Erkenntnis
https://doi.org/10.1007/s10670-021-00497-4

ORIGINAL RESEARCH

A Defence of Manipulationist Noncausal Explanation: The Case for Intervention Liberalism

Nicholas Emmerson[1]

Received: 24 December 2020 / Accepted: 7 November 2021

Abstract

Recent years have seen growing interest in modifying interventionist accounts of causal explanation in order to characterise noncausal explanation. However, one surprising element of such accounts is that they have typically jettisoned the core feature of interventionism: interventions. Indeed, the prevailing opinion within the philosophy of science literature suggests that interventions exclusively demarcate causal relationships. This position is so prevalent that, until now, no one has even thought to name it. We call it "intervention puritanism". In this paper, we mount the first sustained defence of the idea that there are distinctively noncausal explanations which can be characterized in terms of possible interventions; and thus, argue that I-puritanism is false. We call the resultant position "intervention liberalism" (I-liberalism, for short). While many have followed Woodward (Making Things Happen: A Theory of Causal Explanation, Oxford University Press, Oxford, 2003) in committing to I-pluralism, we trace support for I-liberalism back to the work of Kim (in: Kim (ed) Supervenience and mind, Cambridge University Press, Cambridge, 1974/1993). Furthermore, we analyse two recent sources of scepticism regarding I-liberalism: debate surrounding mechanistic constitution; and attempts to provide a monistic account of explanation. We show that neither literature provides compelling reasons for adopting I-puritanism. Finally, we present a novel taxonomy of available positions upon the role of possible interventions in explanation: weak causal imperialism; strong causal imperialism; monist intervention puritanism; pluralist intervention puritanism; monist intervention liberalism; and finally, the specific position defended in this paper, pluralist intervention liberalism.

✉ Nicholas Emmerson
nijachem@cantab.net; nje987@student.bham.ac.uk

[1] University of Birmingham, 26 Assay Lofts, 62 Charlotte Street, Birmingham B3 1BP, UK

Published online: 29 November 2021

1 Introduction

Recent years have seen growing interest in the prospect of modifying interventionist analyses of causal explanation, popularized by James Woodward (2003), in order to characterize explanations which are seemingly *noncausal* in nature.[1] One seemingly odd feature typically shared by such accounts, however, is that they jettison the core feature of interventionism: *interventions*. Indeed, the dominant position within the philosophy of science literature suggests that, roughly speaking, where it is possible to intervene upon *X*, in such a way that changes the value of *Y*, *X causes Y*. Which is to say that interventions exclusively demarcate *causal* explanations.[2]

So prevalent is this position that, until now, no one has seen fit to name it. We call it "intervention puritanism" (*I*-puritanism, for short). While dissenting voices have begun to appear (including Woodward (2018) himself), this paper represents the first sustained defence of the idea that there are distinctively *non*causal explanations which can be characterized in terms of such interventions; in other words, that *possible* interventions do not carve nature at its causal joints.[3] We call this position "intervention liberalism" (*I*-liberalism, for short).

Given the relatively recent emergence of interest in interventionism with respect to noncausal explanation, it might come as some surprise to discover that precedence for *I*-liberalism can be found as far back as the 1970s.[4] In a series of (largely overlooked) papers, Jaegwon Kim argues against *causal imperialism*, the view that *all* explanations track causal relations.[5] In 'Causes and Counterfactuals' (1973) Kim

[1] While Woodward's (2003) interventionist analysis of causation is generally taken to be 'the standard philosophical account' (Wilson, 2018:18), Woodward himself attributes the term "intervention" to Meek & Glymour (1994) and Pearl (2000). Also see: Hitchock (2001), Pearl (2009) and Briggs (2012).

[2] See, for example: Woodward (2003, 2015, 2018; Bokulich 2011; Leuridan 2012; Saatsi and Pexton 2013; Harinen 2014; Pexton 2014; Baumgartner & Gebharter 2015; Rice 2015; Romero 2015; Reutlinger 2016, 2017, 2018; Baumgartner & Casini 2017; French and Saatsi 2018; Khalifa et al., 2018, 2020; Saatsi 2018; Jansson & Saatsi 2019; Lange 2019).

[3] There is currently ongoing debate regarding the explanatory status of *impossible* interventions, with several authors having recently argued for their application in noncausal explanations across mathematics, logic, and metaphysics (see e.g., Schaffer (2016, 2017); Wilson (2016, 2018, 2021); Baron et al (2017); Baron et al (2020); Reutlinger et al (2020); Baron & Colyvan (2021); and Baron (*forthcoming*). It is widely understood that such impossible interventions require the rejection of traditional counterfactual semantics (see e.g., Baron & Colyvan 2021: 564–567). For the purposes of *this* paper, however, the reader ought to assume that where we use the term "intervention", unless explicitly stated otherwise, we mean "physically possible intervention". One *prima facie* reason for limiting our account in this way is that (as we shall soon see) such cases do not require any substantial modification to be made to the typical interventionist methodology, and thus constitute the most robust form of counterexample to *I*-puritanism. What is more, as we discuss further in Sects. 5 and 7, appealing to the role of impossible interventions does not, by itself, constitute a rejection of *I*-puritanism.

[4] In fact, I believe that something like this idea can be traced back to C. S. Peirce (1931–58), who argues that even pure mathematics and logic concern 'operations on diagrams, whether external or imaginary, [which] take the place of the experiments upon real things that one performs in chemical or physical research' (*Collected Papers* 4:530; 1905). Also see Peirce (*Writings* 3:41; 1872). While certainly worthy of further investigation, such a task is strictly beyond the scope of this paper.

[5] Causal imperialism (a term borrowed from Bokulich [2018]) is most closely associated with Railton (1981); Lewis (1986); Strevens (2008); and Skow (2014). This position is to be distinguished from *I*-puritanism, which is neutral with respect to whether all explanations are causal explanations. As we

highlights cases of asymmetric counterfactual dependence which motivate the existence of distinctly noncausal explanations. In 'Noncausal Connections' (1974), Kim goes further, arguing that both causal and noncausal dependence can be characterized in terms of the "bringing about" relation. Where *A* depends upon *B* (causally or otherwise), Kim suggests, we can bring about *B* by first bringing about *A,* but not vice versa.[6]

More recently, Woodward's (2003) manipulationist account of causal explanation has given Kim's intuitive conception of "bringing about" a formal characterization through the notion of a possible intervention; making use of structural equation models to encode an asymmetric pattern of interventionist counterfactuals. In the first two sections of this paper, we revisit Kim's analysis of noncausal explanatory dependence and attempt to reconcile his position within a contemporary interventionist framework.

In Sect. 2, we outline Kim's motivation for suggesting that noncausal counterfactual dependence can be characterized in terms of "bringing about". In Sect. 3 we introduce Woodward's (2003) analysis of causal explanation and demonstrate how Kim's analysis of noncausal explanation can be fruitfully accommodated with the framework of structural equations models and interventionist counterfactuals. As it transpires, the sort of noncausal explanatory dependence highlighted by Kim can be happily cashed out in terms of possible interventions.

As was mentioned at the outset, *I*-liberalism has been staunchly opposed within the recent philosophy of science literature. There are two distinct (but related) debates which have served as focal points of such scepticism. The first concerns a particular type of noncausal explanation: constitutive explanation. In response to Carl Craver's (2007a, 2007b) attempts to characterise constitutive explanation in terms of *mutual manipulability*, one common counter has been that, since manipulability is an essentially causal notion, Craver's account fails as a characterisation of *non*causal explanation.[7]

The second such debate concerns a slew of recent interest in providing a *monistic* account of the asymmetry of causal and noncausal explanation. Here it is once again widely assumed that possible interventions characterize only causal relationships and, as such, that they must be jettisoned when providing an account of explanation

Footnote 5 (continued)
shall see, many hold that while noncausal explanations *do* exist, they are not characterizable in terms of interventionist counterfactuals. We discuss these distinctions in more detail in Sect. 7.

