Philosophical Studies (2024) 181:2091–2112
https://doi.org/10.1007/s11098-022-01886-3

Plumbing metaphysical explanatory depth

Nicholas Emmerson[1]

Accepted: 16 September 2022 / Published online: 28 October 2022
© The Author(s) 2023

Abstract
Recent years have seen increasing interest in interventionist analyses of metaphysical explanation. One area where interventionism traditionally shines, is in providing an account of explanatory depth; the sense in which explanation comes in degrees. However, the literature on metaphysical explanation has left the notion depth almost entirely unexplored. In this paper I shall attempt to rectify this oversight by motivating an interventionist analysis of metaphysical explanatory depth (MED), in terms of the range of interventions under which a metaphysically explanatory generalization remains invariant. After elucidating the notion through a toy-example, I demonstrate the important work which MED can perform in characterizing debate within contemporary metaphysics. Focusing upon rival approaches to explaining the identity and distinctness of concrete objects, I argue that the progress achieved in this debate can be characterized in terms of increasing explanatory depth. Having made an initial case for the utility of MED, I then turn this analysis to the metaphysics of explanation itself. By adopting an interventionist framework with respect to MED, I will show that we can assess the depth of competing theories of explanation. This application has two interesting results: first, it suggests that an interventionist analysis of explanation provides deeper explanations of the connection between explanans and explanandum than rival accounts; and second, it suggests that explanations provided by interventionism become deeper still, if one accepts that this methodology ranges over metaphysical, as well as causal, instances.

Keywords Metaphysical explanation · Explanatory depth · Interventionism · Counterfactuals · Invariance

✉ Nicholas Emmerson
nje987@bham.ac.uk

[1] Department of Philosophy, University of Birmingham, ERI Building, Edgbaston, Birmingham, West Midlands B15 2TT, UK

1 Introduction

Recent years have seen growing interest in the prospect of modifying interventionist analyses of causal explanation in order to characterize *non*causal explanations within metaphysics.[1] One area where interventionism traditionally shines, is in accounting for the sense in which explanation comes in degrees, or as Hitchcock and Woodward (2003b) put it, the *depth* of an explanation. However, an analogous notion of depth is entirely absent from parallel debate concerning metaphysical explanation. In this paper, I seek to rectify this oversight. I argue that *metaphysical* explanatory depth ought to be analogously understood in terms of the range of interventions under which an explanatory generalization remains invariant.

I shall proceed as follows. In the next section I provide a detailed account of the interventionist analysis of explanatory depth, and the benefits of this approach over rival inferential analyses, which define depth in terms of *scope*. In Sect. 3, I apply these contrasting notions of depth to a toy-example of metaphysical explanation and show that the benefits of an interventionist analysis carry across to the explanations within metaphysics.

With the preliminaries out of the way, I put this account to work, applying the interventionist understanding of explanatory depth to two case-studies of metaphysical explanation *in the wild* (as it were). As these case-studies show, an interventionist account of metaphysical explanatory depth provides us with novel tools with which we can characterize live debate within contemporary metaphysics.

In Sect. 4, I focus upon a recent approach to explaining the identity and distinctness of concrete objects, the *quantitative properties proposal*, put forward by Erica Shumener (2020). I argue that Shumener's thesis can be seen as progressive with respect to prior proposals in terms of both *qualitative properties* (Black, 1952; Rocca, 2005) and *weak discernibility* (Saunders, 2006), precisely because it provides greater metaphysical explanatory depth.

In Sect. 5, I argue that this same methodology can be applied to debate surrounding the nature of explanation *itself*. We can, I suggest, think of analyses of explanation as providing metaphysical explanations for the connection between explanans and explanandum. Here, competing accounts of explanation provide contrasting meta-explanatory generalizations of the following form: "*for any x and y, if …, then x explains y*". Through the adoption of an interventionist framework with respect to metaphysical explanation, I will show that we can assess the depth of rival accounts of explanation.

This novel application of the notion of explanatory depth appears to have two interesting results. First, it accurately characterizes the interventionist analysis of explanations as providing greater depth than both inferential (Hempel, 1965; Kitcher, 1981) and conserved quantity (Salmon, 1984, 1994; Dowe, 1992, 2000) accounts. And second, it suggests that interventionism can be shown to provide still deeper explanations of the connection between explanans and explanandum, if one

[1] See e.g., Schaffer (2016, 2017), Wilson (2016, 2018), Reutlinger (2017), Miller and Norton (2022a, b).

additionally accepts that this analysis rangers over metaphysical, as well as causal, instances.

In the final section I discuss three methodological issues which arise as a result of my interventionist analysis of metaphysical explanatory depth. The first concerns how we are to make sense of the distinction between same- and other-object counterfactuals in the context of identity and distinctness; the second, raises the issue of how we are to understand the role of interventions when it comes to metaphysical explanation more generally; and the third asks whether shallow metaphysical explanations are actually explanations at all. In each case, I argue that the interventionist can provide a satisfying response.

2 Explanatory depth: scope vs invariance

For much of the twentieth century, debate surrounding the nature of explanation was dominated by broadly inferential analyses. On Carl Hempel's deductive-nomological model, for example, explanation is centrally concerned with *expectation*. An explanation, on this account, involves a set of auxiliary statements $C_1 \ldots C_n$, asserting the occurrence of events and a law, L, from which we can logically deduce a statement asserting the occurrence of the event to be explained E. Such arguments are intended to show that 'given the particular circumstances and the law in question, the occurrence of the phenomenon *was to be expected*' (Hempel, 1965: 337).

However, Philip Kitcher (1981, 1989) argues that behind Hempel's "official view" of explanation as nomic expectation, there lies an "unofficial" view of explanation as *unification*.[2] As Kitcher interprets this unofficial view, laws are explanatory in virtue of 'showing us how to derive descriptions of many phenomena, using the same patterns of derivation again and again' (1989: 432).[3] Kitcher's key insight here, is that in order for a generalization to be explanatory (as opposed to a merely accidental) it ought to apply to a *range of different cases*.[4]

This idea provides us with a natural way of understanding explanatory depth: the wider the range of cases to which an explanation applies, the deeper the explanation. Where DN and unificationist accounts diverge from an interventionist analysis, however, is with respect to *which* cases are taken to be salient in determining depth. On both the DN and unificationist accounts, this range is understood in terms of *scope*: the set of objects or systems that fall under the antecedent of a given law.

Such laws, according to Hempel (1965) and Kitcher (1981), take the form of universally quantified conditional claims like "All *A*s are *B*s". As Hitchcock and Woodward (2003a) highlight, such conditionals support "other-object" counterfactuals of

[2] Also see Friedman (1974).

[3] As Hempel himself suggests, explanation is achieved 'by a systematic unification, by exhibiting the phenomenon as manifestations of common, underlying structures and processes' (1966: 83).

[4] Brad Weslake highlights that this approach appears to have been, at least tacitly, endorsed by Hempel himself: 'the most natural way to incorporate an account of explanatory depth into the DN account is itself suggested by Hempel (1959: 302–303), who mentions in passing the predictive possibilities afforded by laws in situations other than the one under consideration' (2010: 276).

the following form: 'if some object *o** that is different from *o* and does not possess property *A* were to be an *A*, then it would be a *B*' (2003a: 19).[5] In this sense, the wider the range of *other objects* to which an explanatory generalization applies, the more unifying it will be, and thus the deeper the explanation it will provide.

