' and/ or the angle of sunrays from x_2' to x_2'' by a quasi-intervention (with respect to the flagpole's shadow's length), the flagpole's shadow's length would have been y''.

(CF_2^{qi}) If the flagpole's shadow's length had been changed from y' to y'' and/or the angle of sunrays from x_2' to x_2'' by a quasi-intervention (with respect to the flagpole's height), the flagpole's height would have been x_1''.

Here, asymmetry of counterfactual dependence holds if CF_1^{qi} is true for some values of X_1 and X_2 and CF_2^{qi} false for all values of Y and X_2.[47] Under plausible assumptions, this is the case:

First, consider CF_1^{qi}. Let us focus on changing the flagpole's height. A quasi-intervention on the flagpole's height with respect to its shadow's length is a manipulation of the flagpole's height which changes its shadow's length through changing its height, if at all. Cutting the flagpole would be such a quasi-intervention. Under the same assumptions as before, for at least some cuts of the flagpole, its shadow's length would change. Hence, CF_1^{qi} is true for some values of X_1 and X_2.

Second, consider CF_2^{qi}. Let us first consider quasi-intervening on the flagpole's shadow's length. In contrast to the independence notion, QI3 precludes that cutting the flagpole is a quasi-intervention on the flagpole's shadow's length with respect to the flagpole's height. Rather, a quasi-intervention on the flagpole's shadow's length with respect to the flagpole's height is a manipulation of the flagpole's shadow's length which changes its height through changing its

[47] Alternatively, again, a conservative CEXP with the initial notion of a quasi-intervention would also identify the flagpole explanation as asymmetric if CF_1^{qi} would be true for some values of X_1 and X_2 and it would be false for all values of Y and X_1 that

(CF_3^{qi}) If the flagpole's shadow's length had been changed from y' to y'' and/ or the flagpole's height from x_1' to x_1'' by a quasi-intervention (with respect to the angle of sunrays), the angle of sunrays would have been x_2''.

This is also the case under similar assumptions as in the main text.

shadow's length, if at all. Building a wall between the sun and the flagpole would be such a quasi-intervention. Suppose we were to build the wall. Nothing would happen to the flagpole's height. The same seems to hold for any way of changing the flagpole's shadow's length that does not involve changing the flagpole's height, and thus for any quasi-intervention on the flagpole's shadow's length with respect to its height.

Let us now consider the angle of sunrays. A quasi-intervention on the angle of sunrays is a manipulation of this angle that changes the flagpole's height through changing the angle of sunrays, if at all. Observing the flagpole a few hours later is such a manipulation of the angle of sunrays. Assuming that meanwhile nobody cuts the flagpole or changes it in any other way, no such change in the angle of sunrays would change the flagpole's height. Moreover, the same seems to hold for any way of changing the angle of sunrays that does not involve changing the flagpole's height, and thus for any quasi-intervention on the angle of sunrays with respect to the flagpole's height. Hence, $\mathrm{CF}_2^{\mathrm{qi}}$ is false for all values of Y and X_2.

I conclude that the initial notion of a quasi-intervention yields asymmetry of counterfactual dependence in the flagpole explanation. Correspondingly, a conservative CEXP claiming that X explains Y iff Y counterfactually depends on changes in X brought about by a quasi-intervention yields explanatory asymmetry in the flagpole explanation. Importantly, such a conservative CEXP also accounts for the asymmetry of other explanations:

First, a conservative CEXP with the quasi-interventionist requirement identifies the asymmetry in the explanation of the possibility of stable planetary orbits. This is because QI4 amounts to the same requirement on a change in X as Woodward's independence notion imposes on it (see the definition of $\text{Change}^{\text{IP}}$ in Section 7.1). Under the same assumptions as before, it is true for at least some change in dimensionality that

($\mathrm{CF}_4^{\mathrm{qi}}$) If the dimensionality of space had been changed by a quasi-intervention, stable planetary orbits would have been impossible.[48]

[48] Changing n (dimensionality of space) to 5 is a way of changing space's dimensionality that changes the possibility of stable planetary orbits only via changing the dimensionality, thus a quasi-intervention on space's dimensionality.

And under the same assumptions, it is false that

($\mathrm{CF}_5^{\mathrm{qi}}$) If the possibility of stable planetary orbits had been changed by a quasi-intervention, the dimensionality of space would have been different.[49]

Hence, a conservative CEXP with the quasi-interventionist requirement identifies the asymmetry in this explanation.

Second, a conservative CEXP with the quasi-interventionist requirement identifies asymmetry in causal explanations beyond the flagpole explanation. This is because an intervention is a special case of a quasi-intervention. QI3 is equivalent to I3 in an intervention variable. QI4 is equivalent to I4[50], and QI2 to I2. Finally, I1 is a special case of QI1. I1 requires that an intervention causes X. By contrast, QI1 is a less demanding condition, requiring that X counterfactually depends on the quasi-intervention. It allows for quasi-interventions to also apply to non-causal explanations. For instance, it allows that the dimensionality of space counterfactually depends on changing it in a mathematical calculation, despite such a change not causing space's dimensionality. However, if QI indeed causes X then X counterfactually depends on QI (see also Khalifa et al., 2020, p. 1444). In other words, I1 is a special case of QI1. Hence, an intervention is a special case of a quasi-intervention. As a result, a conservative CEXP with the quasi-interventionist requirement identifies asymmetry whenever the requirement that Y counterfactually depends on changes in X brought about by interventions yields asymmetry. We may take this to mean that a conservative CEXP with the quasi-interventionist requirement accounts for asymmetry in causal explanations.

Let me take stock: In Chapter 6, I argued that if X and Y relate mathematically and with a many-to-one relation between their actual values then an unqualified CEXP already entails that (if at all) X explains Y but not vice versa. This was due to CExp_2 which is asymmetrically satisfied in such explanations. In this chapter, I have thus far argued that qualifying the antecedent in counterfactuals with quasi-interventions yields asymmetry of counterfactual dependence in causal and some non-causal explanations. A conservative CEXP with the quasi-interventionist requirement accounts for

[49] $\mathrm{CF}_5^{\mathrm{qi}}$ has an impossible antecedent if we take the form of gravitational potential and laws of motion to be the other explainer of space's dimensionality (see also Section 7.1).

[50] One might worry about a difference between I4 and QI4: I4 requires that the intervention I be independent of Z (see Section 2.1). By contrast, QI4 requires that a quasi-intervention be such that the change in X it brings about is independent of changes in Z. However, this difference is irrelevant: If I causes X and I is correlated with Z then changes in X will not be independent of changes in Z. I4 excludes this. And this is exactly what QI4 also excludes. Thus, their apparent difference is irrelevant.

explanatory asymmetry in these cases. Based on both insights, one might think that we have a solution to the challenge of asymmetry. The thus far developed conservative CEXP qualifies but upholds $CExp_2$. Therefore, one might think that it does not only yield asymmetry in causal and some non-causal explanations but also in mathematical explanations with many-to-one relations. However, it does not. Instead, this conservative CEXP wrongly claims that X does not explain Y if X and Y relate mathematically, as I argue in the next section.

7.4 A problem with mathematical explanations

A conservative CEXP with the quasi-interventionist requirement wrongly claims that X does not explain Y if X and Y relate mathematically. Here is why: According to this conservative CEXP, X explains Y iff Y counterfactually depends on changes in X brought about by a quasi-intervention on X with respect to Y (P1). But if X and Y relate mathematically, then one cannot quasi-intervene on X with respect to Y (P2).[51] Thus, a conservative CEXP with the quasi-interventionist requirement wrongly claims that X does not explain Y if X and Y relate mathematically.

To illustrate, consider Euler's explanation. According to the proposed conservative CEXP, the bridge system's configuration (X) explains its non-traversability (Y) iff the non-traversability counterfactually depends on changes in the configuration brought about by a quasi-intervention on this configuration with respect to its non-traversability (P1). But X and Y relate mathematically here (see Chapter 6). Therefore, one cannot quasi-intervene on the configuration with respect to non-traversability (P2). A conservative CEXP with the quasi-interventionist requirement wrongly claims that the bridge system's configuration does not explain its non-traversability. Let me explain P2 further.