[6] A popular position within both philosophy of science and metaphysics suggests that 'explanations must depict dependence relations' (Potochnik, 2017:105). While Kim's analysis focuses on causal and noncausal *dependence*, following the likes of Ruben (1990); Kim (1994); Strevens (2008); Audi (2012); Craver (2014) Schaffer (2016); and Kovacs (2017), we shall assume that wherever one finds a dependence relation of the sort in question, an explanation follows. While there remain dissenting voices (Dasgupta 2017; Khalifa et al., 2018; Taylor 2018; Thompson 2018), the majority of those involved in debate surrounding noncausal explanation would at least concede something close to this position.

[7] See, e.g. Craver's (2007a; 2007b); Leuridan (2012); Harinen 2014; Romero 2015; Baumgartner & Gebharter 2016; Cassini & Baumgartner 2016; and Krickel 2018.

which unifies causal and noncausal instances.[8] In both debates, however, motivation for *I*-puritanism is remarkably thin on the ground, principally relying upon Woodward's (2003) own commitment to this position.

In Sect. 4, we demonstrate that *I*-liberalism *can* accurately characterise an archetypal case of constitutive explanation: the nastic movement of *Mimosa Pudica.* In Sect. 5 we argue that, while Woodward (2003) does appear to commit himself *I*-puritan, such an interpretation of his interventionist framework is far from obligatory. What Kim's intuition regarding the "bringing about" relations shows, is that causal relationships do not exhaustively describe the ways in which agents can manipulate the world around them.

In Sect. 6, we consider one of the few arguments against *I*-liberalism which does not rely upon Woodward's own *I*-puritanism. In the process of arguing against explanatory monism, Kareem Khalifa et al (2020) suggest that *I*-liberalism is untenable precisely because it cannot distinguish between cases of genuine noncausal dependence and cases analogous to spurious correlation resulting from some common explanatory source. Conversely, we argue that, while analogous spurious correlations *do* arise with respect to noncausal explanation, constitutive explanations are not among them and, what is more, that *I*-liberalism is perfectly capable of dealing with such cases.

In the final section, we provide a novel taxonomy of available positions upon the role of interventions in explanation. We highlight six such positions: weak causal imperialism; strong causal imperialism; monist intervention puritanism; pluralist intervention puritanism; monist intervention liberalism; and finally, the position defended in this paper, *pluralist intervention liberalism.*

2 Kim on Noncausal Connections

In 'Causes and Counterfactuals', Kim (1973) highlights several cases which appear to undermine the causal imperialist claim that counterfactuals exclusively express *causal* dependencies.[9] Take the relationship between Xanthippe's becoming a widow and the death of her husband, Socrates. Does Socrates' death *cause* Xanthippe's widowhood? There are reasons to think not. First and foremost, these events are spatially "discontiguous". As Kim highlights, to accept that such causal action could be 'propagated instantaneously through spatial distance' would be an unforgiveable afront to physics (1974/1993:13).[10] Second, presuming that individual causal relations instantiate nomic regularities, it is difficult to think of a contingent empirical law capable of subsuming these events (Kim, 1974/1993:13).

[8] E.g. Saatsi and Pexton (2013); Jansson (2015); Reutlinger (2016, 2017); French and Saatsi (2018); Lange (2019); Khalifa et al (2020).

[9] Lewis (1973), being Kim's explicit target here.

[10] It is interesting to note that Woodward believes that his interventionist methodology provides reason for denying that spatiotemporal contiguity is a defining characteristic of causation (2003:36). Although, in more recent work, Woodward (2018) accepts this particular example as an instance of distinctively noncausal dependence.

If Socrates' death is not the cause of Xanthippe's becoming a widow, then perhaps whatever caused Socrates death was *itself* the cause of Xanthippe's widowhood. This interpretation of the situation even seems to allow for a 'nice Humean law' which subsumes hemlock consumption and widowhood: 'given the law, let us assume, that anyone who drinks hemlock dies, we have the law—at least a Humean regularity—that anyone whose husband drinks hemlock becomes a widow' (1974/1993:30).

The problem with this interpretation is that the only route from hemlock to widowhood seems to *go through* death. If Socrates' having ingested hemlock causes Xanthippe's widowhood, it does so only by first causing his death, which puts us back where we started. Consequently, if neither Socrates' drinking hemlock, nor any other apparent cause of his death could count as the cause of Xanthippe's becoming a widow, we can only conclude, according to Kim, that Xanthippe's being widowed has no cause at all.

And yet, these events are obviously connected in some sense. Indeed, notwithstanding those features discussed above, there are some clear similarities between this type of noncausal connection and archetypal causal explanations. First and foremost, Kim argues, the sort of explanatory asymmetry which we might expect from a cause-and-effect relationship can be drawn out when considering the counterfactual conditionals related to these events:

If Socrates had not died at *t*, Xanthippe would not have become a widow at *t*.
If Xanthippe had not become a widow at *t*, Socrates would not have died at *t*.

The counterfactual dependence between these two events is 'irreversible'; while (1) is straightforwardly true, in response to (2), Kim notes, 'we would more likely alter the marital condition of Socrates than tamper with the fact of his death at *t*' (1974/1993:24). This irreversibility becomes clearer when we examine a second form of asymmetry which Kim notes with respect these events: asymmetry in the agency relation. Consider the following counterfactuals:

By bringing about Socrates' death, we could bring about Xanthippe's widowhood.
By bringing about Xanthippe's widowhood, we could bring about Socrates' death.

What Kim's intuition regarding these counterfactuals suggests, is that if we wished to make Xanthippe a widow, facilitating Socrates' death would be the best (indeed only) way to go about doing it. On the other hand, attempting to make Xanthippe a widow would not be an "effective strategy" (to use a phrase of Cartwright's (1979)) to bring about Socrates' death. In much the same way, while we might increase the length of a pendulum to bring about an alteration in its period of swing, altering its period of swing would not be an effective strategy to bring about an increase in its length. Such events are not the result of coincidence or brute fact, but 'are determined by other events; their occurrence is completely dependent on the occurrence of others, but this is not to say that they are causally determined by them' (Kim, 1974/1993:30).

In concluding these observations, Kim tentatively puts forward the thesis that both causal and noncausal connections might be characterised, monistically, in terms of a single unifying relation "*R*": 'a broad relation of dependency that subsumes as special cases the causal relation and other dependency relations' (1974/1993:27). Unfortunately, Kim (1974/1993) does not provide a substantive account of what this relation might be.[11] With the benefit of Woodward's (2003) methodology, however, we believe that a more formal characterization to Kim's intuitions regarding the "bringing about" relation can be given in terms of interventionist counterfactuals and structural equations. It is to this task that we turn in the next section.

3 An Interventionist Account of "Bringing About"

According to Woodward, any attempt to characterise causal dependence ought to begin by considering the practical utility of our notion of causation; what does causal knowledge allow us to achieve that information about mere regularity or correlation, will not? (2003: 28). In answer to this question, Woodward suggests that 'it is heuristically useful to think of explanatory and causal relationships as relationships that are potentially exploitable for the purposes of manipulation and control' (2003: 25). It is manipulability, then, that distinguishes explanatory counterfactuals from nonexplanatory counterfactuals (the latter of which arise as the result of *mere* correlation).[12]

To say that X causes Y, on this picture, is to say that Y would change in value under some suitable intervention that changed the value of X. Where an intervention on X with respect to Y 'changes the value of X in such a way that if any change occurs in Y, it occurs only as a result of the change in the value of X and not from any other source' (Woodward, 2003:14). More formally, I is an intervention on X iff:

I. I causes X;
II. I acts as a switch for all 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;

[11] This is likely due, at least in part, to historical timing. While others (e.g. Collingwood 1944; Gasking 1955; and von Wright 1975) had already argued that 'causes are, as it were, levers for moving effects', such "agential" or "manipulationist" accounts of causal explanation were thought to run into 'intractable difficulties', and were largely abandoned (Hausman, 1982:45). It should come as no surprise then, that Kim's project of accounting for both causal and noncausal explanation in terms "bringing about" found little contemporaneous support. By the turn of the twentieth century, however, the tide had well and truly turned. Thanks, in no small part, to a spirited defence by Menzies and Price (1993), which updated several crucial elements of previous agential theories (such as relinquishing a commitment to determinism), and convincingly circumvented many of the seemingly intractable difficulties mentioned above.