There is, however, a fundamental problem facing a scope-based account of explanatory depth and the traditional understanding of laws the 'universal premisses that occur in explanatory patterns' (Kitcher, 1989: 447). On this interpretation of laws, it becomes difficult to see quite how one generalization could provide for deeper explanations than any other. As Hitchcock and Woodward explain, 'a true generalization is either a universal law, in which case it can facilitate explanations, or it is accidental, in which case it cannot-there are no other options' (2003b: 183).[6]

Fortunately, the interventionist analysis avoids this problem because it does not require generalizations to be universal or exceptionless in order to be explanatory. Rather, according to Hitchcock and Woodward, 'it is only if a generalization is invariant under testing interventions that it conveys information about how one variable depends on another' (2003a: 19). For a generalization to be invariant under testing interventions, it must 'describe a relationship which holds for certain *hypothetical* values of *X* and *Y possessed by the very object o*… where the value of *X* is changed by an intervention' (Hitchcock & Woodward, 2003a: 20).

By interpreting the explanatory character of generalizations in these terms, we see that '[a]mong those generalizations that are invariant, some will be more invariant than others, and they will correspondingly provide deeper explanations' (Hitchcock & Woodward, 2003b: 183–184). As such, the interventionist account of explanatory depth is not concerned with the range of *other objects* for which a given generalization holds, but rather, the range of changes to the *actual object being explained* under which a given generalization remains invariant. In other words, the greater the range of *same-object* counterfactual scenarios under which a generalization holds, the deeper the explanation.

In the introduction, I noted a recent trend which seeks to characterize metaphysical explanation, in interventionist terms, as analogous to causal explanation. Despite what Boris Kment calls the 'far-reaching and structural analogy' (2014: 5) between metaphysical and causal explanation, the role of explanatory depth within metaphysics remains, as yet, unexplored. One initial reason for thinking that this area deserves further investigation is that recent interventionist theories of metaphysical explanation already provide us with all of the methodological tools necessary to characterise a notion of explanatory depth operative in the metaphysical domain.

Jonathan Schaffer, for example, has recently echoed Hitchcock and Woodward (2003a, 2003b) in arguing that the generalizations which govern metaphysical explanation needn't be fundamental or exceptionless. According to Schaffer, to qualify as explanatory a "metaphysical law" must merely support an appropriate pattern

[5] Also see Woodward (2003: 279–288).

[6] Hitchcock and Woodward are not the first to mount this objection. Jaegwon Kim, for example, similarly argues that '[u]nderstanding and explanatoriness are matters of degree… The DN model makes explanation an all-or-nothing affair' (1994: 59).

of counterfactuals, as 'it is through counterfactual-supporting generalizations that one can calculate the impact of potential interventions' (2017: 306). So, just as Hitchcock and Woodward argue with respect to causal explanation, Schaffer (2017) argues that *metaphysical* explanation also requires generalizations which will remain invariant under testing interventions.

From here, it is a small step to arrive at an analogous notion of *metaphysical* explanatory depth. For one metaphysical explanation to be deeper than another, is for the corresponding generalization to be invariant under a wider range of same-object counterfactuals. In the next section, I shall apply this interventionist analysis of explanatory depth to a toy-example of metaphysical explanation, and demonstrate that the benefits of this analysis, over a scope-based account, appear to carry across into the metaphysical domain.

3 A toy-example

Consider, as a preliminary example, the fact that Jeff Bezos is a billionaire.[7] One might well ask why this is the case, what explains Bezos's being billionaire? A natural explanation of this fact could simply highlight his net worth: \$182 B. Another might cite the fact that his net worth is ≥ \$1 B.

(a) For any person x, if x has a net worth of \$182 B, then x is a billionaire.
(b) For any person x, if x has a net worth of ≥ \$1 B, then x is a billionaire.

At least on the face of it both (a) and (b) appear to qualify as candidate generalizations figuring in an explanation of Bezos's being a billionaire. However, it is my contention that the second of these generalizations provides the deeper explanation, and that the interventionist analysis of explanatory depth is better able to make sense of this claim, than a scope-based analysis.

As we saw in the previous section, on a scope-based account, explanatory depth is determined by the range of *other objects* to which an explanation applies. Here,

[7] One might query whether this example is actually an instance of *metaphysical* explanation at all. While some, like Dasgupta (2017) take metaphysical explanation to be synonymous with *constitutive* explanation (a category into which this example appears to fall), it is not clear that metaphysics has a monopoly on such explanations. Within the literature on noncausal explanation more generally, constitutive explanations are typically taken to be a form of *mechanistic* explanation and are distinguished from *etiological* mechanistic explanation. In the latter case, some mechanism explains a phenomenon for which is *causally* responsible, whereas in the former, a phenomenon is explained by the underlying mechanism which *constitutes* it (e.g., Machamer et al,. 2000; Darden & Craver 2002; Craver 2007a, b; Glennan 2010, 2017). For what it's worth, I take *all* explanation to be an instance of *either* metaphysical or causal explanation although, as Wilson (2020) highlights, even this distinction can be difficult to parse. While the above explanation is clearly not causal in nature, for those unwilling to accept it as an instance of metaphysical explanation, I would hope that it serves a useful purpose in helping to draw out the difference between same- and other-object counterfactuals in relation to explanatory depth. In Sects. 4 and 5, I will consider two more complex examples of explanation which are less controversially metaphysical in character.

in assessing the depth of the above explanations, we thus consider the range of other objects under which (a) and (b) hold. The greater the range of objects which can be subsumed under each generalization, the deeper the explanation. This gives rise to a problem, however. It seems that in terms of scope, these candidate generalizations are equally explanatory.

Take Boris Johnson, for example. As far as it is possible to tell, Johnson is not a billionaire; his net worth appears to be somewhere between \$2 M and \$4 M, depending on who you ask. Despite this, *were* Johnson's net worth \$182 B, then he *would* be a billionaire. In this case, the generalization specified by (a) would still hold. In fact, (a) will continue to hold for any counterfactual scenario in which we replace Bezos with another person. *Anyone* would be a billionaire if their net worth were \$182 B. The problem is that the same is true of (b). Were Johnson's, net worth $\geq$ \$1 B, then he would be a billionaire. Were *my* net worth $\geq$ \$1 B then I would be a billionaire. There is no "other object" with which we can replace Bezos which will allows us to draw a distinction between (a) and (b) in terms of scope.[8]

By adopting an interventionist analysis of explanatory depth, however, we can provide a satisfying characterization of why (b) provides a deeper explanation of Bezos's being a billionaire than (a). On this account, rather than observing the range of other objects under which (a) and (b) hold, we must assess changes to the *actual* object (or system) in question. What we are chiefly looking for, in justifying the intuition that (b) provides a deeper explanation than (a), are hypothetical counterfactual scenarios in which (a) is violated but (b) remains invariant. Such cases are not difficult to come up with.

Imagine, for example, that all Amazon warehouse staff unionize and, with investors concerned about the impact of good working conditions on profitability, Amazon's share price falls, leaving Bezos with a net worth of \$150 B. In this scenario, it appears that (a) no longer explains why Bezos is a billionaire; it tells us nothing about counterfactual scenarios in which a person's net worth is anything other than \$182 B. So, while Bezos is obviously still a billionaire in this scenario, (a) cannot be used to explain this fact. On the other hand, (b) will remain invariant under counterfactual scenarios which see interventions upon Bezos's net worth, so long as his net worth remains $\geq$ \$1 B. As such (b) holds under a wider range of counterfactual scenarios than (a) and can thus be considered a deeper explanation of Bezos's being a billionaire.