In Euler's explanation, we cannot quasi-intervene on the bridge system's configuration with respect to its non-traversability. This is due to condition QI3 for a quasi-intervention. QI3 demands that we change the bridge system's configuration without changing its non-traversability directly. However, any change to the bridge system's configuration such that it becomes isomorphic to an Eulerian graph is already (by mathematical necessity) a direct change of its non-traversability. We cannot change the bridge system's non-traversability only *via* changing its configuration rather

[51] I elaborate on P2 below.

than directly. QI3 cannot be satisfied here. Thus, we cannot quasi-intervene on the bridge system's configuration with respect to its non-traversability.

More generally, if X and Y relate mathematically, QI3 cannot be satisfied. QI3 requires that a quasi-intervention on X changes Y via changing X, if at all. However, if X and Y relate mathematically, then X's taking on some value necessitates that Y takes on some value and vice versa (see Chapter 6). This implies that we cannot change X such that Y only changes *via* changing X. Rather, any change of X such that Y changes is already (by mathematical necessity) a direct change of Y. Hence, one cannot change X such as to satisfy QI3. If X and Y relate mathematically, one cannot quasi-intervene on X with respect to Y.

I conclude that a conservative CEXP with the quasi-interventionist requirement wrongly claims that X does not explain Y if X and Y relate mathematically. This is a problem: Cases of mathematical explanations like Euler's explanation are amongst those proponents of CEXP aim to cover. Hence, we need to amend the proposed conservative CEXP such that it covers mathematical explanations. In the next section, I do so by providing a second and final notion of a quasi-intervention and a corresponding quasi-interventionist CEXP.

7.5 A quasi-interventionist theory of explanation

In the last section, I argued that if X and Y relate mathematically, we cannot quasi-intervene on X with respect to Y. To overcome this shortcoming, I suggest adding exemption clauses for mathematical explanations to the notion of a quasi-intervention. This means, if X and Y relate mathematically, a quasi-intervention on X with respect to Y is simply any way of changing X. In particular, a quasi-intervention on X need not anymore change Y only via changing X but may change Y directly. Thus, we circumvent the problem identified in the last section.[52]

[52] I suggest adding exemption clauses to QI2 and QI4 as well. Sometimes, if X and Y relate mathematically one cannot satisfy QI4 and QI2. This happens if Z explains both X and Y and X and Z are also mathematically related. Suppose we take X to denote the configuration of Königsberg's bridge system being isomorphic or not, Z whether this bridge system is even or odd and Y Königsberg's non-traversability. Here, Z explains Y: If the bridge system is even, then it is traversable. If it is odd, it is only traversable in some cases. Moreover, Z explains X: According to Euler's theorem, a bridge system is isomorphic to an Eulerian graph iff it either has only evenly connected or at most two oddly connected parts. Furthermore, in addition to X and Y, Z and X are mathematically related here, via Euler's theorem. This implies that we cannot make X independent of Z, failing to satisfy QI2. And since Z and X relate mathematically, changes in X are not independent of changes in Z, failing to satisfy QI4. However, even though no change to Königsberg's non-isomorphic configuration satisfies QI2 and QI4, we want to uphold that it explains its non-traversability. Adding exemption clauses to QI2 and QI4 allows us to do so.

Here is this final notion of a quasi-intervention (see also Figure 10):

(**QI_f**) QI_f is a quasi-intervention variable for X with respect to Y iff:

QI_f1. X counterfactually depends on QI_f.

QI_f2. Certain values of QI_f are such that when QI_f attains those values, X ceases to depend on the values of other variables that X previously depended on and instead depends only on the value taken by QI_f unless X and Y are mathematically related.

QI_f3. QI_f changes Y if at all via changing X unless X and Y are mathematically related.

QI_f4. QI_f changes X such that any such change in X is independent of any changes in Z which explains Y not via explaining X unless X and Y are mathematically related.

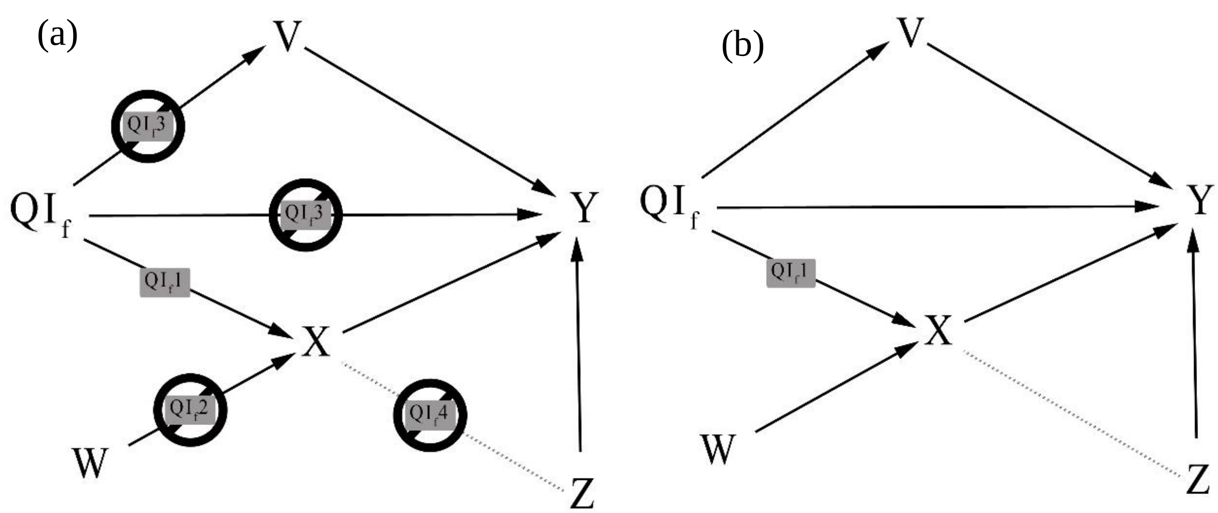


Figure 10: *Illustration of QI_f.*
W, V, X, Y, Z and QI_f denote variables. Solid arrows denote explanatory counterfactual dependence, dashed arrows unexplanatory correlations.
The gray box denotes QI_f1 and the strikethroughs QI_f2, QI_f3 and QI_f4.
(a) X and Y do not relate mathematically.
(b) X and Y relate mathematically.

At this point, it is also convenient to make a hitherto implicit connection explicit. Recall, within the interventionist theory of causal explanation, an intervention is defined in reference to the notion of an intervention variable (see Section 2.1). In analogy, QI_f defines a quasi-intervention variable. Here is the corresponding definition of a quasi-intervention:

Quasi-intervention

QI_f's assuming some value qi_f is a quasi-intervention on X with respect to Y iff

(i) QI_f is a quasi-intervention variable for X with respect to Y.

(ii) X takes on x due to QI_f assuming qi_f.

Due to the added exemption clauses, the final notion of a quasi-intervention can be applied to mathematical explanations. Thereby, it allows us to demand that $CExp_2$ and $CExp_3$ refer to quasi-interventions without excluding mathematical explanations. Here is the resulting quasi-interventionist theory of explanation:

(QuasiCEXP) Suppose that M is an explanandum consisting in the statement that some variable Y takes on y. Then an explanans E for M will consist of (a) (a generalization G describing) a relationship between Y and X (where X may consist of several variables $X_1 \ldots X_n$), and (b) a statement that X takes on x. E is explanatory with respect to M iff

($QCExp_1$) E and M are (approximately) true.

($QCExp_2$) (G describing) the relationship between X and Y correctly describes the actual value y of Y if X takes on its actual value x due to QI_f assuming qi_f.

($QCExp_3$) (G describing) the relationship between X and Y correctly describes that Y would change from y' to y'' if X would change from x' to x'' due to QI_f assuming qi_f', for at least some values of X.

QuasiCEXP yields explanatory asymmetry in two ways:[53]

(a) If X and Y do not relate mathematically, the requirement that Y counterfactually depends on changes in X brought about by a quasi-intervention yields explanatory asymmetry.

[53] Despite yielding explanatory asymmetry in two ways QuasiCEXP is a strongly monist version of CEXP: It qualifies CEXP in the same way for causal and non-causal explanations.