[12] Where it is possible to intervene upon X with respect to Y in this way, one might alternatively say that X is exploitable for the purposes of manipulation Y. As such, we shall use the terms *intervention* and *manipulation* interchangeably in what follows. We shall discuss the sort of nonexplanatory counterfactuals which arise as a result of mere correlation in Sect. 6.

III. 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 $I \to X \to \to Y$ 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;

IV. I is (statistically) independent of any variable Z that causes Y and that is on a direct path that does not go through X. (Woodward, 2003:98).

As we have already seen, Kim is well aware of this agential dynamic to dependence. Indeed, his analysis of the asymmetric relationship between dependent events comes tantalizingly close to the core claim of Woodward's interventionism. Kim argues that the asymmetry of the agency relation, the sense in which 'by bringing about the cause, you bring about the effect', is a *result* of the asymmetry 'between states or events brought about by the action' (1974/1993:25).

While he does not use the term, it is clear that *manipulation* is something like the notion which Kim is intending to highlight with his discussion of the connection between agency and dependence. Indeed, the relationship between Socrates' death and Xanthippe widowhood, appears to fit nicely into the sort of structural equations utilized by interventionists in modelling causation. Such a model consists of:

- A set of variables representing features of reality, in this case:

  C: Whether Socrates dies.

  E: Whether Xanthippe is a widow.

- A set of structural equations linking the values of these variables according to reality's causal structure, where '$\to$' expresses counterfactual dependence:

  $\mathrm{E} \to \mathrm{C}$

- And, an assignment function specifying which values the variables actually take:

  $\mathrm{C} = 1; \mathrm{E} = 1$

For C to be considered a cause of E, it must be possible to intervene upon C, altering its value from '$\mathrm{C}=1$' to '$\mathrm{C}=0$', in such a way that will result in a change in the value of '$\mathrm{E}=1$' to '$\mathrm{E}=0$'. This means that it ought to be possible to intervene upon Socrates death in such a way that also prevents Xanthippe's becoming a widow. Suppose, for example, that Crito was to knock the hemlock from Socrates' hand before it could be consumed.[13] In this scenario, as a direct result of Crito's

[13] While Crito's knocking the hemlock from Socrates' hand is itself a causal process, the dependency relation which it speaks to, holding between Socrates death and Xanthippe's widowhood, is clearly noncausal. The idea that causal processes can give rise to noncausal explanations is not novel. We discuss further instances of this surprising detail, related to constitutive explanation, in Sect. 4.

intervention, Socrates would survive, so ‘C = 0’ (assuming all other variables are held fixed), what is more, as a direct result of Socrates’ survival, Xanthippe would not become widowed, so ‘E = 0’.

This interventionist analysis of Socrates and Xanthippe’s situation also allows us to cash out the sort of asymmetry which Kim highlights as a common factor in cases of both causal and noncausal dependence. In line with *III* above, it is precisely because any possible intervention upon Xanthippe’s widowhood *must* go through Socrates’ death, which suggests that the dependency here is asymmetric. Which is to say, Xanthippe’s widowhood is not exploitable for the purposes of manipulating Socrates’ death. And yet, for Kim, there are clear reasons for thinking that Socrates’ death and Xanthippe’s widowhood are *not* related as cause and effect in any ordinary sense.

What this appears to suggest is that possible interventions cannot stand as a useful dividing line between causal and noncausal explanation. The possibility of intervening simply does not carve nature at its causal joints. This conclusion will obviously come as a blow to *I-puritans*. Since, without possible interventions to play this role, it is not obvious how we are to distinguish these different types of explanation.[14] For those without such pre-theoretical commitments, however, *I*-liberalism ought to hold some intuitive appeal. Afterall, possible interventions stand as an addition to the metaphysical toolkit through which noncausal dependence can be analysed.

One area where our analysis has obvious application is the ongoing discussion surrounding *constitutive* explanation: a popular hunting ground for *I*-puritans. In response to Craver’s (2007a, 2007b) attempts to define constitution in terms of symmetrical interventions, many have argued that this account fails on *I*-puritan grounds. Since interventions are assumed to designate causal relations, and causal relations *alone*, it is argued that Craver cannot make sense of the noncausal dynamic to a phenomenon’s being constituted by its spatiotemporal parts. Although, in the next section, we do the apparently impossible, and successfully apply the *I*-liberalist methodology described above to an example of constitutive explanation.

4 *I*-Liberalism and Constitutive Explanation

Among philosophers of science, constitutive explanations are typically taken to be a form of *mechanistic* explanation, where a mechanism consists of entities/part/objects and their activities/interactions/operations.[15] Constitutive mechanistic explanation is often distinguished from *etiological* mechanistic explanation. In the latter case, some mechanism explains a phenomenon for which it is *causally* responsible, whereas in the former, a phenomenon is explained by the underlying mechanism which *constitutes* it.

[14] See Wilson (2020) for a thorough survey of plausible means by which one might seek to classify causal and noncausal dependence.

[15] See, e.g. Machamer et al. (2000); Craver & Darden (2002); Craver (2007b); Illari & Williamson (2012); Glennan (2017).

Take, for example, the nastic movement of *Mimosa pudica*.[16] Nastic movements occur in plants and fungi as a response to environmental stimuli (*thigmonasty*), with *Mimosa* being the most heralded example due to the dramatic nature of the response. Such movement is constituted by a release of potassium ions in the plant's pulvini cells, which lowers the cell's turgor pressure (pressure exerted on the cell wall due to exosmosis) and, in turn, collapses the cell's parenchyma tissue, constricting the vascular strand serving as a hinge (Esau, 1965).

This explanation allows us to identify the three parts of *Mimosa* that are involved in the phenomena of nastic movement (E*): the potassium ions in the pulvini cells (C^*_1), the turgor pressure of the pulvini cells (C^*_2), and the parenchyma tissue of the pulvini cells (C^*_3). As we saw in the previous section, in order to capture the noncausal dependence at play here, we ought to be able intervene upon the *Mimosa's* pulvini cells in such a way that will also affect the plant's nastic movement, but not vice versa. And this is exactly what we see.

Variables:
C^*_n: Whether the parenchyma tissue of the *Mimosa's* pulvini cells collapse.[17]
E*: Whether the *Mimosa* exhibits nastic movement.
Structural equations:
$E^* \rightarrow C^*_{\mathrm{n}}$
Assignment:
$C^*_n=1$; $E^*=1$

Just as with the case of Socrates and Xanthippe, the nastic movement of the *Mimosa* can be manipulated through an intervention upon its pulvini cells. For example, administering potassium channel blockers (such as peptides containing the integrin-binding sequence RGD [Arg-Gly-Asp] [Jaffe et al., 2002]), restricts potassium ions in the pulvini cells from affecting the cell's turgor pressure and, as such, prevents nastic movement from occurring; so here '$C^*_n=0$' and, as a result, '$E^*=0$'.

This interventionist approach to constitutive explanation also allows us to cash out the sort of explanatory asymmetry which Kim highlights as a common factor in cases of both causal and noncausal dependence. Once again, in line with condition *III*, it is precisely because there is no possible intervention upon the *Mimosa's* nastic

[16] Other examples of constitutive mechanistic explanation abound: Spatial memory (Bechtel 2008; Craver 2007b); action potential (Craver 2007b); the heart (Bechtel and Abrahamsen 2005; Glennan 2010; Craver & Darden 2013); cells synthesizing proteins (Craver and Darden 2013; Darden 2002; Machamer et al. 2000); long-term potentiation at synapses of neurons (Craver & Darden 2001; Craver & Darden 2013; Craver 2007b; Machamer et al., 2000).

[17] Here, we have condensed C^*_1–C^*_3 into a single variable. The relationship between spatio-temporal parts of a constitutive mechanism is typically taken to be causal (i.e. release of potassium ions in the pulvini cells causes the cell's turgor pressure to drop). However, our principal interest is in the *non*causal relationship between the constitutive mechanism and the phenomenon to be explained, as such combining these variables allows us to maintain the noncausal character of the arrow in the structural equations. Presuming, for the sake of brevity, that under experimental conditions we can guarantee that the pulvini cells' parenchyma tissue will collapse *only* when the release of potassium ions decreases the turgor pressure within the cell, this ought to make no difference to our overall argument.

movement that does not *go through* the *Mimosa's* pulvini cells, which suggests that the dependence here is asymmetric. Which is to say, the *thigmonsasty* of the *Mimosa* is not exploitable for the purposes of manipulating its pulvini cells.