[8] One might worry that the generalizations which are specified in (a) and (b) are not *universal*. Since it was specified above that inferential accounts takes laws to be universal generalizations, it could be argued that this example is unfairly tipped in the interventionist's favour. I would point out, however, that these generalizations can be easily rephrased in order to apply universally: "(a*) *For any x, if x has a net worth of \$182 B, then x is a billionaire"*. However, this formulation unnecessarily complicates matters, opening up the possibility of counterfactual scenarios which become impossible to interpret with any clarity. For example, what if *x* is an inanimate object, or an animal? Would a rock, a record player or a rhinoceros be a billionaire if they had a net worth of \$182 B? Hitchcock and Woodward (2003a) argue that it is mysterious what such counterfactual scenarios are even supposed to *mean*. More importantly, such scenarios still do not allow us to draw a distinction, in terms of depth, between (a) and (b). Where one generalization breaks down, so will the other.

To my mind, examples such as this do a good job of motivating the idea that there is a notion of explanatory depth, analogous to that identified by Hitchcock and Woodward (2003b) operative in the metaphysical domain. What's more, as was noted above, several well fleshed-out interventionist theories of metaphysical explanation are already on the table (e.g., Schaffer, 2016, 2017; Wilson, 2016, 2018; Reutlinger, 2017; Miller and Norton 2022a, b), theories which would require little alteration in order to make use of this notion.

However, the recent literature is also replete with attempts to characterize metaphysical explanation in analogy with inferential accounts of scientific explanation. Dasgupta (2017), for example has recently motivated an 'analogue of the "DN model"' upon which 'the thing to be explained follows from the explainer together with something like a "metaphysical law"' (2017: 80).[9] Similarly, David Kovacs (2020) follows Kitcher (1981, 1989) in claiming that metaphysical explanation results from seeing 'how a large number of phenomena are the consequences of a small number of basic facts, from which they can be derived using relatively few and similar patterns of derivation' (2020: 1673).[10]

Insofar as any adequate analysis of explanation ought to be able to account for the sense in which explanation comes in degrees, an interventionist methodology clearly outshines its inferential counterparts. This is not to say that those who provide inferential accounts of metaphysical explanation explicitly endorse a scope-based account of depth; in fact, any reference to depth appears to be entirely absent from this literature. Nonetheless, until an alternative to a scope-based account is put forward, the superiority of an interventionist analysis of explanatory depth, over its inferential rivals, represents a key motivation for its adoption in cases of both causal and metaphysical explanation.

Admittedly, the example discussed in this section may strike the reader as an uninteresting application of the notion at issue. As a result, my analysis of explanatory depth within metaphysics might seem like a relatively modest benefit of an interventionist methodology. In response to this thought, I now want to turn to two much more interesting cases of metaphysical explanatory depth *in the wild,* as it were. In the next section, I shall demonstrate that *progress* within a lively area of contemporary metaphysics research, concerning the identity and distinctness of concrete objects, can be interpreted through the framework provided above: in terms of

[9] Also see Wilsch (2015, 2016). Reutlinger similarly notes that 'a friendly amendment of the covering law account may even allow for… metaphysical covering laws (such as general statements about one kind of facts grounding another kind of facts)' (2017: 241). Unlike Dasgupta (2017), however, Reutlinger quickly dismisses this account as a viable option, owing to 'well-known problems of the covering-law account' (2017: 241). See Salmon (1989) for a survey of these problems.

[10] Baron and Norton (2019) also defend a unificationist account of metaphysical explanation which utilizes Friedman (1974) and Kitcher's (1981, 1989) analogous models of scientific explanation. Of course, I would not want to claim that unification is an irrelevant factor in assessing competing explanation. The important point, however, is that unification is better understood in interventionist terms as being a result of invariance under testing interventions; the wider this range, the more unifying the explanation. See Hitchcock and Woodward (2003b: 192–194) and Woodward (2003: 356–371) for more on the role of interventions in characterising the unificatory dynamic of explanation.

progress through metaphysical explanations of increasing depth. In Sect. 5, I then turn this methodology to debate surrounding the nature of explanation itself.

4 Explaining identity and distinctness

In a recent paper, Shumener (2020) seeks to provide a novel metaphysical explanation of the identity and distinctness of concrete objects. Before presenting her own account, Shumener discusses two others. The first, the *qualitative properties proposal* suggests that identity facts of the form [$x = y$] are explained by the fact that x and y share all of their qualitative properties.[11] The second, the *weak discernibility proposal,* suggests that such identity facts are explained by the fact that x and y stand in only *reflexive* relations to one another.[12] According to Shumener's own *quantitative properties proposal* on the other hand, identity facts are explained by the fact that x and y stand in *quantitative* relations to each other *non-fundamentally*.[13]

Unfortunately, a critical analysis of Shumener's argument is beyond the scope of this paper. My interest in this account is, rather, in the dialectical trajectory of the debate; an intuitive interpretation of which can be given in interventionist terms, as progression through metaphysical explanations of increasing depth. First, note that each of the above explanations of identity, and the converse explanations of distinctness, constitute something approaching an explanatory generalization. We can reinterpret each to give a generalization which fits better with the example already discussed:

(c) The qualitative properties proposal: for any objects x and y, if x and y share all of their qualitative features, then x is identical to y; and if x has some qualitative feature that y lacks, then x and y are distinct.
(d) The weak discernibility proposal: for any objects x and y, if x and y only stand in reflexive relations to one another, then x is identical to y; and if x stands in an irreflexive relation to y, then x and y are distinct.
(e) The quantitative properties proposal: for any objects x and y, if for any quantitative relation R that x and y stand in, x and y stand in R to one another non-fundamentally, then x is identical to y; and if for at least one quantitative relation Q that x and y stand in, x and y stand in Q to one another fundamentally, then x and y are distinct.

[11] A precise definition of a qualitative feature is difficult to come by, although, Shumener suggests that: 'qualitative features are those that do not involve the identity relation or involve specific relations. So, for example, *5 km mass, adjacent to, same colors as* are qualitative features' (2020: 2079). By specifying that only qualitative properties ground identity, we avoid the trivial possibility that the property of identity *itself* grounds facts about what is identical to what. See Black (1952: 11).

[12] See Saunders (2006).

[13] Quantitative properties or features are, like qualitative features, difficult to formally define. However, of specific interest to us here, are those taken from our physical theories, having a determinate-determinable structure and admitting of degrees: 'examples of determinate quantitative relations include: *five meters away from, twice as massive as, opposite charge as* and the like' (Shumener, 2020: 2084).

By adopting the interventionist analysis of depth, we can provide a satisfying characterisation of why (e) provides a deeper metaphysical explanation of identity facts than (d) and, in turn, why (d) provides a deeper metaphysical explanation of identity facts than (c). Once again, what we are chiefly looking for, in justifying the claim that the *quantitative properties proposal* provides the deepest explanation of identity and distinctness, are hypothetical counterfactual scenarios in which one generalization is violated while another remains invariant.

In order to show how the *weak discernibility proposal* provides greater metaphysical explanatory depth than the *qualitative properties proposal*, we require a counterfactual scenario in which it can be shown that qualitatively identical objects can be numerically distinct. Such a scenario will be one under which the generalisation specified in (c) is *not* invariant. This scenario must, of course, be one in which it remains the case that the objects involved stand only in irreflexive relations to one another, thus securing the invariance of (d). As luck would have it, Max Black (1952) has popularized a case which fits the bill.