Due to (a), QuasiCEXP accounts for explanatory asymmetry in causal explanations, including the flagpole explanation, and some non-causal ones, including the explanation of the possibility of stable planetary orbits and plausibly the one of liquid crystals' fluid behaviour.[54] Note that due to the exception clauses for mathematical explanations, the quasi-interventionist requirement in QuasiCEXP does not ensure explanatory asymmetry in mathematical explanations.

> (b) If X and Y relate mathematically and their actual values with a many-to-one relation, then QuasiCEXP entails that (if at all) X explains Y but not vice versa.

(b) holds because QuasiCEXP upholds $CExp_2$ for mathematical explanations. If X and Y relate mathematically, there is no difference between $QCExp_2$ and $CExp_2$. This is because there is no difference between a change in X brought about by a quasi-intervention and simply any change in X if X and Y relate mathematically. Due to (b), QuasiCEXP accounts for asymmetry in Euler's explanation (see Chapter 6) and the strawberry explanation.[55]

Notably, there is a class of non-causal explanations QuasiCEXP claims to be symmetric: mathematical explanations with one-to-one relations between the actual values of the explanans and the explanandum. In the next chapter, I illustrate why and propose to accept the symmetry of such explanations. Before doing so, I want to address one objection:

One could accuse me of failing to provide a semantics for the kinds of counterfactuals QuasiCEXP refers to, namely quasi-interventionist counterfactuals. In response, my project has been to qualify which counterfactuals CEXP should refer to such that a thus qualified CEXP accounts for explanatory asymmetry. Even without a semantics of quasi-interventionist counterfactuals, we can judge whether such counterfactuals are true or false, considering various dependence claims we

[54] Take X to represent the rod-like shape of the molecules with x^1 representing them having this shape and x^2 them not having this shape. Y represents the fluid behavior with y^1 representing anisotropic fluid behavior and y^2 not anisotropic fluid behaviour. X and Y do not relate mathematically here. Due to (a), QuasiCEXP yields asymmetry here: First, a quasi-intervention on the shape of the molecules with respect to the fluid behaviour would be to chemically change their shape to be not rod-like. Such a change plausibly does not change the fluid behaviour other than through the change in the molecules. And plausibly the fluid behaviour would change (i.e. not continue to be anisotropic) under such a quasi-intervention. Second, a quasi-intervention on the fluid behaviour with respect to the shape of the molecules would be to cool the fluid turning it into a solid. Such a change plausibly does not change the shape of the molecules directly. And such a change of the fluid behaviour does not change the shape of the molecules at all. QuasiCEXP claims that the shape of the molecules explains the fluid behaviour but not vice versa.

[55] Take X to represent the numbers of strawberries with x^i representing a particular number. Take Y to have two values, y^1 representing indivisibility and y^2 divisibility. Then, the actual value of X relates only to the actual value of Y, but the actual value of Y relates to several values of X. 23 strawberries are not evenly divisible, but several numbers of strawberries are not evenly divisible.

assume to hold. Thus, we can judge whether QuasiCEXP accounts for explanatory asymmetry. The further project of providing a semantics for quasi-interventionist counterfactuals should be addressed in the future.[56]

[56] Reutlinger suggests how to amend Lewisian semantics, Goodmanian semantics and suppositional theory for interventionist counterfactuals (Reutlinger, 2013, Chapter 3). It could be possible to develop a semantics for quasi-interventionist counterfactuals in a similar fashion. Note that it would also have to be applicable if X and Y do not represent events, unlike standard Lewisian semantics (see Chapter 4).

8 Towards a quasi-interventionist solution to the revised challenge

In this chapter, I propose that QuasiCEXP is a promising solution to the revised challenge of (a)symmetry rather than the challenge of asymmetry. To this end, I first explain why mathematical explanations with a one-to-one relation between the actual values of the explanans and the explanandum are symmetric according to QuasiCEXP (Section 8.1). Then, I provide initial motivation for accepting these explanations as symmetric and outline open questions on QuasiCEXP as solving the revised challenge (Section 8.2). Finally, I argue that the approaches discussed in Part II do not solve the revised challenge (Section 8.3).

8.1 Explanatory symmetry in the quasi-interventionist theory

According to QuasiCEXP, X explains Y and Y explains X if X and Y relate mathematically and with a one-to-one relation between their actual values. To see this, let us discuss the only explanation we have neglected thus far: the explanation of the scaling exponent $\frac{3}{4}$ in Kleiber's law, which relates the basal metabolic rate of organisms to their body mass (see Section 2.2.4). Recall, biologists explain this exponent using the dimensionality of the organisms covered by Kleiber's law. The explanation of this law is a mathematical explanation with a one-to-one relation:

First, the actual values of X and Y relate one-to-one. X represents the dimensionality of organisms with x^i denoting different dimensionalities and x^1 three-dimensionality (the actual value of X). Y represents the scaling exponent with y^i denoting different scaling exponents and y^1 the scaling exponent $\frac{3}{4}$ (the actual value of Y). Moreover, d-dimensional organisms have a scaling exponent of $\frac{d}{d+1}$. This implies that three-dimensionality is associated only with a $\frac{3}{4}$ scaling exponent and vice versa. The actual values of X and Y relate one-to-one.

Second, X and Y relate mathematically. Here, the scaling exponent is derived from a geometrical feature (the fractal-likeness) of the organisms' life-distributing networks, and thereby from their dimensionality (see Section 2.2.4). According to this derivation, under certain assumptions, if organisms are d-dimensional and employ corresponding fractal-like networks, their scaling exponent must be $\frac{d}{d+1}$ for geometrical reasons. Similarly, under the same assumptions, if the scaling exponent of some organisms is $\frac{d}{d+1}$, they must be d-dimensional for geometrical reasons. X and Y relate mathematically here.

Now, recall, QuasiCEXP accounts for explanatory asymmetry in two ways:

> (a) If X and Y do not relate mathematically, the requirement that Y counterfactually depends on changes in X brought about by a quasi-intervention yields explanatory asymmetry.
>
> (b) If X and Y relate mathematically and their actual values with a many-to-one relation, then QuasiCEXP entails that (if at all) X explains Y but not vice versa.

Neither (a) nor (b) holds in the case of mathematical explanations with one-to-one relations like the explanation of Kleiber's law. Instead, according to QuasiCEXP, these explanations are symmetric:

First, consider $QCExp_2$, the condition that the relationship between X and Y correctly describes the actual value y of Y if X takes on its actual value x due to QI_f assuming qi_f. If X and Y relate mathematically, then $QCExp_2$ amounts to the condition that Y takes on its actual value y if X takes on its actual value x. This is satisfied in both directions in mathematical explanations with one-to-one relations like the explanation of Kleiber's law: It is true that if organisms are three-dimensional, their scaling exponent is $\frac{3}{4}$. And it is true that if their scaling exponent is $\frac{3}{4}$, the organisms are three-dimensional.

Second, consider $QCExp_3$. This condition requires that the relationship between X and Y correctly describes that Y would change from y' to y'' if X would change from x' to x'' due to QI_f assuming qi_f' for at least some values of X. Again, if X and Y relate mathematically, then $QCExp_3$ amounts to the condition that Y counterfactually depends on changes in X for some values of X. Again, this is satisfied in both directions in mathematical explanations with one-to-one relations: In such explanations, there is some change in X (namely a change to the actual value of X) that necessitates a corresponding change in Y. And there is some change in Y (a change to its actual value) that necessitates a corresponding change in X. For instance, changing the organisms' dimensionality from two to three dimensions necessarily changes the corresponding scaling exponent from $\frac{2}{3}$ to $\frac{3}{4}$ and vice versa. Thus, $QCExp_3$ is satisfied in both directions.

To conclude, $QCExp_2$ and $QCExp_3$ are satisfied in both directions in mathematical explanations with one-to-one relations. According to QuasiCEXP, X explains Y and Y explains X in such cases.

The dimensionality of organisms explains the scaling exponent and vice versa. In the next section, I propose to accept this result.