However, in his own interventionist account of constitutive explanation, Craver's (2007a, 2007b) has suggested that it is in fact *mutual* manipulability that defines such noncausal mechanisms. Craver argues that while 'one can change the explanandum phenomenon by intervening to change a component [of a mechanism]', one can *also* (contrary to our account) 'manipulate the component by intervening to change the explanandum phenomenon' (2007b:153). As such, Craver concludes that all constitutive dependency relationships are "bidirectional".

Yet, Craver's account is clearly problematic on two related fronts.[18] First, as Romero (2015), Baumgartner and Gebharter (2016), and Krickel (2018) highlight, and the example of *Mimosa* demonstrates, manipulations of the latter variant, whereby a component is manipulated via the explanandum phenomenon, are impossible by Woodward's (2003) definition of an intervention.[19] Second, supposing such "top-down" interventions were possible, given that interventions are intended to characterise explanatory relations, this would suggest that constitution entails explanatory *symmetry* (Schindler, 2013). And, as Khalifa et al. note, 'a surefire way to embarrass a theory of explanation is to show that it fails to respect the commonsense idea that explanation is an asymmetric relation' (2018:1).[20]

The *I*-liberal interpretation of such mechanisms presented above, can easily avoid both of these issues, preserving the explanatory asymmetry which forms the heart of Kim's desire for a unifying account of causal and noncausal dependence while, at the same time, ruling out the sort of top-down intervention which Craver (problematically) believes to be characteristic of constitutive explanation.[21] There is a further

[18] For further criticisms, not discussed here, see Harinen (2014).

[19] Romer (2015) and Baumgartner & Gebharter (2016) argue that such interventions are actually 'fat-handed' rather than outright impossible. A fat-handed intervention is an intervention which violates "III." in as much as it manipulates both the mechanistic components and the phenomena in question *at the same time*, effectively serving as a common cause of both. While Woodward's (2003) definition of an intervention can, according to Romer (2015) and Baumgartner & Gebharter (2016) be altered to accommodate such a notion, Krickel (2018) has argued that this approach has severe limitations. In so far as our own position takes such interventions to be impossible, and thus beyond the scope of *I*-liberalism, we preserve more of the core of Woodward's (2003) original definition, and as such, stay truer to the character of his manipulationist account of explanation.

[20] Given the obvious difficulties which is raises, one might well wonder why Craver introduces the notion of mutual manipulation at all. His motivation is principally to try and make sense of important "top down" research strategies within the life sciences, distinguishing between interference experiments, stimulation experiments and activation experiments (2007b: 146–157). However, Baumgartner and Gebharter have recently argued that '[e]mpirical evidence does not only consist in correlational evidence resulting from suitable manipulations' and, furthermore, that top-down experimentation can be made sense of without the need for top-down interventions.

[21] The question of what distinguishes constitutive explanation from both other forms of noncausal dependence, and causal dependence, is an interesting one. Unfortunately, we do not have the space here to discuss this topic at length. However, we would point out that there are obvious features of constitutive explanation which could serve as useful distinguishing characteristics, e.g. a mechanism and the phenomena which it explains are typically taken to share the same spatio-temporal location, which distinguishes such cases from noncausal explanations like Kim's example of Socrates death and Xanthip-

apparent issue with Craver's account, however, which is of much greater interest to us. In 'Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms' Leuridan (2012) suggests that Craver's account entails that constitutive explanations are, in fact, *causal* explanations.

Leuridan's (2012) argument is that one cannot get away with embedding an account of constitutive explanation within an interventionist framework and emerge with a characterisation of *non*causal explanation: interventions, in other words, highlight *only* causal relations. Indeed, Baumgartner and Gebharter (2016) agree, noting the following slogan from Woodward: 'no causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference' (2003:61). This slogan is, of course, the central tenet of *I*-puritanism.

The literature surrounding constitutive explanation is not the only area where we find support for *I*-puritanism. Another topic which has elicited a great deal discussion surrounding the essentially causal character of interventions is the broader project of providing a *monistic* account of explanation: 'an analysis that accommodates causal and noncausal explanations, and accounts for the asymmetries of both' (Khalifa et al, 2018). Here too, Woodward's (2003) interventionism has taken centre stage (e.g. Reutlinger, 2016, 2017, 2018).

Debate surrounding explanatory monism, and debate surrounding mechanistic explanation are closely connected. For example, a successful account of explanatory monism would, presumably, apply to mechanisms (both constitutive and etiological) as a limiting case.[22] It is no surprise, then, to find that both debates have motivated their commitment to *I*-puritanism along very similar lines. The principal motivation for the essentially causal character of dependencies characterized by interventions, is that Woodward (2003) himself assumes such an *I*-puritan stance. In the next section, however, after discussing explanatory monism in more detail, we argue that, although Woodward (2003) does support *I*-puritanism, this conclusion is not *entailed* by his analysis of causal explanation.

5 Woodward's *I*-Puritanism

An *almost* universal feature of recent attempts to provide a monistic account of explanation has been the abandoning of structural equation models and interventionist counterfactuals with respect to noncausal explanation.[23] Indeed, this appears to be a rare point of agreement, even among those who take the monist framework to be something other than counterfactual (e.g. Khalifa et al., 2018), and those who

Footnote 21 (continued)

pe's widowhood, which are not so constrained (See: Leuridan 2012; Romer 2015; Craver 2007b; Baumgartner and Gebharter 2016).

[22] Although, this is not to say that the existence of noncausal mechanistic explanation is committal with respect to monism.

[23] E.g. Saatsi and Pexton (2013); Jansson (2015); Reutlinger (2016, 2017); French and Saatsi (2018); Lange (2019); Khalifa et al (2020).

advocate explanatory pluralism (e.g. Lange, 2019). To take a popular example, Reutlinger (2017) proposes the following non-interventionist counterfactual criteria for a monistic explanation, where ‘G_1,…, G_m’ comprise generalizations, and ‘S_1,…, S_n’ comprise auxiliary statements:

I. *Veridicality condition*: G1, …, Gm, S1, …, Sn, and E are (approximately) true.
II. *Implication condition*: G1, …, Gm and S1, …, Sn logically entail E or a conditional probability P(E|S1, …, Sn) – where the conditional probability need not be ‘high’ in contrast to Hempel’s covering-law account.
III. *Dependency condition*: G1, …, Gm support at least one counterfactual of the form: had S1, …, Sn been different than they actually are (in at least one way deemed possible in the light of the generalizations), then E or the conditional probability of E would have been different as well.

And yet, as Roski (2020) highlights, in stripping Woodward’s (2003) account of the mechanism which characterizes causal asymmetry, namely interventions, Reutlinger’s account appears to suffer from the same embarrassing explanatory symmetry which plagues Craver’s (2007a, 2007b) account of constitutive mechanistic explanation. Reutlinger is not alone along monists here.[24] As Lange similarly argues with respect to many other recent attempts to characterize explanatory monism:

> ‘These attempts recognize that even when there is explanatory asymmetry, there may be symmetry in counterfactual dependence. Therefore, something more than mere counterfactual dependence is needed to account for explanatory asymmetry’ (2019: 1).

In light of the predicament in which symmetry places monist accounts of explanation, it would seem natural to expect to find some convincing reasons for rejecting the story which was told in the first half of this paper; that interventions stand to characterize certain explanatory asymmetries across both causal and noncausal instances. Strangely, however, there is very little in the way of argument put forward in defence of this stance.

The principal motivation for this position appeals to Woodward’s (2003) claim that possible interventions serve to illuminate exclusively causal dependencies. This intuition is frequently deployed within the recent literature. As Lange himself suggests that ‘the notion of an intervention is a causal notion and so is not obviously applicable to non-causal explanation’ (2019: 2).[25] Jansson also argues that interventions cannot help in characterising the asymmetry of noncausal explanation, since

[24] We used this analysis as an example because Reutlinger’s (2016, 2017, 2018) account stands as an exception to the rule that ‘precise formulations of [explanatory monism] are few and far between’ (Khalifa et al, 2018: 2).