Black (1952) imagines a possible world containing only two spatially separated objects, *A* and *B,* which possess *different* qualitative properties. For the sake of argument, let's assume that *A* is spherical, and *B* is cuboid. Now, according to the generalization specified by (c), if *x* and *y* share all of their qualitative properties, then they are identical; and if *x* has some qualitative feature which *y* lacks, then *x* and *y* are distinct. Since *A* possesses the property of 'being spherical', which *B* lacks, (c) appears to be a candidate explanation of the distinctness of *A* and *B*.

In order to assess the *depth* of this explanation, however, we need to assess the range of same-object counterfactuals under which the relevant generalization will remain invariant. So, now imagine that we intervene upon *A* or *B* (or both), resulting in a counterfactual scenario in which they now share all of their qualitative properties (they are the same size, shape, mass etc.), yet remain spatially separated. They are, in other words, indistinguishable in terms of their qualitative properties.

In such a scenario, (c) is no longer explanatory. *A* does not possesses any qualitative properties which *B* lacks, and yet *A* and *B* are not identical (since they remain spatially separated). As Shumener argues, '[t]he Qualitative Properties Proposal cannot account for the distinctness of the spheres because there is no qualitative feature that one sphere has that the other lacks' (2020: 2080). Here then, we have a testing intervention under which (c) is not invariant. Since *x* and *y* share all of their qualitative features and yet they are not identical.

Conversely, as Saunders (2006) has argued, (d) remains invariant in such cases. This is because, while all of the *qualitative* relations in which *x* and *y* stand to one another are reflexive, 'the spheres stand in irreflexive relations like *five meters away from* to one another' (Shumener, 2020: 2080). Thus, the explanatory generalization specified by (d) remains invariant under a testing intervention which (c) is not. As a result, on Hitchcock and Woodward's (2003b) understanding of explanatory depth, the *weak discernibility proposal* appears to provide a deeper explanation of the identity and distinctness of objects than the *qualitative proposal*.[14]

[14] The reader may well question whether (c) is actually an *explanatory* generalization at all. Doesn't showing that (d) provides greater explanatory depth than (c) also involve showing that (c) is, in fact, false? Assuming that explanation is factive, this means that the qualitative properties proposal doesn't

Just as before, in arguing that Shumener's (2020) *quantitative properties proposal* provides a deeper metaphysical explanation of identity and distinctness than the *weak discernibility proposal*, we need to locate a counterfactual scenario in which (e) remains invariant while (d) does not. Shumener provides us with another example which fits the bill: Deluxe Max Black cases, involving 'metaphysically possible scenarios in which there are co-located, qualitatively indiscernible objects' (2020: 2081).

So, if we return to our qualitatively indiscernible objects *A* and *B*, and now imagine a counterfactual scenario which involves them sharing the exact same spatiotemporal location, then (d) will no longer be explanatory. Since *A* and *B* will now be co-located, we cannot differentiate between them based upon the irreflexive spatiotemporal relations which they stand in to one another.[15] As such, on an interventionist account of metaphysical explanation, (d) is not invariant under testing interventions resulting in Deluxe Max Black scenarios.

Shumener's (2020) *quantitative properties proposal* relies upon a novel distinction between an object having properties fundamentally and non-fundamentally, which differs from a property *itself* being either fundamental or non-fundamental. While this distinction is of crucial importance to Shumener's argument in favour of (e), a detailed discussion of the intricacies of this approach is beyond the scope of this paper.

What is important, for our purposes, is that Shumener argues that the *quantitative properties proposal* can correctly characterize the objects involved in both Max Black and Deluxe Max Black cases as being distinct; despite sharing all of their qualitative properties in the former case, and additionally not standing in irreflexive spatiotemporal relations to one another in the latter. In both scenarios, the objects can be distinguished on the basis that they stand in *quantitative* relations to one another *fundamentally*. Were they identical, then they would stand in such relations to one another *non-fundamentally*.

To briefly recap, (c) is not invariant under interventions resulting in either Max Black and Deluxe Max Black cases. The *qualitative properties proposal* is thus invariant under the narrowest range of testing interventions, providing the shallowest of the three candidate metaphysical explanations highlighted by Shumener (2020). Because (e) holds with respect to both Max Black and Deluxe Max Black cases, Shumener's own *quantitative properties proposal* is invariant under the widest range

Footnote 14 (continued)
merely provide a shallower explanation of the identity and distinctness of objects, it provides no explanation at all. As a result, one might think that my account of explanatory depth differs from Hitchcock and Woodward's in an important respect. I am entirely sympathetic to this concern and, in Sect. 6, I explain that this situation is exactly what we ought to expect from analogous interventionist analyses of explanation within methodologically divergent domains.

[15] As Shumener points out, certain "symmetrized states" of quantum particles, appear to be cases of the Deluxe Max Black variety, containing multiple subatomic particles 'which are not distinguished on the basis of their positions' (2020: 2082). See, e.g. French (1989).

of same-object counterfactual scenarios and, as a result, can be seen to provide the deepest metaphysical explanation of the identity and distinctness of concrete objects. And finally, since (d) is invariant under counterfactual scenarios resulting Max Black cases, but not Deluxe Max Black cases, the *weak discernibility proposal* occupies the middle-ground in terms of metaphysical explanatory depth. (d) has a wider range of invariance than (c), but a narrower range of invariance than (e).

I believe that this application of metaphysical explanatory depth naturally gestures towards another, even more interesting application. This second application becomes apparent once we notice that debate surrounding the nature of explanation is *itself* metaphysical in character. As I shall argue in the next section, what this means is that by adopting an interventionist account of metaphysical explanation, along with the account of metaphysical explanatory depth put forward in this paper, we can show that the explanations provided by interventionism are deeper than those supplied by rival accounts.

What is more, I will argue that in the very act of adopting interventionism with respect to metaphysical explanation, the range of counterfactual scenarios under which such meta-explanatory generalizations remain invariant, widens. Which is to say, interventionism provides deeper explanations of the connection between explanans and explanandum, if one additionally accepts that the interventionist analysis ranges over metaphysical, as well as causal, instances.

5 Depth in the metaphysics of explanation

So far, we have encountered two opposing positions upon the nature of explanation. On the one hand, DN and unificationist accounts suggests that explanation is *inference*. On the other, the interventionist account suggests that explanation is intimately connected to *manipulation*. However, a third theory, popularized by Salmon (1984, 1989, 1994) and Philip Dowe (1992, 2000), suggests that 'causal processes, causal interactions and causal laws provide the mechanisms by which the world works' (Salmon, 1984: 132). On this, *conserved quantity* (CC) account, an explanation of y in terms of x is the result of a causal interaction resulting in the exchange of a conserved quantity (e.g., energy, momentum, charge etc.).

In the previous section we saw that Shumener (2020) takes conflicting accounts of the identity and distinctness of concrete objects to give competing explanations for facts of the form [$x = y$]. Similarly, one can view each of the above theories as attempting to motivate alternative explanations for facts of the form [x explains y]. Here, differing accounts of what it takes to explain a phenomenon can be seen to provide meta-explanatory generalizations:

(f) The Inferential Proposal: For any x and y, if y can be logically deduced from x and the laws in question, then x explains y.
(g) The Conserved Quantity Proposal: For any x and y, if x and y causally interact in such a way that facilitates the exchange of a conserved quantity from x to y, then x explains y.

(h) The Interventionist Proposal: For any x and y, if it is possible to intervene on x with respect to y, in such a way that changes the value of y if at all, only through x, then x explains y.