8.2 Towards a quasi-interventionist solution to the revised challenge

Because QuasiCEXP entails explanatory symmetry in some non-causal explanations it does not solve the challenge of asymmetry. However, QuasiCEXP is a promising solution to the revised challenge of (a)symmetry, which demands to account for the symmetry of some non-causal explanations and the asymmetry of causal and other non-causal ones. Accepting QuasiCEXP as solving the revised challenge requires accepting that the mathematical explanations with one-to-one relations it claims to be symmetric are actually symmetric. Here are two initial motivations to accept them as symmetric:

First, intuitions about explanatory asymmetry in explanations like the one of Kleiber's law seem to be fragile. Indeed, it seems to be difficult to tell whether it is intuitive that the dimensionality of organisms explains the particular proportional relation of their body mass to their basal metabolic rate. Similarly, it seems to be difficult to tell whether it is intuitive that the particular proportional relation of the organisms' body mass to their basal metabolic rate explains their dimensionality. I admit that statements about intuitions are themselves fragile. Yet, the fragility of intuitions concerning asymmetry in mathematical explanations with one-to-one relations provides some motivation to take seriously that these explanations might be symmetric.

Second, the symmetry of mathematical explanations with one-to-one relations aligns well with claims about symmetric non-causal explanations in the literature. In Section 3.2, I mentioned that some authors accept the symmetry of non-causal explanations. They do so for mathematical explanations with one-to-one relations: Reutlinger sees Euler's explanation as symmetric, understanding it such that the actual values of X and Y relate one-to-one (Reutlinger, 2018, p. 92).[57] And Saatsi and Pexton (2013) claim the explanation of Kleiber's law to be symmetric. QuasiCEXP would agree with their views.

[57] As Reutlinger understands the explanans X, it represents whether a bridge system is isomorphic to an Eulerian graph (x^2) or not (x^1) (Reutlinger, 2018, p. 92). Y represents that the bridge system is non-traversable (y^1) or traversable (y^2). In this version of Euler's explanation, the actual values of X and Y relate one-to-one: x^1 is associated only with y^1 and vice versa. A non-isomorphic bridge system is non-traversable, and a non-traversable one non-isomorphic.

These initial motivations to accept the symmetry of mathematical explanations with one-to-one relations do not suffice to defend such an acceptance. A proper defence would require establishing that scientists and laypeople view such explanations as symmetric. I leave this project for further research. Here, I only pose two open questions this further defence should address:

1) Would QuasiCEXP identify more explanations as symmetric when considering other explanations failing to satisfy QI3, beyond mathematical ones?

Beyond the literature this thesis deals with, there are also other explanations failing to satisfy QI3 in the initial notion of a quasi-intervention, e.g., so-called constitutive explanations (see e.g., Kästner & Andersen, 2018; Kostić, 2018). One could aim to incorporate them into QuasiCEXP by adding further exemption clauses to the notion of a quasi-intervention such as to allow for such explanations to be covered. If so, then whenever the actual values of the explanans and the explanandum in these further explanations relate one-to-one, these explanations might be symmetric according to an amended QuasiCEXP. It is a topic for further research how to incorporate these additional explanations into QuasiCEXP and whether such a version of QuasiCEXP entails their symmetry.

2) Is it plausible that the directionality of mathematical explanations changes when we understand their explanans differently?

Whether a mathematical explanation is asymmetric or symmetric according to QuasiCEXP depends only on whether the actual values of the explanans and the explanandum relate many-to-one or one-to-one. However, we often can understand an explanation either way. For instance, we can understand Euler's explanation such that the bridge system's configuration (X) relates many-to-one to its non-traversability (Y), as in Chapter 6. But we can also understand X as a binary variable with x^1 denoting the bridge system's configuration being isomorphic to an Eulerian graph and x^2 denoting that it is not. Here, the actual value of X relates one-to-one to the actual value of Y. In the former case, QuasiCEXP entails the explanation to be asymmetric, and in the latter symmetric. Whether this feature of QuasiCEXP is plausible should be discussed in the future.

8.3 Again against Lewisian semantics and the inference schemes

In the last section, I proposed that QuasiCEXP is a promising solution to the revised challenge. One might wonder though: Are amendments of CEXP using Lewisian semantics or the inference schemes also promising solutions to this challenge? They are not.

First, consider a conservative CEXP with Lewisian semantics. In Chapter 4, I argued that it is unclear how a conservative CEXP with Lewisian semantics would solve the challenge of asymmetry. This was because it is unclear how Lewisian semantics would yield asymmetry of counterfactual dependence in non-causal explanations involving explanantia and/ or explananda that are not events. The reason was that Lewisian semantics yields no clear verdict on the truth value of counterfactuals not involving events. For the same reason, it is unclear how a conservative CEXP with Lewisian semantics would solve the revised challenge. Lewisian semantics does not identify counterfactuals not involving events as true. But a conservative CEXP with Lewisian semantics demands that such counterfactuals be true to be explanatory. As a result, it is unclear how such a conservative CEXP would establish either explanatory symmetry or asymmetry in non-causal explanations without events. Therefore, it is unclear how a conservative CEXP with Lewisian semantics would solve the revised challenge.

Second, consider the modal fact principle. Neither a conservative nor a radical CEXP with the modal fact principle solves the revised challenge. This is because this principle never establishes explanatory symmetry. In its more promising version, the modal fact principle says that X explains Y if X can be the target of an intervention and Y cannot be. Suppose X can be the target of an intervention (e.g., the bridge system's configuration) and Y cannot (e.g., its non-traversability). The principle tells us that X explains Y. But it will never tell us that Y also explains X unless Y can be the target of an intervention and X cannot be. And if that were the case then the principle would no longer claim that X explains Y. It would never yield explanatory symmetry. Thus, neither a conservative nor a radical CEXP with the modal fact principle solves the revised challenge.

Third, consider the double explanation principle. Neither a conservative nor a radical CEXP with the double explanation principle solves the revised challenge. This principle says that X explains Y if (i) Z causally explains X, (ii) X and Y are statistically dependent, (iii) Y and Z are statistically dependent and (iv) there are no other explanations of these dependencies. However, as argued in Section 5.2, there are non-causal explanations in which either of (i), (iii) or (iv) is not satisfied and still X explains Y, as in Euler's explanation. In such cases, a conservative or radical CEXP with the double explanation principle fails to claim that X explains Y. Hence, neither a conservative nor a radical CEXP with the double explanation principle solves the revised challenge.

I conclude that amendments of CEXP using Lewisian semantics or the inference schemes do not solve the revised challenge. QuasiCEXP, by contrast, is a promising solution to this challenge. Whether one can adequately defend the symmetry of the explanations that are symmetric according to QuasiCEXP remains to be seen in the future.

9 Conclusion

We typically explain phenomena by citing their causes. However, some explanations do not fit this scheme of a causal explanation. Or so several philosophers have argued. For those philosophers, some such non-causal explanations instead make essential use of mathematics like Euler's theorem. In others, they argue, it is simply hard to understand the explanans as causing the explanandum. This thesis started from accepting the need to defend a theory of explanation that accounts for the explanatory power of such non-causal explanations.

In Part I, I defined a counterfactual theory of causal and non-causal explanation motivated by a common view of several authors: CEXP extends the interventionist theory of causal explanation to capture the explanatory power of causal and non-causal explanations. Just like the interventionist theory, CEXP claims that X explains Y iff Y counterfactually depends on changes in X. However, unlike the interventionist theory, CEXP does not require that Y counterfactually depends on changes in X brought about by interventions. I also introduced the challenge of asymmetry for CEXP: how to qualify CEXP such that it claims all explanations it covers to be asymmetric. I distinguished this challenge from an alternative one. The revised challenge of (a)symmetry demands to also account for some symmetric non-causal explanations. The distinction of both challenges allows us to specify when a theory fails to account for explanatory asymmetry and when it successfully identifies a symmetric explanation.

In Part II, I rejected two potential avenues to amend CEXP such as to solve the challenge of asymmetry: I argued that it is unclear how a conservative CEXP with Lewisian semantics would solve the challenge of asymmetry. This is because it is unclear how Lewisian semantics would establish asymmetry of counterfactual dependence for counterfactuals that do not involve events. And in many non-causal explanations, the explanantia and/ or explananda are not events. Moreover, neither a radical nor a conservative CEXP with two inference schemes (Woodward's (2018, 2020) double explanation principle and Lange's (2019) modal fact principle) solves the challenge of asymmetry. Overall, Part II established the need for an alternative avenue to deal with explanatory asymmetry as a proponent of CEXP.