[25] It is important to note, however, that Lange does not support explanatory monism. Rather he suggests that ‘the order of explanatory priority is fixed by different considerations in different non-causal explanations’ (2019: 24).

'the solution is given in terms of interventions, [and] these are cashed out in causal terms in Woodward [2003]' (2015: 22, fn. 48). Similarly, Saatsi and Pexton argue that although Woodward 'happily welcomes the possibility that the counterfactual aspect of his account may come apart from its causal aspect', this element of Woodward's account 'should not be wedded to a causal manipulationist interpretation of explanatory modal information' (2013: 614).[26]

However, while Woodward's (2003) does indeed support such an *I*-puritan reading of his interventionist framework, an *I*-liberal interpretation is by no means ruled out. In summarizing the interventionist mantra, he suggests that *any* successful explanation ought to be accompanied by 'a hypothetical or counterfactual experiment that shows us that and how manipulation of the factors mentioned in the explanation… would be a way of manipulating or altering the phenomenon explained' (Woodward, 2003: 11). What the discussion of Kim's example from the previous section suggests, we have maintained, is that causal relationships do not exhaustively describe the ways in which the world might be manipulated or altered.

Woodward (2003) holds that what distinguishes explanatory counterfactuals from non-explanatory counterfactuals (which highlight *mere* correlations) is that only the former allow for the possibility of manipulation. This is not to say, of course, that all interventionist counterfactuals pick out causal relations per se*,* but instead, that no interventionist counterfactuals pick out relations of mere correlation.[27] This understanding of interventions is perfectly compatible with the idea that *some* interventionist counterfactuals highlight possible manipulations which are not distinctly causal in nature.

The idea that interventions might stand as a useful distinguishing factor between causal and noncausal explanations can also be traced to Woodward (2003: 220–221). While both causal and noncausal patterns of dependence ought to be able to support counterfactuals (which in turn support "what-if-things-had-been-different" questions), in the latter case, according to Woodward, these counterfactuals cannot be interpreted in terms of interventions. It is important to note that this conclusion is reached on the basis of a single example: the dependence of the stability of planetary orbits on the dimensionality of space–time. Woodward argues that 'it seems implausible to interpret such derivations as telling us what will happen under interventions on the dimensionality of space–time' (2003: 220).

[26] As Woodward himself highlights, 'Woodward (2003) (tacitly and without explicit discussion) adopted the common philosophical view that causal (and causal explanatory) relationships contrast with relationships of dependence that hold for purely conceptual, logical, or mathematical reasons' (2018: 121).

[27] While, as Baumgartner and Gebharter (2016) point out, Woodward (2003: 61) suggests that there can be 'no difference in manipulability relations without a causal difference' this claim is *not* entailed by the definition of an intervention discussed in Sect. 3, of this paper. If this position is correct, it is not obviously so.

The exact reason for this implausibility is not mentioned, although the most obvious candidate is that such an intervention is nomologically impossible.[28] However, Woodward (2003) does not consider any interventions of the type discussed in the previous sections. There is nothing nomologically impossible, for example, about the prospect of intervening upon the *Mimosa's* pulvini cells in order to elicit *thigmonasty*. And, as we saw in Sect. 2, intervening upon Socrates' death for the purposes of manipulating Xanthippe's widowhood is, not only possible, but provides an accurate characterization of the explanatory asymmetry which interests Kim (1974:1993).

Given the frequency with which Woodward (2003) is appealed to in defence of *I*-puritanism, there a quiet irony in his having recently weakened his own commitment to this stance. In 'Some Varieties of Non-Causal Explanation', Woodward briefly discusses cases of noncausal explanations where 'at least some of the variable figuring in the candidate *explanans* are possible targets for manipulation… but the *connection* between these and candidate *explanandum* seems (in some sense) purely mathematical' (2018: 130).

However, as Reutlinger et al. explain, '[i]n the mathematical case, this involves supposing that mathematical facts were different. But on the standard philosophical accounts of mathematics, mathematical truths are necessary', as such intervening in such cases 'would seem to be deeply problematic' (2020: 10). Lange (2019) explicitly references the example used by Woodward (2018) (the traversibility of Königsberg's famous bridges), as requiring interventions which are impossible to perform. This is not to say that there have not been attempts to make sense of the explanatory potential of such impossible interventions, but such accounts are strictly speaking beyond the scope of this paper.[29]

Despite this, there are reasons to think that Woodward (2018) would be sympathetic to the central aim of this paper: providing the first robust defence of the claim that there are distinctively noncausal explanations which can be characterized in terms of *possible* interventions. Woodward *does* now consider interventions of the type discussed in the previous sections: 'there is an obvious sense in which it is true that by manipulating whether or not Socrates dies, one can alter whether Xantippe is a widow' (2018: 121). Further suggesting that such explanations are more naturally described by locutions such as "brings about by" than "causes" (2018: 131). While Woodward does not provide a great deal of detail concerning how he expects such interventions to fit within his earlier interventionist framework, we can see little reason for him to reject the methodology laid out in this paper.

More recently still, Khalifa et al (2020) refer to a position very close to our *I*-liberalism, by the (none-too-pithy) title *quasi-interventionist change-relating*

[28] Interestingly, elsewhere Woodward seems to have little issue with the notion of impossible interventions. He argues, for example, that '[e]ven in purely theoretical contexts, causal claims should be understood as telling us about the results of hypothetical manipulations; it is just that we cannot, at least at present, carry out these manipulations' (2003: 37).

[29] See e.g., Baron et al (2017); Baron et al (2020); Reutlinger et al (2020); Baron & Colyvan (2021); and Baron (*forthcoming*). We discuss how impossible interventions fit into the landscape of stances on the nature of the distinction between causal and noncausal explanation in more detail in Sect. 7.

counterfactual monism (QCM). They note that the only revision to Woodward's original characterization of an intervention which is required here, is replacing *I* with something like 'X counterfactually depends on *I*' (Khalifa et al, 2020: 6).[30] This interpretation seems to capture the spirit of "*R*" and Kim's desire to subsume 'as special cases the causal relation and other dependency relations' (1974/1993: 27). As Khalifa et al. suggest, 'if this quasi-interventionist approach captured every kind of explanation, causes would just be a limiting case' (2020: 6).[31]

Interestingly, however, Khalifa et al (2020) introduce QCM in the process of arguing that an account of explanatory monism based upon it is untenable, because such methodology is incapable of dealing with a familiar type of noncausal explanation: constitutive explanation. The principal motivation for this stance being that possible interventions are apparently incapably of distinguish between constitutive explanations and spurious correlations resulting from common explanatory dependence.

While a defence of monism itself is beyond the scope of this paper, in Sect. 4, we argued that a benefit of our own position is that it *can* account for the explanatory asymmetry of constitutive explanation. As such, in the next section, we demonstrate that *I*-liberalism is perfectly capable of drawing a distinction between genuine noncausal constitutive dependence and spurious correlations arising as a result of a common explanatory source.

6 Noncausal Interventions and Common Explanatory Dependencies

Khalifa et al (2020) argue, citing Ylikoski (2013), that while 'causal relata are metaphysically independent entities, constitutive relata 'are not independent existences, so one cannot think of an intervention on the basis that would not also be an intervention on the system' [2013: 284]' (2019: 7). As a result, they suggest that an intervention upon a system's components or organization (the explanans), would also be a direct intervention upon the explanandum. Consequently, according to Khalifa et al (2020), an intervention in this case would violate Woodward's principle that '[a]ny directed path from *I* to *Y* goes through *X*' (2003: 93). This is because such an intervention would apparently act as a "common cause" of both *X* and *Y*.

If true, this would indeed be a troubling result. Difficulties in distinguishing instances where variables are spuriously correlated, owing to a common explanatory dependency, have historically plagued theories of explanation. Indeed, it is a principal motivation for the adoption of interventionist methodology that, where *A* and

[30] It is worth noting that this revision is a much less dramatic one, than that proposed by Romero (2015) and Baumgartner and Gebharter (2016) in order to accommodate fat-handed interventions.