Using the analysis detailed above, I believe that we can show that explanations which make use of (h) are deeper than those which make use of (f) and (g). Our first step is to show that interventionism provides deeper explanations than inferential accounts.[16] This is not a difficult task since such accounts face well known problems in characterizing the intuitive asymmetry of explanation. Sylvain Bromberger (1965) provides the most familiar example of such a case, concerning a flagpole and its shadow.[17]

Bromberger's example illustrates that, given the length of the flagpole's shadow (*x*) (and the angle of elevation of the sun), we deduce the height of the flagpole (*y*). And yet, the length of a flagpole's shadow does not *explain* its height. What this means, in the parlance of Hitchcock and Woodward (2003b), is that the generalization specified by (f), will not remain invariant in such cases, where *y* can be deduced given *x* and the laws in question, despite the fact that *x* does not explain *y*.

The interventionist generalization specified by (h) fairs much better here. Part of the central motivation for interventionism itself stems from its ability to accurately characterise the asymmetric character of such explanations. As (h) suggests, while intervening upon the height of the flagpole would allow us to manipulate the length of its shadow, the converse relation does not hold. That is, one cannot manipulate the height of the flagpole by intervening upon its shadow. Any such intervention would, itself, have to *go through* the height of the flagpole.[18] Consequently, (h) correctly characterises Bromberger's example as unexplanatory. Thus, the interventionist account will remain invariant under a wider range of counterfactual scenarios than inferential accounts and can be considered to provide a deeper *metaphysical* explanation of the connection between explanans and explanandum as a result.

The CC proposal faces similar problems to inferential accounts.[19] As Henk de Regt argues, Salmon's concept of causality is problematic at the deepest level of physical reality, where standard interpretations of quantum mechanics leave no room for 'continuous space-time trajectories along which energy and momentum are transported' (2017: 61). Similarly, Woodward argues that:

> [t]here are explanations, such as those involving causation by omission or by double prevention, that do not involve a physically interesting form of action at

[16] While the DN and unificationist pictures of explanation do differ in several respects, de Regt highlights that 'Kitcher's unificationist model turns out to be a sophisticated version of Hempel's deductive-nomological model, preserving the basic features of deductive argument and subsumption under laws' (2017: 53). As a result, for our current purposes it is not necessary to draw a meaningful distinction between the two theories.

[17] Also see Barnes (1992).

[18] See Woodward (2003: 98–102).

[19] Hitchcock (1995) notes that the CC proposal is, in fact, unable to make sense of many of the counterexamples to the DN model put forward by Salmon (1984) himself.

a distance, but are nonetheless cases of causal connection without intervening spatiotemporally continuous processes or transfer of energy momentum from cause to effect' (2003: 353).

Such cases, of course, will constitute interventions in which the criteria specified by (g) are not invariant since, although the connection between x and y is (at least widely accepted to be) explanatory, the criteria specified by the relevant explanatory generalization are not met. Woodward (2003: 350–373), however, argues that interventionism can correctly characterize such instances, providing principled reason to deny that explanation is connected to the transfer of a conserved quantity along continuous space-time trajectories. As such, we can once again see that the range of interventions under which (h) will remain invariant is wider than the equivalent range for (g). Thus, interventionism provides deeper explanations of the connections between explanans and explanandum than the CC proposal.

So, I have argued that by adopting an interventionist reading of metaphysical explanation, along with the account of metaphysical explanatory depth outlined in this paper, we can show that interventionism provides the deepest account of the nature of explanation *itself*. However, in applying interventionism at this meta-explanatory level, to generalizations purporting to account for facts of the form [x explains y], it might not be immediately obvious what the "object" of the requisite "same-object" counterfactuals actually is.

On the traditional account of interventionism, discussed in Sect. 2, an intervention involves altering the value of an explanans variable in order to manipulate an explanandum variable. In the case discussed above, we can intervene upon the hight of the flagpole in order to manipulate the length of the shadow. It is precisely because flagpoles can be used to manipulate shadows that we take flagpole height to explain shadow length, and not *vice versa*. In terms of same-object counterfactuals, the "object" here is the flagpole.

However, interventionism (on my view, at least) does not merely claim to be a theory of the relationship between flagpoles and shadows, it claims to be a theory about the relationship between explanans and explanandum *tout court.* As such, in attempting to test the invariance of (f), (g) and (h) as accounts of facts of the form [x explains y], the "object" of the relevant "same-object" counterfactuals cannot be the flagpole alone. Briefly revisiting the example from the previous section will help us to get a grip on what is happening at this meta-explanatory level.

Recall that the generalizations (c), (d) and (e) purport to explain something about the relationship between x and y, namely [$x = y$]. I have argued that on an interventionist account of metaphysical explanation, we can assess the depth of these generalizations by considering the range of interventions under which they remain invariant. However, such interventions are admittedly slightly different from those described in Sect. 2. In the case of the identity and distinctness of concrete objects, we do not intervene on *A with respect to B*; our goal is not to attempt to manipulate one of our spheres *by* intervening upon the other.

Rather, we intervene upon the features of A and B and their relations to one another (x), in order to attempt to "manipulate" the relation of identity and

distinctness itself (*y*). In this sense, the "object" of the same-object counterfactuals in question ought to be thought of as the combination of *A*, *B* and the features they possess. It is by intervening on this system, constituting *x*, that we are able to assess the range of interventions under which (c), (d) and (e) remain invariant and thus, the depth of the explanations they provide. As I see it, the "object" in the meta-explanatory case is the same.

On an *inferential proposal*, for example, the fact that we can deduce the length of the flagpole's shadow from its height and the relevant laws, itself explains why we take flagpoles to explain shadows. In this sense, the object of the relevant counterfactual (*x*) is the system constituted by the flagpole, its shadow and the deductive relationship in which they stand; with the explanation relation itself being the thing which we are attempting to manipulate (*y*). As we have already seen, however, the fact that we can construct counterfactual scenarios, involving the same objects, in which we can deduce the height of the flagpole from the length of its shadow suggest that (f) provides for relatively shallow explanations of the relationship between explanans and explanandum here.[20]

As the above discussion shows, examples of causal explanation are enough to prove that (h) will remain invariant under a wider range of interventions that either (f) or (g). However, the *true* depth of the interventionist account of explanation becomes apparent only once one acknowledges that metaphysical instances ought to contribute to the range of counterfactual scenarios against which competing theories of explanation are measured.

To take an archetypal example, given the existence of Socrates, and a relevant law (e.g., *set formation* as embedded in Zermelo–Fraenkel set theory, see Shaffer 2017: 309–310), we can deduce that the set {Socrates} existed; there are no possible worlds in which Socrates existed, but the set {Socrates} did not. However, the necessary connection here runs in both directions. Given that the set {Socrates} existed, Socrates' existence can also be deduced. So, according to (f), *x* explains *y* and *y* explains *x*. Yet the consensus suggests that it is the existence of Socrates explains the existence of the set {Socrates}, but not *vice versa*. Once again, an inferential account will not be invariant in such cases, where the occurrence of *x* can be deduced, given *y* and the laws in question, despite the fact that *y* does not explain *x*.

While the DN model sees explanation where we typically take there to be none, the CC account faces the opposite dilemma. For reasons which we have already seen, the CC account will have great difficulty making sense of the explanatory connection between the existence of Socrates and the set {Socrates}. The connection

[20] It is important to note that while the antecedents of such counterfactuals consist of multiple different objects and relations these objects stand in to one another, the objects themselves remain the same in each counterfactual considered. While we can alter these objects in various ways in order to assess the impact upon the relation of explanation which holds (or doesn't) between them, this is a far-cry from the sort of scope based "other-object" counterfactual model discussed in Sect. 2. It would perhaps be more accurate to call the counterfactuals involved in instances of metaphysical explanation "same-object[*s*]" counterfactuals, however I think that this would prove more confusing than helpful. For more on the possibility of the same/other-object distinction breaking down in the case of metaphysical explanation, see Sect. 6.

between x and y, here is not mediated by spatiotemporally continuous processes of conserved quantity transfer. Indeed, Woodward is cognizant of this difficulty: '[t]here are reasons to doubt that [the CC account] is an extensionally adequate theory, in the sense that it correctly distinguishes between causal and noncausal interactions' (2003: consensus suggests 30).