In Part III, I developed a quasi-interventionist version of CEXP, promising as a solution to the revised challenge: QuasiCEXP claims that X explains Y iff Y counterfactually depends on changes in X brought about by a quasi-intervention. QuasiCEXP accounts for explanatory asymmetry in two ways: First, if X and Y relate mathematically and with a many-to-one relation between their actual values, then an unqualified CEXP already entails that (if at all) X explains Y but not vice versa. This feature also holds for QuasiCEXP. Second, if X and Y do not relate mathematically, the requirement in QuasiCEXP that Y counterfactually depends on changes in X brought about by a quasi-intervention yields explanatory asymmetry. Furthermore, QuasiCEXP entails explanatory symmetry if X and Y relate mathematically and with a one-to-one relation between their actual values. I provided initial motivation to accept the symmetry of such explanations and outlined topics for further research on QuasiCEXP as solving the revised challenge. Overall, QuasiCEXP is a promising solution to the revised challenge, in contrast to amendments of CEXP using Lewisian semantics or the inferences schemes.

To summarize, the main contributions of this thesis to the literature on explanation are:

1. I distinguish between the challenge of asymmetry and the revised challenge of (a)symmetry that demands to identify some non-causal explanations as symmetric.
2. I reject versions of a counterfactual theory of explanation relying on Lewisian semantics or two inference schemes as solutions to the challenge of asymmetry and the revised challenge of (a)symmetry.
3. I provide a quasi-interventionist theory of explanation as a promising solution to the revised challenge of (a)symmetry.

There are several directions for future research emerging from this thesis: Firstly, as discussed in Chapter 8, one should thoroughly defend the symmetry of the non-causal explanations QuasiCEXP identifies as symmetric and address the open questions on this issue. Secondly, one should develop how mathematical explanations with several explanantia can relate many-to-one such as to exhibit actual value asymmetry (and thus explanatory asymmetry according to QuasiCEXP). Thirdly, one should test how successfully QuasiCEXP identifies the (a)symmetry in a wide range of causal and non-causal explanations beyond those considered in this thesis. This includes explanations discussed in the literature (see e.g., examples in Craver & Povich, 2017; Reutlinger, 2018) and undiscussed ones. Testing QuasiCEXP would also require case studies of how scientists and laypeople explain. Fourthly, as mentioned in Section 7.5, one should develop a semantics for quasi-interventionist counterfactuals. Finally, one should address two discussions that were left aside in this thesis: (i) A counterfactual theory of causal and non-causal explanation like QuasiCEXP faces problems beyond accounting for asymmetry (see e.g., Khalifa et al., 2020; Kuorikoski, 2021). (ii) A unified theory of explanation like QuasiCEXP is not the only tenable response to examples of non-causal explanations (for a pluralist view see Lange, 2017; for a cluster view see Rice & Rohwer, 2020). I leave all of these directions to further research.

Bibliography

Baron, S., Colyvan, M., & Ripley, D. (2017). How Mathematics Can Make a Difference. *Philosophers' Imprint*, *17*(3). http://hdl.handle.net/2027/spo.3521354.0017.003

Berto, F., & Jago, M. (2018). Impossible Worlds. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2018). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/fall2018/entries/impossible-worlds/

Bokulich, A. (2011). How scientific models can explain. *Synthese*, *180*(1), 33–45. https://doi.org/10.1007/s11229-009-9565-1

Bromberger, S. (1966). Why-questions. In R. Colodny (Ed.), *Mind and Cosmos: Essays in contemporary science and philosophy* (pp. 86–111). Pittsburgh University Press.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.

Craver, C. F., & Povich, M. (2017). The directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science Part A, 63,* 31–38. https://doi.org/10.1016/j.shpsa.2017.04.005

Diestel, R. (2017). *Graph theory* (fifth). Springer.

Elliott, K., & Lange, M. (2021). Running it up the flagpole to see if anyone salutes: A response to Woodward on causal and explanatory asymmetries. *THEORIA. An International Journal for Theory, History and Foundations of Science*, *0*(0), Article 0. https://doi.org/10.1387/theoria.22351

French, S., & Saatsi, J. (2018). Symmetries and explanatory dependencies in physics. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 185–205). Oxford University Press.

Held, C. (2019). Towards a Monist Theory of Explanation. *Journal for General Philosophy of Science*, *50*(4), 447–475. https://doi.org/10.1007/s10838-019-09458-6

Hitchcock, C., & Woodward, J. (2003a). Explanatory Generalizations, Part I: A Counterfactual Account*. *Noûs, 37*(1), 1–24. https://doi.org/10.1111/1468-0068.00426

Hitchcock, C., & Woodward, J. (2003b). Explanatory Generalizations, Part II: Plumbing Explanatory Depth. *Noûs*, *37*(2), 181–199. JSTOR. https://doi-org.eur.idm.oclc.org/10.1111/1468-0068.00435

Jansson, L. (2015). Explanatory Asymmetries: Laws of Nature Rehabilitated. *The Journal of Philosophy*, *112*(11), 577–599. https://doi.org/10.5840/jphil20151121138

Jansson, L. (2020). Network explanations and explanatory directionality. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *375*(1796), 20190318. https://doi.org/10.1098/rstb.2019.0318

Jansson, L., & Saatsi, J. (2019). Explanatory Abstractions. *The British Journal for the Philosophy of Science*, *70*(3), 817–844. https://doi.org/10.1093/bjps/axx016

Kästner, L., & Andersen, L. M. (2018). Intervening into mechanisms: Prospects and challenges. *Philosophy Compass*, *13*(11), e12546. https://doi.org/10.1111/phc3.12546

Khalifa, K., Doble, G., & Millson, J. (2020). Counterfactuals and Explanatory Pluralism. *The British Journal for the Philosophy of Science*, *71*(4), 1439–1460. https://doi.org/10.1093/bjps/axy048

Khalifa, K., Millson, J., & Risjord, M. (2021). Inference, explanation, and asymmetry. *Synthese*, *198*(4), 929–953. https://doi.org/10.1007/s11229-018-1791-y

Kitcher, P., & Salmon, W. (1987). Van Fraassen on Explanation. *The Journal of Philosophy*, *84*(6), 315–330. https://doi.org/10.2307/2026782

Kostić, D. (2018). Mechanistic and topological explanations: An introduction. *Synthese*, *195*(1), 1–10. https://doi.org/10.1007/s11229-016-1257-z

Kou, J., Chen, Y., Zhou, X., Lu, H., Wu, F., & Fan, J. (2014). Optimal structure of tree-like branching networks for fluid flow. *Physica A: Statistical Mechanics and Its Applications*, *393*, 527–534. https://doi.org/10.1016/j.physa.2013.08.029

Kuorikoski, J. (2021). There Are No Mathematical Explanations. *Philosophy of Science*, *88*(2), 189–212. https://doi.org/10.1086/711479

Lange, M. (2013). What Makes a Scientific Explanation Distinctively Mathematical? *The British Journal for the Philosophy of Science*, *64*(3), 485–511. https://doi.org/10.1093/bjps/axs012

Lange, M. (2017). *Because without cause: Non-causal Explanations in Science and Mathematics*. Oxford University Press.

Lange, M. (2019). Asymmetry as a challenge to counterfactual accounts of non-causal explanation. *Synthese*, *198*(4), 3893–3918. https://doi.org/10.1007/s11229-019-02317-3

Lewis, D. (1979). Counterfactual Dependence and Time's Arrow. *Noûs*, *13*(4), 455–476. https://doi.org/10.2307/2215339

Lewis, D. (1986). Counterfactuals and Comparative Possibility. In *Philosophical Papers Volume II* (pp. 3–32). Oxford University Press. https://doi.org/10.1093/0195036468.003.0001

Menzel, C. (2017). Possible Worlds. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/possible-worlds/

Menzies, P., & Beebee, H. (2020). Counterfactual Theories of Causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2020). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2020/entries/causation-counterfactual/

Pincock, C. (2018). Accommodating explanatory pluralism. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 39–56). Oxford University Press.