[31] It is not clear to us that replacing the claim that '*I* causes *X*' with the claim that '*I* counterfactually depends on *X*' is at all necessary for *I*-liberalism. As we mentioned in footnote 12, we are perfectly happy with the relationship between *I* and *X* being a causal one. In the case of Socrates' death and Xanthippe's widowhood, any intervention upon Socrates' death *will* be a causal process. What is important for *I*-liberalism, is that such an intervention establishes a subsequent *noncausal* explanatory dependence between the death of Socrates and Xanthippe's becoming a widow. In this sense, *I*-liberalism appears to require even less modification to Woodward's original manipulationist framework than QCM.

B are correlated, it allows for the distinction to be reliably drawn between scenarios where: *A* causes *B*; *B* causes *A*:or *A* and *B* are both caused by *C*.[32]

While analogous spurious correlations *do* arise with respect to noncausal dependence, it is clear to us that instances of constitutive explanation are not among them. If Khalifa et al. were correct in their claim that an interventionist account of noncausal explanation will misdiagnose instances of constitutive explanation as spurious correlation, then the *I*-liberalist would not to be able to draw a distinction between these cases. As we shall now demonstrate, however, our analysis is perfectly capable of distinguishing between spurious (noncausal) correlations and constitutive explanations.

As an example of a genuinely spurious noncausal correlation, take the relationship between Xanthippe's widowhood, and the existence of Socrates' Singleton. It seems that there is a necessary inverse correlation here, there are no possible worlds in which Xanthippe is a widow and Singleton Socrates exists. Similarly, there are no possible worlds in which Singleton Socrates does not exist and Xanthippe is not a widow. Yet, we would not want to say that Xanthippe's widowhood *depends* upon, or is *explained by*, the nonexistence of Singleton Socrates (or vice versa).

Indeed, the correct story here seems rather obvious: the existence of Socrates' Singleton and Xanthippe's widowhood are *both* determined by a single common factor: the existence of Socrates. Where Socrates exists, it is necessarily the case that Single Socrates exists and that Xanthippe is not a widow (and vice versa). And, while Socrates' existence explains both Singleton Socrates' existence and Xanthippe's not being widowed, neither of the latter facts explain each other. This case, then, looks like an instance of genuine noncausal common dependence of the type which Khalifa et al (2020) intend to highlight.

That the correlation between the existence of Singleton Socrates and Xanthippe's widowhood is spurious can be quite happily cashed out in terms of interventions. In line with *III*, try as one might, it is simply not possible to intervene upon either Xanthippe's widowhood, or the existence of Singleton Socrates, in order to manipulate the other; any such intervention *must* go through Socrates' existence. This tells us that it is Socrates' existence which is doing all of the determining here, and hence,

[32] Ylikoski's (2013) argument is specifically aimed at Craver's (2007a; 2007b) mutual manipulability account of constitutive explanations. In this context, the argument that constitutive explanation underdetermines explanatory relations in virtue of being unable to distinguish such cases from spurious correlations arising from common causes is, as far as we are concerned, perfectly sound. As we highlighted in footnote 16, the type of "top-down" interventions which result from Craver's mutual manipulability, *can* be interpreted in such a way that they do appear to be the result of common causes. However, Khalifa et al (2020) take this argument further, seemingly arguing that even the sort of "bottom up" intervention which we have taken to be unproblematic gives rise to this same explanatory confusion. Woodward (2018) has also suggested (although in vaguer terms) that in abandoning interventions, explanatory monists face something like this problem. However, as we shall see below, by adopting *I*-liberalism, this result can be avoided.

all of the explanatory work.[33] For the sake of simplicity, let's consider another (more basic) example of constitutive explanation:

Variables:
C: Whether the diamond's constituent carbon atoms are thus-and-so arranged.
E: Whether the diamond is hard.
Structural equations:
E → C
Assignment:
C=1; E=1

We can now see that this case of constitutive explanation is quite different to the spurious correlation highlighted above, and that this difference can be drawn out in terms of possible interventions. If Khalifa et al. (2020) are correct, and the arrangement of the diamond's constituent carbon atoms, and its hardness, are wrongly characterised by *I*-liberalism as being jointly depend upon some common factor, it ought to be the case that there are no possible interventions upon either variable that goes through the other.

This line of thought mirrors Kim's, in considering Socrates' death and Xanthippe's widowhood as common effects of Socrates' having ingested hemlock. In this case, Kim suggests that the problem with this idea is that the only route from hemlock to widowhood seems to go through death. In the case of Xanthippe's widowhood and the existence of Singleton Socrates, on the other hand, any possible intervention which attempts to manipulate the former using the latter, is *mediated* by the existence of Socrates. If Khalifa et al. (2020) are correct, then cases of constitutive explanation ought to look more like the relationship between Xanthippe's widowhood and the existence of Singleton Socrates, that the relationship between Xanthippe's widowhood and Socrates' death. However, this is not what we find.

Just as in the case of Socrates' death and Xanthippe's widowhood, in line with *III*, any possible intervention upon the hardness of a cut diamond *must* go through its constituent carbon atoms.[34] It is simply impossible to alter the hardness of a diamond without altering the arrangement of its constituent carbon atoms. This is just what we saw in relation to the *Mimosa's* in Sect. 4, any intervention upon the plant's nastic movement must go through its pulvini cells. Were these cases of constitutive explanation the result of a common explanatory source, there would be some third variable doing the actual explanatory work. But this is not so. Thus, an account of noncausal explanation which makes use of interventions *can,* in fact, draw an

[33] In 'Supervenience as a Philosophical Concept' (1990/1993) Kim uses a very similar argument to defend the thesis that supervenience, like correlation, is not an explanatory relation. Just as correlation is insufficient to establish an explanatory causal dependence relationship between two variables, Kim argues that supervenience is insufficient to establish an explanatory *noncausal* dependence relationship between two variables. The analogy between supervenience and correlation could well be an illuminating one, especially given a recent resurgence of interest in the explanatory status of supervenience (e.g. Kovacs 2019). Unfortunately, however, we do not have the space to explore this connection here.

[34] Such an intervention could be performed by subjecting our diamond to around 10 million times ordinary atmospheric pressure, for example (see e.g. Knudson et al., 2008).

illuminating distinction between constitutive explanation and spuriously correlated variables resulting from a common explanatory dependence relation.

7 Concluding Remarks on Taxonomy

In this paper, we have attempted to mount the first sustained defence of *I*-liberalism, against its more popular rival, *I*-puritanism. In Sect. 2, we introduced Kim's argument against causal imperialism, the claim that all explanation is essentially causal in nature. Causal imperialism is false, according to Kim, because there are clear cases of asymmetric explanatory counterfactual conditionals which are not the result of causal relationships. Crucially, what these causal and noncausal counterfactuals share, is a close connection to the notion of "bringing about": where *A* depends upon *B*, we can bring about *A* by first bringing about *B*.

In Sect. 3, we introduced Woodward's interventionist analysis of causal explanation and argued that Kim's intuition regarding the asymmetry of noncausal explanation and the "bringing about" relation can be neatly characterized in terms of interventionist counterfactuals and structural equation models. What this shows, we argued, is that the notion of a possible intervention does not line up with the distinction between causal and noncausal explanation.

Having characterized the core claim of *I*-liberalism, that possible interventions can characterize certain noncausal explanations, in Sect. 4, we moved on to apply our methodology to an archetypal instance of such explanation: a constitutive mechanism. We argued that *I*-liberalism avoids two central issues facing Craver's own mutual manipulability analysis: the need for impossible interventions; and the 'embarrassing' explanatory symmetry which results. We also observed that debate surrounding constitutive explanation has been a key breeding ground of *I*-puritan sentiments, with the likes of Leuridan (2012) having dismissed Craver's account *tout court* because it attempts to characterise noncausal explanation in terms of interventions.

As we also highlighted, however, this is not the only area where we find such dismissive attitudes. More recently still, debate surrounding the viability of explanatory monism has invoked similar responses. In Sect. 5, we noted that the predominating *I*-puritan stance has led to difficulty in characterising the obvious asymmetry of noncausal explanation. Despite this difficulty, we showed that motivation for *I*-puritanism among those involved in characterizing *both* monistic and constitutive explanation typically appeals to Woodward's (2003) own defence of this position. In response, we argued that, even though Woodward (2003) supports *I*-puritanism, nothing in his account mandates this interpretation. Indeed, Woodward's (2018) most recent foray into the topic of noncausal explanation appears to roll back this stance and take a significant step towards *I*-liberalism.