What this means, of course, is that while (f) is violated in cases where metaphysical explanation is intuitively asymmetric, (g) will fail to capture the explanatory character of metaphysical explanations *altogether*. Since, as has been seen at some length, (h) accurately characterises such explanations with metaphysical character, every such instance represents a widening of the range of interventions under which (h) will be invariant when compared to (f) and (g). Thus, any metaphysical explanation will be an addition to range of invariance for (h), but not (f) and (g).

There is, however, a further benefit of accepting that metaphysical explanations ought to contribute to the range of counterfactual scenarios against which depth is measured. In the very act of adopting interventionism with respect to metaphysical explanation, the range of testing interventions under which explanations of the connection between explanans and explanandum remain invariant, widens. Which is to say, interventionism *itself* provides even deeper explanations of the connection between explanans and explanandum if one additionally accepts that this analysis ranges over metaphysical, as well as causal instances.

As a result, the account of metaphysical explanatory depth which I have attempted to elucidate appears to provide some motivation for those interventionists still on the fence with respect to metaphysical explanation. Accepting that metaphysical explanations can be given an interventionist treatment significantly wides the range of counterfactual scenarios under which (h) remains invariant. Thus, explanations provided by interventionists who reject an analogous analysis of metaphysical explanation will be *shallower* than those provided by interventionists who accept this analysis.

6 Methodological concerns

In this final section, I wish to discuss several methodological questions which arise as a result of the account of metaphysical explanatory depth which I have attempted to motivate. The first concerns how we are to make sense of the distinction between same- and other-object counterfactuals in the context of identity and distinctness; the second, raises the issue of how we are to understand the role of interventions when it comes to metaphysical explanations more generally; and the third asks whether shallow metaphysical explanations are actually *explanations* at all.

With regards to the first concern, the worry here is that when considering explanations of identity and distinctness, the border between same- and other-object counterfactuals breaks down. If so, then it appears that the interventionist analysis would collapse into a scope-based account. In order to see why this is not the case, we will need to assess how a scope-based account of depth would cope in such scenarios.

So, let's return to our possible world containing spatiotemporally separated qualitatively discernible objects: *A,* a sphere; and *B* a cube.

According to the *qualitative properties proposal*, since *A* and *B* are qualitatively discernible, they are distinct. However, in order to assess the depth of this explanation, on a scope-based account, we are invited to assess the range of other-object counterfactuals under which (c) will continue to hold. Imagine, for example that we substitute *A* and *B*, for *C* and *D*; objects which are *indiscernible* in terms of their qualitative properties.

Despite the fact that *C* and *D* share all of their qualitative properties, *were* it the case that *C* possessed the qualitative properties of *A*, and *D* possessed the qualitative properties of *B*, *C* and *D* would be distinct. In this scenario, because *C* would possess the property of being spherical, which *D* lacks, (c) would continue to hold. The problem for scope-based accounts, is that on an interventionist analysis of explanatory depth in terms of *same*-object counterfactuals, (c) *does not* remain invariant in scenarios concerning qualitatively indiscernible, spatially separated objects.

On an interventionist analysis, we do not substitute *A* and *B* for other objects in order to assess the scope of the relevant generalization, but rather imagine intervening upon *A* and *B* (or both) in such a way that results in a situation in which they share all of their qualitative properties. Here (c) is violated; it will mischaracterize *A* and *B* as being identical when we know them to be distinct (being spatially separated). So, it appears that a scope-based account is once again unable to draw a meaningful distinction between these competing accounts of the identity and distinctness of concrete objects in terms of depth; a distinction which *can* be drawn if we adopt an interventionist notion of depth.[21] As a result, it is clear that the contrast between same- and other-object counterfactual analyses of depth holds firm, even in scenarios involving the identity and distinctness of concrete objects.

In order to address the second methodological concern, I would first like to make an admission: I am what Nicholas Emmerson (2021) has recently labelled an "intervention liberal". It is my view that interventions do not carve nature at its causal joints. I take it that, in at least *some* cases, interventions provide us with a useful tool in characterising metaphysical explanation. As we saw in Sect. 3, the hypothetical interventions used to assess the depth of competing explanations for Jeff Bezos's being a billionaire, are entirely "possible" in the requisite sense, corresponding to 'conceptually possible or well-defined physical manipulations' (Woodward, 2018: 122).

Even in the (admittedly more contentious) case concerning the identity and distinctness of concrete objects, I see little difficulty in conceptualizing the relevant interventions. It seems entirely possible to imagine a hypothetical scenario in which,

[21] It is important to note that the same situation holds with respect to the *weak discernibility proposal.* Supposing that *A* and *B* are qualitatively indiscernible, spatially separated objects, but that *C* and *D* are qualitatively indiscernible, co-located objects. *Were* it the case that *C* and *D* stood in the same relations to one another as *A* and *B*, then *C* and *D* would be distinct according to (d). On an interventionist analysis (d) would be violated in this case, while (e) would remain invariant. As such, on a scope-based account, we are unable to draw meaningful distinction between the *weak discernibility* and *quantitative properties proposals* in terms of their explanatoriness.

at t_1, only a single object exists (to which *A* and *B* both refer), and then to consider the implications of a manipulation which sees this single object split into two spatially separated objects, at t_2; objects which are qualitatively discernible from each other, *and* from the original object which existed at t_1.[22]

Despite this, there is not universal agreement on this point. Others, labelled "intervention puritans" by Emmerson (2021), believe that interventions exclusively serve to demarcate causal relationships, and that the notion becomes problematic in the context of noncausal explanation.[23] The worry is that, in cases of metaphysical explanation, the requisite interventions are neither well-defined, nor (in some cases at least) logically or metaphysically possible.[24] It is, however, beyond the scope of this paper to mount a sustained defence of the role of interventions with respect to noncausal explanation. For my purposes it suffices that there are already well developed interventionist approaches to metaphysical explanation on the table, approaches which are able to accommodate the account of metaphysical explanatory depth which I have provided above.[25]

And this brings us to our final methodological concern: whether shallow metaphysical explanations are actually explanatory at all. Consider an example used by Hitchcock and Woodward (2003b) to illustrate the notion of depth with respect to causal explanation: the laws of Newtonian mechanics and Einstein's relativistic correction to those laws. When applied to objects with a velocity that is relatively small compared to that of light, generalizations generated by Newtonian mechanics will remain invariant under a range of interventions, *R*, on that velocity.

However, the special relativistic correction to these laws will remain invariant under a much wider range of interventions R^*, where R^* strictly contains *R*, but also contains interventions upon velocities closer to that of light. In this sense the special relativistic corrections to Newtonian mechanics provide for deeper explanations

[22] It seems that I am in good company on this point. As I mentioned in Sect. 2, Schaffer (2017) explicitly characterises "metaphysical laws" in terms of invariance under testing interventions. Similarly, Wilson (2018) describes a wide variety of metaphysical explanations which can be accurately characterized using interventionist counterfactuals and structural equation models. Also see Schaffer (2016), Wilson (2016), and Miller and Norton (2022a, b).