Povich, M. (2018). Minimal Models and the Generalized Ontic Conception of Scientific Explanation. *The British Journal for the Philosophy of Science*, *69*(1), 117–137. https://doi.org/10.1093/bjps/axw019

Povich, M. (2021). The Narrow Ontic Counterfactual Account of Distinctively Mathematical Explanation. *The British Journal for the Philosophy of Science*, *72*(2), 511–543. https://doi.org/10.1093/bjps/axz008

Price, H. (1992). Agency and Causal Asymmetry. *Mind*, *101*(403), 501–520. https://www.jstor.org/stable/2253900

Reutlinger, A. (2013). *A Theory of Causation in the Social and Biological Sciences*. Palgrave Macmillan.

Reutlinger, A. (2017a). Explanation beyond causation? New directions in the philosophy of scientific explanation. *Philosophy Compass*, *12*(2), e12395. https://doi.org/10.1111/phc3.12395

Reutlinger, A. (2017b). Does the counterfactual theory of explanation apply to non-causal explanations in metaphysics? *European Journal for Philosophy of Science*, *7*(2), 239–256. https://doi.org/10.1007/s13194-016-0155-z

Reutlinger, A. (2018). Extending the counterfactual theory of explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 74–95). Oxford University Press.

Reutlinger, A., & Saatsi, J. (Eds.). (2018). *Explanation beyond causation: Philosophical perspectives on non-causal explanations*. Oxford University Press.

Rice, C. (2015). Moving Beyond Causes: Optimality Models and Scientific Explanation. *Noûs*, *49*(3), 589–615. https://doi.org/10.1111/nous.12042

Rice, C., & Rohwer, Y. (2020). How to Reconcile a Unified Account of Explanation with Explanatory Diversity. *Foundations of Science*. https://doi.org/10.1007/s10699-019-09647-y

Saatsi, J., & Pexton, M. (2013). Reassessing Woodward's Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations. *Philosophy of Science*, *80*(5), 613–624. https://doi.org/10.1086/673899

Salmon, W. C. (1984). Scientific Explanation: Three Basic Conceptions. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, *1984*, 293–305. https://www.jstor.org/stable/192510

Salmon, W. C. (1989). *Four Decades of Scientific Explanation*. University of Pittsburgh Press.

Schindler, S. (2014). Explanatory fictions—For real? *Synthese*, *191*(8), 1741–1755. https://doi.org/10.1007/s11229-013-0362-5

Singh, S. (2002). *Liquid crystals: Fundamentals*. World Scientific.

Skow, B. (2014). Are There Non-Causal Explanations (of Particular Events)? *The British Journal for the Philosophy of Science*, *65*(3), 445–467. https://doi.org/10.1093/bjps/axs047

Strevens, M. (2008). *Depth: An Account of Scientific Explanation*. Harvard University Press.

Weatherson, B. (2016). David Lewis. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2016/entries/david-lewis/

West, G. B., Brown, J. H., & Enquist, B. J. (1997). A General Model for the Origin of Allometric Scaling Laws in Biology. *Science, 276*(5309), 122–126. https://doi.org/10.1126/science.276.5309.122

West, G. B., Brown, J. H., & Enquist, B. J. (1999). The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms. *Science*, *284*(5420), 1677–1679. https://doi.org/10.1126/science.284.5420.1677

Widom, M., & Mahan, G. D. (Invalid Date). Liquid crystal. In *Encyclopedia Britannica*. https://www.britannica.com/science/liquid-crystal

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.

Woodward, J. (2015). Methodology, ontology, and interventionism. *Synthese*, *192*(11), 3577–3599. https://doi.org/10.1007/s11229-014-0479-1

Woodward, J. (2018). Some varieties of non-causal explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations* (pp. 117–137). Oxford University Press.

Woodward, J. (2019). Scientific Explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2019/entries/scientific-explanation/

Woodward, J. (2020). Flagpoles anyone? Causal and explanatory asymmetries. *THEORIA. An International Journal for Theory, History and Foundations of Science*, *0*(0), Article 0. https://doi.org/10.1387/theoria.21921

Appendix A: How interventions account for explanatory asymmetry

In Chapter 3, I introduce the challenge of asymmetry for CEXP. As mentioned in Section 3.2, the interventionist theory of causal explanation does not address explanatory asymmetry in a way that proponents of CEXP could use to solve this challenge. This is because the interventionist theory accounts for explanatory asymmetry by using interventions. And proponents of CEXP dispense with interventions. Below, I explain how the interventionist theory establishes explanatory asymmetry using interventions in more detail.

The interventionist theory accounts for explanatory asymmetry by ruling out non-explanatory counterfactuals as false (see Woodward, 2003, p. 197f; the flagpole example is mine). It does so by counting only true counterfactuals whose antecedent is brought about by interventions as explanatory. Here are the so-called interventionist counterfactuals in the flagpole explanation:

($\mathrm{CF}_1^{\mathrm{i}}$) If the flagpole's height had been changed from x_1' to x_1'' and/ or the angle of sunrays from x_2' to x_2'' by an intervention (with respect to the flagpole's shadow's length), the flagpole's shadow's length would have been y''.

($\mathrm{CF}_2^{\mathrm{i}}$) If the flagpole's shadow's length had been changed from y' to $y''$and/ or the angle of sunrays from x_2' to x_2'' by an intervention (with respect to the flagpole's height), the flagpole's height would have been x_1''.

Recall, an intervention on X with respect to Y is a manipulation of X that changes Y only via the change in X, if at all (see Section 2.1). This notion of an intervention is defined such that $\mathrm{CF}_1^{\mathrm{i}}$ is true for at least some values of X_1 and X_2, while $\mathrm{CF}_2^{\mathrm{i}}$ is false for all values of Y and X_2, yielding asymmetry of counterfactual dependence.

First, consider $\mathrm{CF}_1^{\mathrm{i}}$. Let us focus on manipulating the flagpole's height. An intervention on the flagpole's height with respect to its shadow's length is a manipulation of the flagpole's height which changes its shadow's length through the change in the flagpole's height, if at all. Cutting the flagpole would be such an intervention. Suppose we assume that we otherwise observe the flagpole as in the original case. There is no wall between the sun and the flagpole and so on. We also assume that the sun is still somewhere above the horizon. Under these assumptions, for at least some cuts of the flagpole, its shadow's length would change. $\mathrm{CF}_1^{\mathrm{i}}$ is true for some values of X_1 and X_2.

Second, consider CF_2^i. Let us first focus on manipulating the flagpole's shadow's length. An intervention on the flagpole's shadow's length with respect to the flagpole's height is a manipulation of the flagpole's shadow's length which changes its height through the change in its shadow's length, if at all. Building a wall between the sun and the flagpole would be such an intervention. Suppose we were to build the wall. Nothing would happen to the flagpole's height. The same seems to hold for any intervention on the flagpole's shadow's length.

Let us now consider the angle of sunrays. An intervention on the angle of sunrays would be a manipulation of this angle that changes the flagpole's height via changing the angle of sunrays, if at all. Observing the flagpole a few hours later is such a manipulation of the angle of sunrays. Assuming that nobody meanwhile came to cut the flagpole or change it in any other way, no such change in the angle of sunrays would change the flagpole's height. Moreover, the same seems to hold for any intervention on the angle of sunrays. Thus, CF_2^i is false for all values of Y and X_2.

As outlined, the qualification that only true interventionist counterfactuals count as explanatory ensures asymmetry of counterfactual dependence. The flagpole's shadow's length counterfactually depends on changes in its height and the angle of sunrays brought about by interventions. But the flagpole's height does not counterfactually depend on changes in its shadow's length or the angle of sunrays brought about by interventions. For this reason, according to the interventionist theory, the angle of sunrays and the flagpole's height explain its shadow's length, but not the flagpole's shadow's length and the angle of sunrays its height. The interventionist theory yields explanatory asymmetry (as defined in Section 3.1).

Notably, the interventionist way to account for asymmetry and the quasi-interventionist one discussed in Section 7.3 and 7.5 are equivalent in the flagpole explanation (and other causal explanations). As mentioned in Section 7.3, this implies that quasi-interventions will be as successful in accounting for explanatory asymmetry in causal explanations as interventions are.

Appendix B: A primer on Lewisian semantics

In Chapter 4, I argue that it is unclear how a conservative CEXP with Lewisian semantics would solve the challenge of asymmetry. To this end, I describe how Lewisian semantics yields asymmetry of counterfactual dependence if asymmetry of overdetermination holds. Throughout Chapter 4 I assume that readers are familiar with Lewisian semantics. For readers unfamiliar with Lewisian semantics, I describe how it evaluates counterfactuals in more detail below.