In Sect. 6, we discussed an argument, recently put forward by Khalifa et al (2020), which suggests that an interventionist account of noncausal explanation would be unable to distinguish between genuinely explanatory relationships and spurious correlations resulting from common causes. On the contrary, we argued the *I*-liberalist is perfectly capable of drawing a distinction between genuine explanatory noncausal dependence relations on the one hand, and unexplanatory spurious correlations arising from a common dependence relation, on the other. In what remains of this final section, we wish to provide something of a taxonomy of the various positions which have been discussed in this paper and highlight exactly what our own position commits us to.

While the idea that all explanation is causal explanation has largely fallen out of favour, it is worth noting that causal imperialism is fully compatible with *I*-puritanism as we have described it. One might think (against the current consensus, of course) that all explanation is causal explanation *and* that all such explanation can be characterized in interventionist terms. Let's call this position "strong causal puritanism". Causal imperialists need not necessarily think that all causal explanation is characterizable in interventionist terms, just that wherever one *can* intervene upon *X* with respect to *Y* in such a way that changes the value of *Y*, *X* causes *Y*. This leaves open the possibility that, while all explanations are causal explanations, some such explanations defy interventionist analysis. Let's call this position "weak causal puritanism".

The former of these positions, strong causal puritanism, implies explanatory monism. If all explanation is causal explanation, and all causal explanation is characterizable in terms of interventions, then we have a single unifying (interventionist) account of explanation. Conversely, the latter position, weak causal puritanism, implies explanatory pluralism. Even if all explanation is causal in nature, if some such explanations are not characterizable in interventionist terms, then explanatory monism must be false. As far as we are aware, however, no one has committed themselves to either of these positions in the recent literature.

Those of an *I*-puritan persuasion are likely to reject causal imperialism on the grounds that noncausal explanations are possible. Monist *I*-puritans will argue that although causal and noncausal explanation can be captured using a single unifying thesis, said thesis will not reference interventions.[35] Pluralist *I*-puritans, on the other hand, will accept that noncausal explanation is possible, but reject the idea that both types of explanation can be captured in a single unifying thesis, as Lange puts it: 'the order of explanatory priority is fixed by different considerations in different non-causal explanations' (2019: 24).

I-liberalism is obviously incompatible with causal imperialism on two fronts. First, *I*-liberalism presupposes that noncausal explanations are possible, and second, it argues (*contra* *I*-puritanism) that (at least) some noncausal explanations can be characterized in terms of the possibility of intervening upon the explanans variable. Given that causal imperialism consists in the denial the first of these claims, the second is clearly a nonstarter. Although, as with causal imperialism and *I*-puritanism, *I*-liberalism is noncommittal with respect to explanatory monism vs explanatory pluralism.

[35] As we have seen such theories typically take counterfactuals to be the central unifying feature of causal and noncausal explanation, although Khalifa et al (2018) have recently argued for a monistic *inferential* account of explanation.

A monist *I*-liberalist methodology would imply, not only that possible interventions can characterize *certain* instances of noncausal explanation, but that possible interventions are capable of characterizing *all* instances of explanation. The most obvious reason for rejecting this hard-line *I*-liberalism, are scenarios of the sort highlighted by Woodward (2003) as a reason for rejecting *I*-liberalism all together. These involve counterfactual conditionals whose antecedents hold with necessity. Intervening upon a variable which holds its value of necessity (like the dimensionality of space–time) will, of course, be *at least* nomologically (although often also logically and/or metaphysically) impossible.

Such counterpossible counterfactuals have received a great deal of attention within the philosophy of science literature. Indeed, one of the most interesting recent developments has been the idea that we can cash out the explanatory potential of counter*possibles* in terms of interventions, even though such interventions are impossible. For example, Baron et al (2017), Reutlinger et al (2020) and Baron et al (2020) have argued that mathematical explanations can be understood in terms of interventions which are, strictly speaking, (metaphysically) impossible to perform.[36]

It is important to note, however, that this position is compatible with *I*-puritanism. Indeed, a pluralist *I*-puritan might well accept that certain noncausal explanations are characterizable in terms of impossible interventions, but nonetheless maintain their central thesis, that wherever it is *possible* to intervene upon X in such a way that changes the value of Y, then X causes Y. In this sense, there would remain hope for the *I*-puritan that a neat dividing line can be drawn between causal and noncausal explanation in terms of *possible/impossible* interventions. It is this idea which we have sought to undermine.[37]

In this paper, we have attempted to mount a thorough defence of only a *weak* form of *pluralist I*-liberalism, which suggests that possible interventions do not highlight causal explanatory relations *alone*. In other words, our pluralist *I*-liberalist thesis suggest merely that *some* noncausal explanations, of the types highlighted herein, are characterizable in terms of interventions which are possible to perform. This weaker *I*-liberalist thesis is, of course, sufficient to prove the falsity of *I*-puritanism. If even a single noncausal explanation permits of a manipulationist analysis, then it simply cannot be the case that possible interventions serve only to characterize causal relations.[38]

[36] Schaffer (2016, 2017) and Wilson (2016, 2018) have argued that metaphysical explanations similarly require the analysis of impossible interventions, and Baron & Colyvan (2021) and Baron (*forthcoming*) have argued that certain ontological and logical explanations (respectively) are in the same boat. In all these cases, adopting such a stance requires a commitment to the non-triviality of counterpossibles and the abandoning of the traditional semantics for counterfactuals (see e.g., Stalnaker 1968; Lewis 1973). However, as Schaffer (2016) highlights, there are already good reasons for thinking that counterpossible scenarios require non-trivial evaluation (see, e.g. Restall 1997; Goodman 2004; Priest 2005; and Jago 2015).

[37] Our thanks go to an anonymous reviewer at this journal for pressing this important point.

[38] Thanks to Alastair Wilson for invaluable guidance on previous incarnations of this paper; also, to two anonymous reviewers at this journal. This paper forms part of the project "A Framework for Metaphysical Explanation in Physics" (FraMEPhys), hosted by the University of Birmingham and funded through the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 757295).

References

Audi, P. (2012). A clarification and defences of the notion of grounding. In F. Correia & B. Schneider (Eds.), *Metaphysical grounding: understanding the structure of reality.* (pp. 101–121). Cambridge: Cambridge University Press.

Baron, S. (*forthcoming*). Counterfactuals of ontological dependence. *Journal of the American Philosophical Association*, 1–22.

Baron, S., & Colyvan, M. (2021). Explanation impossible. *Philosophical Studies, 178*(2), 559–576.

Baron, S., Colyvan, M., & Ripley, D. (2017). How mathematics can make a difference. *Philosopher's Imprint, 17*(3), 1–19.

Baron, S., Colyvan, M., & Ripley, D. (2020). A counterfactual approach to explanation in mathematics. *Philosophia Mathematica, 28*(1), 1–34.

Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science, 84*(2), 214–233.

Baumgartner, M., & Gebharter, A. (2016). Constitutive relevance, mutual manipulability, and fat-handedness. *British Journal for the Philosophy of Science, 67*(3), 731–756.

Bechtel, W. (2008). Mechanisms in cognitive psychology: What are the operations. *Philosophy of Science, 75*(5), 983–994.

Bechtel, W., & Abrahamson, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science of Biological and Biomedical Sciences, 36*(2), 421–441.

Bokulich, A. (2011). How scientific models can explain. *Synthese, 180*(1), 33–45.

Bokulich, A. (2018). Searching for noncausal explanations in a sea of causes. In A. Reutlinger & J. Saatsi (Eds.), *Explanation Beyond Causation: Philosophical Perspectives on Non-Causal Explanations.* (pp. 141–161). Oxford: Oxford University Press.

Briggs, R. (2012). Interventionist counterfactuals. *Philosophical Studies, 160*(1), 139–166.

Cartwright, N. (1979). Causal laws and effective strategies. *Noûs, 13*(4), 419–437.

Collingwood, R. G. (1940). *An essay on metaphysics*. Oxford: Oxford University Press.