[23] Proponents of intervention puritanism include: Bokulich (2011), Leuridan (2012), Saatsi and Pexton (2013), Harinen (2014), Pexton (2014), Jansson (2015), Rice (2015), Romero (2015), Baumgartner and Gebharter (2016), Baumgartner and Casini (2017), French and Saatsi (2018), Jansson and Saatsi (2019), Khalifa et al., (2018, 2020), Reutlinger (2018), Saatsi (2018), Lange (2019).

[24] To the extent that intervention liberalism appears to require a commitment to *counterpossible non-triviality,* Schaffer (2016) highlights that there are already good reasons for thinking that counterpossible scenarios require non-trivial evaluation (see e.g., Restall 1997, Goodman 2004, Priest 2005, Berto and Jago 2013, Jago 2015). What's more, largely as a result of their perceived utility in scientific explanation, recent years have seen a dramatic increase in attempts to motivate non-trivial counterpossibility (e.g., Baron and Colyvan 2021; Baron et al., 2020; Kimpton-Nye 2020; Reutlinger et al., 2020; Tan, 2019; Wilson 2021). Consequently, it is clear that counterpossibles pose a problem of interventionism *in general*, not just for interventionist interpretations of metaphysical explanation.

[25] As a final note on this topic, I would highlight that regardless of whether one is able to *imagine,* or *conceive of*, the sort of hypothetical manipulation required by an interventionist analysis of metaphysical explanation, Schaffer argues that there is nothing formally problematic here: '[t]he mathematics doesn't "know" if an intervention is countermetaphysical or counter logical. It just sees adjusted values to variables and adjusted functions, which it solves as before' (2016: 71).

insofar as they remain invariant under a wider range of testing interventions, R^*, despite the fact that Newton's laws *are* explanatory within the narrower range R.

In this example, it seems that we are comparing two *explanatory* generalizations; the "laws" of Newtonian mechanics appear to provide explanations within a given domain (concerning objects with velocities which are relatively small compared to that of light), despite failing to be invariant when this domain is expanded. In the cases of metaphysical explanation discussed throughout this paper, however, one might think that something rather different is going on. We have not been comparing *explanatory* generalizations at all. Such generalizations *compete* and, as a result, only one of them can be true and thus, explanatory.

By showing that the *quantitative properties proposal* provides deeper explanations than both the *qualitative properties proposal* and the *weak discernibility proposal*, what we are actually doing here is showing that the latter theories are false, and thus couldn't have been explanatory in the first place. Presuming that explanation is factive, by demonstrating that (e) is invariant under a wider range of interventions, we provide *counterexamples* to both (c) and (d), which means they must be false and cannot qualify as metaphysically explanatory generalizations.

Woodward (2021) has recently argued that we should not think of invariance as *evidence* of truth. Rather, 'invariance in relationships is a matter of the holding of certain kinds of truths–truths that we regard as particularly important to discover, rather than something that competes with truth or is evidence of truth' (Woodward, 2021: 266). That x, if true, would best explain y, is no reason to think that x *is* true, according to Woodward, because to show that the premises of an explanation are true 'we need to appeal to independent evidence in support of such truth claims' (2021: 266). However, in metaphysics, such independent evidence will typically underdetermine which, if any, of the premises of our candidate explanations are true.

Instead, in analysing competing metaphysical theories, we are typically required to "grant" or "assume" their truth *for the sake of argument*, and then assess what each theory would commit us to were it, in fact, true. Such theorizing is not uncommon and arises throughout scientific practice. As Alastair Wilson (2021) highlights, physics is difficult, and false theories abound. If we are to have any hope of progressing towards the correct fundamental theory, thinking critically about various competing possibilities, and evaluating them by contrasting their consequences, is a methodological imperative. However, we are not required to judge hypothetical scenarios as *objectively* possible in order to investigate such theories. Wilson (2021) draws a helpful analogy with reductio arguments in mathematics to make this point.

Classically, mathematical statements are taken to be true if possible. As a result, in order to reason nontrivially about false mathematical claims, we must be able to reason nontrivially about the impossible: '[m]athematicians may use a reductio argument to establish the falsity of a claim that they already know to be false (e.g., when teaching students)' (Wilson, 2021: 1121). Reasoning in this way clearly requires that mathematicians be able to temporarily grant that the claim in question is true and hence, possible. In this sense, physicists, mathematicians (and metaphysicians) can 'adopt a noncommittal pretence of possibility for the sake of argument' (Wilson, 2021: 1121).

When presented with the *qualitative properties* and *weak discernibility proposals*, the metaphysician does not have epistemic access to the truth or falsity of (c) and (d). Without the relevant empirical evidence to help us, the metaphysician can instead adopt a "noncommittal pretence of possibility for the sake of argument" and then proceed to examine the range of testing interventions under which each remains invariant. In this way, the metaphysician can demonstrate that the *weak discernibility proposal* provides deeper potential explanations of the identity and distinctness of concrete objects that the *qualitative properties proposal*, without needing to accept either theory as true.

Of course, exactly how we are to cash out the process involved in adopting a noncommittal pretence of possibility is another question. We could, for example, follow Toby Handfield (2004) and embed problematic counterfactuals in indicative conditionals. Alternatively, we might, as Wilson puts it, "go metatheoretical" and replace counterfactual reasoning with 'direct theorizing about models' (2021: 1119); or appeal to fictionalism in the make-believe style of Kendall Walton (1990), Roman Frigg (2010), and Sam Kimpton-Nye (2020).

While the reader might not be enamoured with any of these options (and others are available), the purpose of this discussion is not to promote any particular methodology.[26] Rather my aim is merely to highlight that this process, granting the truth of a theory for the sake of argument, is by no means uncommon. Given that the practices of both mathematics and physics appear to require us to account for such theorizing already, I take it as no concession at all that the analysis of metaphysical explanatory depth which I have provided here might also be able to make use of such an account.

To return to the original point, it is clear that the connection between truth and depth is much stronger in the case of metaphysical explanation than in the case of causal explanation. While Woodward (2021) argues that, in the causal case, depth and truth are entirely disconnected, in the metaphysical case, depth itself can provide (defeasible) reason to believe that a given theory is true. When we adopt a noncommittal pretence of possibility with respect to a metaphysical theory, and can find no testing intervention which violates it, this provides us with at least *some* justification for our belief in its truth.

Conversely, if the metaphysician comes across a testing intervention under which the theory appears to be violated, they have good reason to believe the theory to be

[26] More controversially, one might consider adopting something like the notion of *acceptance* put forward by Cohen (1992). According to Cohen, one *accepts* that *p* when one treats it as given, i.e., when one 'adopts a policy of… including [*p*] among one's premises for deciding what to do or think in a particular context' (1992: 4). For alternative accounts of acceptance, and how this notion might differ from belief, see e.g., Van Fraassen (1980); Bratman (1992) and Maher (1993). Finnur Dellsén (2017), has recently argued that *understanding*, the cognitive achievement which results from grasping an explanation, can be accompanied by mere acceptance, rather than full-blown belief: 'belief and acceptance will coincide in most cases. However, they can come apart, viz. when one decides to adopt a policy of treating something as given despite being indisposed to feel that it is true' (2017: 14). Dellsén argues that we can "treat" a theory as given and use it in our explanations of various natural phenomenon–thus accepting it for explanatory purposes-despite not believing it to be true and even, in some cases, where we believe it to be false.

false; although this needn't prevent us from cogently talking about the structure of the world that the theory describes. However, while the notion of depth appears to play a different role within metaphysics and science, this difference does not lead to a difference in the methodology of *assessing* or *calculating* explanatory depth across these domains. As I have shown, Hitchcock and Woodward's interventionist analysis requires little alteration in order to characterize the depth of both causal and metaphysical explanations.