Consider the example from Section 4.1 again (see e.g., Menzies & Beebee, 2020):

> ***Suzy's rock-throwing***
>
> At time t, Suzy throws a rock at a window. The rock flies through the air, hits the window and the window shatters at time t*.

Consider the following counterfactual:

> (CF_L) If Suzy had not thrown the rock, the window would not have shattered.

CF_L is a counterfactual involving the non-occurrence of an event in the antecedent and consequent, respectively. Against the background of the example, CF_L is true. But under what conditions are counterfactuals like CF_L true? Lewisian semantics provides the following answer:

> ***Lewis' counterfactual analysis***
>
> "A counterfactual "If it were that A, then it would be that C" is (non-vacuously)[58] true if and only if some (…) world where both A and C are true is more similar to our actual world, overall, than is any world where A is true but C is false." (Lewis, 1979, p. 465, see also 1986)

For example, CF_L is true iff some possible world in which Suzy does not throw the rock and the window does not shatter is more similar to the actual world than any world in which Suzy does not throw the rock but the window still shatters. Note that Lewis' counterfactual analysis demands to compare possible worlds in which the antecedent is true, e.g., in which Suzy does not throw the rock. Lewis' counterfactual analysis then raises two questions:

[58] For Lewis, a counterfactual is vacuously true if there are no possible A worlds, i.e. if it has an impossible antecedent (Lewis, 1986, p. 18f).

First, what is a possible world? For our purposes, we can think about possible worlds as follows (see also Menzel, 2017, Introduction): We regularly think that our world could have been different from how it actually is. Suzy could not have thrown the rock, the window could not have broken, the rock could have been soft and so on. The way the world actually is is just one amongst many possibilities, one of many possible worlds.[59]

Second, when is a possible world more similar to the actual world than another one? Lewis' answer is (Lewis, 1979, pp. 465–467): it depends. More precisely, Lewis argues that there are different adequate similarity measures for different contexts. What it takes for a possible world to be more similar to the actual one than another possible world differs across the contexts in which counterfactuals are expressed.

Nevertheless, as mentioned in Section 4.1, Lewis provides a standard similarity measure designed for "our usual sort of counterfactual reasoning" (Lewis, 1979, p. 457). For our purposes, the standard similarity measure is adequate, because it yields asymmetry of counterfactual dependence. Here it is:

> ***Lewis' standard measure of similarity***[60]
> "(1) It is of the first importance to avoid big, widespread, diverse violations of law.
> (2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
> (3) It is of the third importance to avoid even small, localized, simple violations of law.
> (4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly." (Lewis, 1979, p. 472)

Lewis' standard measure is best understood when put to use. Let us consider how one can use it to evaluate $\mathtt{CF}_\mathrm{L}$. Lewis' counterfactual analysis demands to ask: Are there any possible worlds in which Suzy does not throw the rock and the window does not shatter that are more similar to the actual one than all possible worlds in which Suzy does not throw the rock and the window nevertheless shatters? To answer this question, we need to look at possible worlds in which Suzy

[59] For several views of what possible worlds are see (Menzel, 2017, Section 2).
[60] As mentioned in Section 4.1, in deriving his standard measure, Lewis assumes that the actual laws of nature are deterministic (Lewis, 1979, p. 460f). This implies that two possible worlds which obey these deterministic laws perfectly are either exactly alike throughout time or not exactly alike for any period of time. Just like in Section 4.1, I throughout assume determinism in this sense.

does not throw the rock. Lewis' standard measure of similarity tells us which of these possible worlds are more similar to the actual one than others. Let us denote the actual world with W_0 and compare the following possible worlds, also mentioned in Section 4.1:

> ***W_1***: W_1 is exactly like W_0 until right before t when a small violation of law occurs such that Suzy does not throw the rock. Afterwards, W_1 unfolds without any further violations of law. W_1 then differs from W_0 in spatio-temporal facts after the small miracle right before t. In particular, in W_1, the window does not shatter.

> ***W_2***: W_2 has a different past than W_0. In particular, in W_2, Suzy does not throw the rock at time t. However, in W_2, a whole range of small violations of law occur right after t such as to make W_2 exactly like W_0 after t. The air is pushed away, Suzy remembers throwing the rock and so on. In particular, in W_2, the window shatters.

Both W_1 and W_2 are worlds in which the antecedent of CF_L, that Suzy does not throw the rock, is true. Is W_1 or W_2 then more similar to W_0?

According to Lewis' standard measure, W_1 is more similar to W_0 than W_2. In W_1, a small violation of law (a small miracle) occurs. And there is a perfect match of spatio-temporal facts with W_0 up until right before this small miracle. In W_2, a whole range of violations of law (a big miracle) occur. And there is a perfect match of spatio-temporal facts with W_0 after t. According to Lewis' standard measure, worlds that avoid big miracles are more similar to the actual world than ones that do not (see 1 in Lewis' standard measure). Thus, W_1 is more similar to W_0 than W_2.

Moreover, Lewis' standard measure is designed such that there are no possible worlds more similar to W_0 than worlds like W_1. This is because W_1 is a world with a small miracle making Suzy not throw the rock, conformity to laws after this miracle and to spatio-temporal facts of the actual world for the longest possible period until the miracle (Weatherson, 2016, p. 17).[61] W_1 is a world in which the window does not shatter. Hence, according to Lewis' counterfactual analysis, $\mathrm{CF_L}$ is true.

Thus far, I have described how Lewisian semantics evaluates counterfactuals that involve events. But how does Lewisian semantics yield an asymmetry of counterfactual dependence? To see this, let us look at the following counterfactual:

($\mathrm{CF_L^b}$) If the window had not shattered, Suzy would not have thrown the rock.

For there to be an asymmetry of counterfactual dependence in the case of Suzy's rock-throwing, $\mathrm{CF_L}$ should be true while $\mathrm{CF_L^b}$ false. The shattering of the window would then counterfactually depend on Suzy's rock-throwing but not vice versa. According to Lewis, we indeed usually hold so-called backtracking counterfactuals like $\mathrm{CF_L^b}$ to be false (Lewis, 1979, p. 457).[62] Here is how his counterfactual analysis and standard measure of similarity yield this verdict:

$\mathrm{CF_L^b}$ is true iff there is some possible world in which the window does not shatter and Suzy does not throw the rock that is more similar to the actual world than all possible worlds in which the window does not shatter and Suzy still throws the rock. Again, denoting W_0 as the actual world, consider W_1 and the following possible world, also mentioned in Section 4.1:

> ***W₃***: W_3 starts exactly like W_0. In particular, in W_3, Suzy throws the rock. However, right before t*, a small miracle occurs such that the window does not shatter. Afterwards, W_3 unfolds without any further violations of law. W_3 then differs in spatio-temporal facts from W_0 after the small miracle before t*.

Both W_1 and W_3 are worlds in which the window does not shatter. According to Lewis' standard measure, W_3 is more similar to W_0 than W_1. W_3 matches the spatio-temporal facts of W_0 for a longer period of time than W_1 (see 2 in Lewis' standard measure). Moreover, there are no possible

---

[61] More precisely: Assuming that the laws in the actual world are deterministic (as Lewis does) and that Suzy throws the rock in the actual world, one has two options to prevent Suzy's rock-throwing: (a) break one law to prevent Suzy's rock-throwing as in W_1 or (b) allow for the past to differ from the actual one such that Suzy does not throw the rock as in W_2. In case (b) one either needs to introduce a whole range of small miracles to make W_2 align perfectly again with W_0 after Suzy's failure to throw (says Lewis, see Section 4.1 on asymmetry of overdetermination) or one lets W_2 be different than W_0. In both cases, W_2 is less similar to W_0 than W_1, according to Lewis' standard measure (particularly according to 1 and 2). Hence, worlds with small miracles, conformity to laws thereafter and to spatio-temporal facts before are amongst the most similar ones in which the antecedent is true.