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

Craver, C. (2007b). Constitutive explanatory relevance. *Journal of Philosophical Research, 32*, 3–20.

Craver, C. (2014). The ontic account of scientific explanation. In M. Kaiser, O. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the special sciences: the case of biology and history.* (pp. 27–52). Berlin: Springer.

Craver, C., & Darden, L. (2001). Discovering mechanisms in neurobiology: The case of spatial memory. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and Method in Neuroscience.* (pp. 112–137). Pittsburgh: University of Pittsburgh Press.

Craver, C., & Darden, L. (2002). Strategies in the interfiled discovery of the mechanism of protein synthesis. *Studies in History and Philosophy of Science Part C: Studies in the History and Philosophy of Biological and Biomedical Sciences, 33*(1), 1–28.

Craver, F., & Darden, L. (2013). *Search of Mechanisms. Discoveries across the Life Sciences*. Chicago: University of Chicago Press.

Darden, L. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), 342–352.

Dasgupta, S. (2017). Constitutive explanation. *Philosophical. Issues, 27*(1), 74–97.

Esau, K. (1965). *Plant anatomy*. John Wiley.

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: Oxford University Press.

Gasking, D. (1955). Causation and recipes. *Mind, 64*(256), 479–487.

Glennan, S. (2010). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research, 81*(2), 362–381.

Glennan, S. (2017). *The new mechanical philosophy*. Oxford University Press.

Goodman, J. (2004). An extended Lewis/Stalnaker semantics and the new problem of counterpossibles. *Philosophical Papers, 33*(1), 35–66.

Harinen, T. (2014). Mutual manipulability and causal inbetweenness. *Synthese, 195*(1), 35–54.

Hausman, D. M. (1982). Causal and explanatory asymmetry. *PSA: Proceedings of the biennial meeting of the philosophy of science association* (vol. 1982, pp. 43–54).

Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy, 98*(6), 273–299.

Illari, P., & Williamson, J. (2012). What is a Mechanism? Thinking about mechanisms across sciences. *European Journal for Philosophy of Science, 2*(1), 119–135.

Jaffe, J., Leopold, A., & Staples, R. (2002). Thigmo responses in plants and fungi. *American Journal of Botany, 89*, 375–382.

Jago, M. (2015). Hyperintensional propositions. *Synthese, 192*(3), 585–601.

Jansson, L., & Saatsi, J. (2019). Explanatory abstractness. *British Journal for the Philosophy of Science, 70*(3), 817–844.

Khalifa, K., Doble, G., & Millson, J. (2020). Counterfactuals and explanatory pluralism. *British Journal for the Philosophy of Science, 71*(4), 1439–1460.

Khalifa, K., Millson, J., & Risjord, M. (2018). Inference, explanation, and asymmetry. *Synthese*. https://doi.org/10.1007/s11229-018-1791-y

Kim, J. (1974/1993). Noncausal connections. In J. Kim (Ed.), *Supervenience and mind*. Cambridge: Cambridge University Press.

Kim, J. (1973). Causes and counterfactuals. *Journal of Philosophy, 70*(17), 570–572.

Kim, J. (1990). Supervenience as a philosophical concept. *Metaphilosophy, 21*(1–2), 1–27.

Kim, J. (1994). Explanatory knowledge and metaphysical dependence. *Philosophical Issues, 5*, 51–69.

Knudson, M., Desjarlais, D., & Dolan, D. (2008). Shock-wave exploration of the high-pressure phases of carbon. *Science, 322*(5909), 1822–1825.

Kovacs, D. (2017). Grounding and the argument from explanatoriness. *Philosophical Studies, 174*(12), 2927–2952.

Kovacs, D. (2019). The myth of the myth of supervenience. *Philosophical Studies, 176*(8), 1967–1989.

Krickel, B. (2018). Saving the mutual manipulability account of constitutive relevance. *Studies in History and Philosophy of Science Part A, 68*, 58–67.

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

Leuridan, B. (2012). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *British Journal for the Philosophy of Science, 63*(2), 399–427.

Lewis, D. (1973). *Counterfactuals*. Blackwell.

Lewis, D. (1986). *On the plurality of words*. Wiley-Blackwell.

Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.

Meek, C., & Glymour, C. (1994). Conditioning and interventing. *British Journal for the Philosophy of Science, 45*(4), 1001–1021.

Menzies, P., & Price, H. (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science, 44*(2), 187–203.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge University Press.

Pearl, J. (2009). Causal inference in statistics: An overview. *Statistical Surveys, 3*, 96–146.

Peirce, C. S. (1900). In E. Moore (Ed.), *The writings of Charles S. Peirce: A chronological edition*. Bloomington: Indiana University Press.

Peirce, C. S. (1931–58). In C. Hartshorne & P. Weiss (Eds.), *Collected papers of charles sanders peirce*. (vols. i-vi), A. Burks (vols. vii & viii). Cambridge MA: Belknap Press.

Pexton, M. (2014). How dimensional analysis can explain. *Synthese, 191*(10), 2333–2351.

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.

Railton, P. (1981). Probability, explanation, and information. *Synthese, 48*(2), 233–256.

Restall, G. (1997). Ways things can't be. *Notre Dame Journal of Formal Logic, 38*(4), 583–596.

Reutlinger, A. (2016). Is there a monist theory of causal and non-causal explanations? The counterfactual theory of scientific explanation. *Philosophy of Science, 83*(5), 733–745.

Reutlinger, A. (2017). Does the counterfactual theory of explanation apply to non-causal explanation in metaphysics? *European Journal for Philosophy of Science, 7*, 1–18.

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: Oxford University Press.

Reutlinger, A., Colyvan, M., & Krzyżanowska, K. (2020). The prospects for a monist theory of non-causal explanation in science and mathematics. *Erkenntnis*. https://doi.org/10.1007/s10670-020-00273-w

Rice, C. (2015). Moving beyond causes: Optimality models and scientific explanation. *Noûs, 49*(3), 589–615.

Romer, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese, 192*(11), 3731–3755.

Roski, S. (2020). Metaphysical explanations and the counterfactual theory of explanation. *Philosophical Studies*. https://doi.org/10.1007/s11098-020-01518-8

Ruben, D. (1990). *Explaining explanation*. Routledge.

Saatsi, J. (2018). On explanations from geometry of motion. *British Journal for the Philosophy of Science, 69*(1), 253–273.

Saatsi, J., & Pexton, M. (2013). Reassessing Woodward's account of explanation: Regularities, counterfactuals, and noncausal explanations. *Philosophy of Science, 80*(5), 613–623.

Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies, 173*(1), 49–100.

Schindler, S. (2013). Mechanistic explanation: asymmetry lost. In K. Dieks (Ed.), *Recents Progress in Philosophy of Science: Perspectives and Foundational Problems.* Berlin: Springer.

Skow, B. (2014). Are there non-causal explanations (of Particular Events)? *British Journal for the Philosophy of Science, 63*(3), 445–467.

Stalnaker, R. (1968). A theory of conditionals. In N. Rescher (Ed.), *Studies in Logical Theory (American Philosophical Quarterly Monographs 2).* . (pp. 98–112). Oxford: Blackwell.

Strevens, M. (2008). *Depth: An account of scientific explanation*. Harvard University Press.

Taylor, E. (2018). Against explanatory realism. *Philosophical Studies, 175*(1), 197–219.

Thompson, N. (2018). Irrealism about grounding. *Royal Institute of Philosophy Supplement, 82*, 23–44.

von Wright, G. H. (1975). *Causality and Determinism*. New York: Columbia University Press.

Wilson, A. (2021). Counterpossible reasoning in physics. *Philosophy of Science*, *88*(5).

Wilson, A. (2018). Metaphysical causation. *Noûs, 50*(4), 1–29.

Wilson, A. (2020). Classifying dependencies. In D. Glick, G. Darby, & A. Marmodoro (Eds.), *The foundation of reality: Fundamentality, space and time.* (pp.46–59). Oxford: Oxford University Press.

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

Woodward, J. (2015). Interventionism and causal exclusion. *Philosophy and Phenomenological Research, 91*(2), 303–347.

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.117–141). Oxford: Oxford University Press.

Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis, 78*(2), 277–297.