Acknowledgements For helpful comments on previous incarnations of this paper, my thanks go to: Samuel Andrews; Joaquim Giannotti; Michael Townsen Hicks; Kerry McKenzie; James Norton; Noelia Iranzo Ribera; Katie Robertson; Stephan Roski; and Alastair Wilson. My thanks also go to both organizers and audiences at: the "Difference-Making and Explanatory Relevance" workshop, hosted by the Relevance Team at the University of Hamburg; and the 2021 Society for the Metaphysics of Science Annual Conference.

Funding This paper forms part of the project "A Framework for Metaphysical Explanation in Physics", 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).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Barnes, E. (1992). Explanatory unification and the problem of asymmetry. *Philosophy of Science, 59*(4), 558–571.

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

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

Baron, S., & Norton, J. (2019). Metaphysical explanation: The kitcher picture. *Erkenntnis, 86*(1), 187–207.

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.

Berto, F., & Jago, M. (2013). *Impossible worlds*. Oxford University Press.

Black, M. (1952). The identity of indiscernibles. *Mind, 61*(242), 153–164.

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

Bratman, M. (1992). Practical reasoning and acceptance in context. *Mind, 101*(401), 1–16.

Cohen, L. J. (1992). *An essay on believe and acceptance*. Clarendon 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.

Darden, L., & Craver, C. (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.

Dasgupta, S. (2017). Constitutive explanation. *Philosophical Issues, 27*(1), 74–97.
Dellsén, F. (2017). Understanding without justification or belief. *Ratio, 30*(3), 239–254.
Dowe, P. (1992). Wesley Salmon's process theory of causation and the conserved quantity theory. *Philosophy of Science, 59*(2), 195–216.
Dowe, P. (2000). *Physical causation*. Cambridge University Press.
Emmerson, N. (2021). A defence of manipulationist noncausal explanation: The case for intervention liberalism. *Erkenntnis*. https://doi.org/10.1007/s10670-021-00497-4
French, S. (1989). Identity and individuality in classical and quantum physics. *Australasian Journal of Philosophy, 67*(4), 432–446.
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.* Oxford University Press.
Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy, 71*(1), 5–19.
Frigg, R. (2010). Models and fiction. *Synthese, 172*(2), 168–251.
Glennan, S. (2017). The new mechanical philosophy. Oxford University Press.
Glennan, S. (2010). Mechanisms, causes, and the layered model of the world. *Philosophy and Phenomenological Research, 81*(2), 362–381.
Goodman, J. (2004). An extended lewis/stalnaker semantics and the new problem of counterpossibles. *Philosophical Papers, 33*(1), 35–66.
Handfield, T. (2004). Counterlegals and necessary laws. *Philosophical Quarterly, 54*(216), 402–419.
Harinen, T. (2014). Mutual manipulability and causal inbetweenness. *Synthese*, *195*(1), 35–54.
Hempel, C. (1959). The logic of functional analysis. In L. Gross (Ed.), *Symposium on sociological theory* (pp. 271–287). Harper & Row.
Hempel, C. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. The Free Press.
Hempel, C. (1966). *Philosophy of Natural Science. Prentice-Hall Foundations of Philosophy Series*. New Jersey: Prentice-Hall.
Hitchcock, C. (1995). Salmon on explanatory relevance. *Philosophy of Science, 62*(2), 304–320.
Hitchcock, C., & Woodward, J. (2003a). Explanatory generalizations, Part I: A counterfactual account. *Noûs, 37*(1), 1–24.
Hitchcock, C., & Woodward, J. (2003b). Explanatory generalizations, Part II: Plumbing explanatory depth. *Noûs, 37*(2), 181–199.
Jago, M. (2015). Hyperintensional propositions. *Synthese, 192*(3), 585–601.
Jansson, L. (2015). Explanatory asymmetries: Laws of nature rehabilitated. *Journal of Philosophy, 112*(11), 577–599.
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
Kimpton-Nye, S. (2020). Necessary laws and the problem of counterlegals. *Philosophy of Science, 87*(3), 518–535.
Kitcher, P. (1981). Explanatory unification. *Philosophy of Science, 48*(4), 507–531.
Kitcher, P. (1989). Explanatory unification and the causal structure of the world. In P. Kitcher & W. Salmon (Eds.), *Scientific explanation* (pp. 410–505). University of Minnesota Press.
Kovacs, D. (2020). Metaphysical explanatory unification. *Philosophical Studies, 177*(6), 1659–1683.
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.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Maher, P. (1993). *Betting on theories*. Cambridge: Cambridge University Press.
Miller K., & Norton, J. (2022a). *Everyday Metaphysical Explanation*. Oxford: Oxford University Press.
Miller K., & Norton, J. (2022b). Non-cognitivism about metaphysical explanation. *Analytic Philosophy*. https://doi.org/10.1111/phib.12258
Pexton, M. (2014). How dimensional analysis can explain. *Synthese*, *191*(10), 2333–2351.

Priest, G. (2005). *Towards non-being: The logic and metaphysics of intentionality*. Oxford University Press.
Restall, G. (1997). Ways things can't be. *Notre Dame Journal of Formal Logic, 38*(4), 583–596.
Reutlinger, A. (2017). Does the counterfactual theory of explanation apply to non-causal explanation in metaphysics? *European Journal for Philosophy of Science, 7*(2), 239–256.
Reutlinger, A. (2018). Extending the counterfactual theory of explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations.* 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.
Rocca, M. (2005). Two spheres, twenty spheres, and the identity of indiscernibles. *Pacific Philosophical Quarterly, 86*(4), 480–492.
Romero, F. (2015). Why there isn't inter-level causation in mechanisms. *Synthese, 192*(11), 3731–3755.
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.
Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton University Press.
Salmon, W. (1989). 4 Decades of scientific explanation. *Minnesota Studies in the Philosophy of Science, 13*, 3–219.
Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science*, 61(2), 297–312.
Saunders, S. (2006). Are quantum particles objects? *Analysis, 66*(1), 52–63.
Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies, 173*(1), 49–100.
Schaffer, J. (2017). Laws for metaphysical explanation. *Philosophical Studies, 27*(1), 303–321.
Shumener, E. (2020). Explaining identity and distinctness. *Philosophical Studies, 177*(7), 2073–2096.
Tan, P. (2019). Counterpossible non-vacuity in scientific practice. *Journal of Philosophy, 116*(1), 32–60.
Van Fraassen, C. (1980). *The scientific image*. Oxford University Press.
Walton, K. (1990). *Mimesis as make-believe: On the foundations of the representational art*. Harvard University Press.
Weslake, B. (2010). Explanatory depth. *Philosophy of Science, 77*(2), 273–294.
Wilsch, T. (2015). The nomological account of ground. *Philosophical Studies, 172*(12), 3293–3312.
Wilsch, T. (2016). The deductive-nomological account of metaphysical explanation. *Australasian Journal of Philosophy, 94*(1), 1–23.
Wilson, A. (2016). Grounding entails counterpossible non-triviality. *Philosophy and Phenomenological Research, 92*(3), 716–728.
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. 46–68). Oxford University Press.
Wilson, A. (2021). Counterpossible reasoning in physics. *Philosophy of Science, 88*(5), 1113–1124.
Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford University Press.
Woodward, J. (2018). Some varieties of non-causal explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation beyond causation: Philosophical perspectives on non-causal explanations.* Oxford University Press.
Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology*. Oxford University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.