[62] Lewis is careful to say that there are contexts in which we hold backtracking counterfactuals to be true (Lewis, 1979, p. 457). These are outside of the scope of his standard measure of similarity.

worlds with intact windows more similar to W_0 than worlds like W_3. This is again because W_3 is a world with a small miracle, conformity to laws after this miracle and to facts of the actual world for the longest possible period (Weatherson, 2016, p. 17). In W_3, Suzy throws the rock. Thus, according to Lewis' counterfactual analysis, $\mathrm{CF}_\mathrm{L}^\mathrm{b}$ is false. Lewisian semantics establishes asymmetry of counterfactual dependence in this example.

To summarize, the key elements of Lewisian semantics are Lewis' counterfactual analysis and his standard measure of similarity. Moreover, according to this semantics, CF_L is true while $\mathrm{CF}_\mathrm{L}^\mathrm{b}$ is false. In other words, Lewisian semantics yields asymmetry of counterfactual dependence in the case of Suzy's rock-throwing, and more generally in counterfactuals involving events (if asymmetry of overdetermination holds as discussed in Section 4.1).

Appendix C: Against other strategies to solve the challenge of asymmetry

In Part II, I argue against two attempts to solve the challenge of asymmetry, amendments of CEXP using Lewisian semantics or Woodward's (2018, 2020) and Lange's (2019) inference schemes. As mentioned, beyond these two strategies, some other avenues to solve the challenge of asymmetry are not promising for proponents of CEXP. These either conflict with core commitments of the interventionist theory (and thus presumably of CEXP) or struggle with being generalizable to all explanations proponents of CEXP are interested in. Below, I elaborate on these shortcomings.

Here are two strategies to solve the challenge of asymmetry that one might propose:

1. **Pragmatic avenue**: One could appeal to pragmatic considerations to account for explanatory asymmetry. The directionality explanations exhibit derives from the explanatory context, the interests of the explaining agent or the like, so the argument would go.
2. **Ontic avenue**: One could argue that different ontic relations like causal relations or grounding relations are simply asymmetric such as to support asymmetry of counterfactual dependence (see Povich, 2018, and also 2021). As a result, a conservative CEXP yields asymmetry of counterfactual dependence, and thus explanatory asymmetry.

However, both strategies conflict with core commitments of the interventionist theory of explanation, and thus presumably also of CEXP.

First, the pragmatic avenue conflicts with the limited role of pragmatics in the interventionist theory, and thus presumably also in CEXP. As an ontic theory of explanation, pragmatic considerations play a limited role in the interventionist theory, mainly one of determining the explanandum (Woodward, 2003, p. 226ff). Presumably, proponents of CEXP aim to maintain this non-pragmatic commitment. By contrast, the pragmatic avenue demands to emphasize pragmatic aspects in CEXP beyond a limited role, e.g., that the interests of explaining agents account for the directionality of explanations. Therefore, the pragmatic avenue conflicts with the non-pragmatic commitment in the interventionist theory, and presumably also in CEXP.

Second, the ontic avenue conflicts with the non-metaphysical commitment of the interventionist theory, and at least of some proponents of CEXP. Even though the interventionist theory is an ontic theory, interventionists aim to stay agnostic about the metaphysics of causal relations (see e.g., Woodward, 2015). They demand to cite causal relations in explanations but aim to exclude an ontological discussion about causation. Instead, they use interventions to distinguish causal relationships and non-causal correlations from a methodological point of view, i.e. the point of view that such a distinction using interventions contributes to goals of causal reasoning like the discovery of manipulable relations (Woodward, 2015, p. 3581f). Likewise, at least some proponents of CEXP express a hesitation to discuss the metaphysics of ontic relations (see Jansson & Saatsi, 2019, p. 834; Woodward, 2015).[63] However, it seems that the ontic avenue would require discussing the metaphysics of ontic relations. To account for explanatory asymmetry by merely pointing to asymmetric ontic relations requires establishing under what conditions ontic relations are asymmetric. And this implies discussing their ontology, contrary to the non-metaphysical attitude at least some proponents of CEXP express. Thus, the ontic avenue conflicts with a core commitment of the interventionist theory, and at least of some proponents of CEXP.[64]

Finally, Jansson (2015) proposes another avenue to account for explanatory asymmetry. She argues that explanatory asymmetry derives from the sensitivity of the explanandum but not the explanans to changes in the conditions of applicability of a law-like generalization. For instance, in the flagpole explanation, Jansson views the presence of a light source as a condition of applicability for the trigonometric relationship between the flagpole's height, its shadow's length and the angle of sunrays (Jansson, 2015, p. 585). Here, the flagpole's shadow's length but not its height is sensitive to the presence of a light source. Thus, Jansson claims, the flagpole's height explains its shadow's length and not vice versa (Jansson, 2015, p. 587).[65]

[63] Notably, Povich is a proponent of CEXP, *and* he proposes the ontic avenue, expressing less of a non-metaphysical attitude. Note though that Povich does not analyse under what conditions ontic relations are asymmetric (Povich, 2018, 2021). As a result, his ontic avenue seems incomplete. It remains open whether Povich would object to adding such a metaphysical analysis to CEXP.

[64] Perhaps one could characterize all ontic relations from a methodological point of view, akin to how interventions are used to distinguish causal relations from non-causal correlations within the interventionist theory. Note though that this suggestion moves beyond the ontic avenue. This is because such a view would not simply appeal to the asymmetry of ontic relations and argue that this asymmetry grounds explanatory asymmetry. Instead, this suggestion would identify features of ontic relations which distinguish those that support asymmetry of counterfactual dependence from those that do not. For instance, one could argue that the quasi-interventionist version of CEXP proposed in Part III provides a notion of manipulability (using quasi-interventions) that characterizes ontic relations supporting asymmetry of counterfactual dependence. This quasi-interventionist solution, however, would not simply appeal to the asymmetry of ontic relations to ground explanatory asymmetry. Rather, it would appeal to the manipulability of some ontic relations using quasi-interventions.

[65] Jansson's treatment of the flagpole explanation differs from the one adopted in this thesis. As introduced in Section 2.1, the flagpole explanation has two explanatia, the angle of sunrays and the flagpole's height. By contrast, Jansson does not seem to think of the angle of sunrays as an explanans in this explanation.

However, Jansson's proposal struggles with being generalizable towards all explanations proponents of CEXP are interested in. This is because not all explanatory generalizations seem to have conditions of applicability. For instance, mathematical generalizations like Euler's theorem hold universally (see also Lange, 2019, pp. 3903–3906; and Jansson, 2020 for a reply).[66] And proponents of CEXP are also interested in explanations with generalizations that seem to not have conditions of applicability, such as Euler's explanation. Thus, Jansson's proposal struggles with being generalizable towards all explanations proponents of CEXP are interested in.

I conclude that the pragmatic avenue, the ontic one and Jansson's proposal are not promising as solutions to the challenge of asymmetry for proponents of CEXP.

[66] Jansson's reply relies on understanding conditions of applicability more broadly than originally as referring also to conditions of applying an explanatory model to some target phenomenon (Jansson, 2020). For instance, according to Jansson, to use Euler's theorem in an explanation of Königsberg's non-traversability requires some conditions that make the graph-theoretical model of Königsberg (see Figure 3b in Section 2.2) apply to the actual bridge system (Jansson, 2020, pp. 3–5). For example, Königsberg's bridges must be such that they can be physically crossed. And Königberg's non-traversability is sensitive to changes in such conditions but not the configuration of its bridge system. Whether Königsberg is traversable depends on whether it is physically possible to cross its bridges. But the actual configuration of Königsberg's bridge system does not depend on whether it is physically possible to cross its bridges. Thus, on Jansson's account, the configuration of Königsberg's bridge system explains its non-traversability but not vice versa.

Whatever the merits of Jansson's reply in the case of Euler's explanation, it does not seem to establish that her account is generalizable towards all explanations proponents of CEXP are interested in. In some such explanations, there seems to neither be a law-like generalization nor a model with conditions of applicability as needed for Jansson's account. For instance, it remains hard to see what the conditions of applicability are for the relationship of liquid crystals' fluid behavior and their molecule's shape to be explanatory of the fluid behavior but not of the molecule's shape on Jansson's view. Or what the conditions of applicability of the mathematical generalization that 23 is not evenly divisible by 3 would be that establish that mother's number of strawberries explains their indivisibility but not vice versa on Jansson's view. In both cases, there seems to neither be a law-like generalization nor a model with conditions of applicability.