Topoi
https://doi.org/10.1007/s11245-026-10462-x

Extending Causation Beyond Manipulation: Analogical Reasoning in Two Theories of Causation

Liyuan Jiao[1,2]

Received: 12 January 2026 / Accepted: 16 June 2026
© The Author(s) 2026

Abstract
Analogical reasoning has played an important role in philosophical theory, including manipulationist approaches to causation. Manipulationism ties causal judgments to the idea that causes function as handles for manipulating effects. This emphasis gives rise to the problem of unmanipulable causes: how should causal relations be understood in cases where manipulation is not possible? This problem poses a challenge for two influential versions of manipulationism, namely agency theory and interventionist theory, both of which seek to extend causal judgments beyond manipulable cases. This paper advances three claims. First, despite their differences, both approaches extend causal claims beyond contexts of direct manipulation in a way that presupposes analogical reasoning, even in cases where this is not explicitly acknowledged. Second, neither agency theory nor interventionist theory, as it stands, satisfies a condition on analogical warrant, since they fail to supply independent support for attributing a shared mechanism across contexts. Third, the paper examines whether appeals to laws of nature could supply such warrant and argues that this strategy ultimately fails. The analysis clarifies a structural limitation of manipulationist approaches to causation and points to the need for a more explicit account of how such analogical warrant may be established.

Keywords Causation · Manipulationism · Analogical reasoning · Unmanipulable causes · Warrant · Laws of nature

1 Introduction

Analogical reasoning is commonly employed in both ordinary and scientific contexts, often by drawing on similarities between familiar situations and unfamiliar ones that are taken to involve causal relations. Within philosophy, analogy also plays a role within theories of causation themselves, including manipulationist accounts, where causal claims are extended from manipulable to unmanipulable cases. This raises a further question: how is such analogical extension to be warranted when the causal features of the unmanipulable cases are not independently available, but are instead what the analogy is meant to establish? This paper examines this question by focusing on the manipulationist approach to causation and treatments of unmanipulable cases within the approach.

The manipulationist approach to causation has long been influential in both philosophy and the sciences. Early philosophical accounts along these lines can be found in the work of Collingwood (1940) and von Wright (1971;1975), while related ideas have also been developed in statistics, social science, and experimental design. In general terms, the central claim of this approach is that X causes Y if and only if, were X to be suitably manipulated, a corresponding change in Y would follow (cf. Woodward 2009, p. 234). This emphasis on manipulation gives rise to *the problem of unmanipulable causes*: how are causal relations between events that are not manipulable to be understood?

In contemporary literature, this problem poses a challenge for two main versions of the manipulationist approach, namely agency theory and interventionist theory. Agency theory explains causation in terms of what agents can freely bring about; interventionist theory abandons a reductive analysis in terms of agency and instead introduces the notion of an intervention as an abstract counterpart of manipulation.

✉ Liyuan Jiao
liyuan.jiao@durham.ac.uk

[1] Department of Philosophy, Durham University, Durham, UK

[2] Centre for Philosophy of Epidemiology, Medicine, and Public Health, University of Johannesburg, Johannesburg, South Africa

Published online: 23 July 2026

The notion of what is "unmanipulable" is understood differently across these accounts. In agency-based theories, it concerns limits on what agents can freely bring about. In interventionist frameworks, it is tied to whether a well-defined intervention can be specified: causal claims may be extended to cases in which interventions cannot in practice be carried out, provided that a well-defined intervention can still be specified within a structural model (cf. Woodward 2003, pp. 128–131).[1]

Despite this difference, both approaches seek to extend causal claims beyond cases of direct manipulation as a way of responding to the problem of unmanipulable causes. Agency theory does so explicitly by appeal to analogy. Interventionist theory does not characterise its strategy in these terms, but a closer examination suggests that a structurally similar move is involved. In Woodward's framework, causal claims about unmanipulable phenomena are assessed by evaluating the relevant counterfactuals under hypothetical interventions within structural models. As will be argued below, the construction and application of such models in unmanipulable cases depend on structural information and associated assumptions derived from manipulable contexts, and this dependence constitutes an implicit appeal to analogy.

This paper argues that such an appeal follows naturally from a manipulationist account of causation. If causal claims are grounded in manipulation, then extending those claims beyond manipulable cases requires drawing connections between manipulable and unmanipulable contexts. The central claim defended here is that both agency theory and interventionist theory rely on this form of extension, but that neither, as it stands, provides adequate grounds for treating the relevant analogical reasoning as warranted.

Before proceeding, it is important to clarify the limited aims of this paper. I do not seek to offer a general theory of causation, nor to provide a comprehensive defence or rejection of manipulationism. Rather, the focus is on one specific issue: how analogical reasoning functions within manipulationist approaches to address unmanipulable causes, and why its warrant remains insufficiently articulated.

The paper is organised as follows. Section 2 introduces agency theory and interventionist theory. Section 3 examines their respective strategies for addressing unmanipulable causes. Section 4 argues that these strategies, as they stand, do not supply adequate grounds for the analogical reasoning involved. Section 5 considers an appeal to laws of nature as a potential source of warrant. Section 6 addresses possible objections, and Section 7 concludes.

2 Manipulationist Approach to Causation

Manipulationist approaches to causation share the core intuition that causal relations are closely connected to the possibility of manipulation. On this view, to say that one event A causes another B is, in some sense, to say that manipulating A would make a difference to B. This basic idea has been developed in different ways, most prominently in agency theory and interventionist theory. Although these approaches differ in how they characterise manipulation, both face a common difficulty once the focus shifts to causal relations involving events that are not manipulable.

2.1 Agency Theory

Agency theory of causation analyses causation in terms of the sort of manipulations that agents might perform. Based on earlier work by Collingwood (1940), Gasking (1955), and von Wright (1971;1975), Peter Menzies and Huw Price (hereafter MP) (1993) have developed an account on which causation is reduced to agency.[2] The central claim of this theory is articulated by Menzies and Price (1993, p. 187), who write that "an event A is a cause of a distinct event B just in case bringing about the occurrence of A would be an effective means by which a free agent could bring about the occurrence of B."

In what follows, I reformulate this claim in variable terms by introducing a situation variable S to capture the relevant background conditions. On this formulation,

> An event A causes a distinct event B in a situation S if and only if bringing about A in S by a free agent would be an effective means to bring about B in S.

The appeal of this view lies in its close connection to everyday causal reasoning. Take turning on a lamp as an example: under normal circuit operating conditions (S), pressing the button of a lamp (A) causes the illumination of this lamp (B), for one freely pressing the button of this lamp is an effective means to bring about the illumination of this lamp when the circuit operates normally. In such cases, causal claims appear to be grounded in the perspective of agents who act on the world and observe the resulting changes.

Two notions play a crucial role in this account: that of a free agent and that of a free act. The notion of a free act

[1] Cases in which no well-defined intervention can be specified fall outside the scope of Woodward's framework. Such cases are treated not as instances of unmanipulable causation but as involving ill-defined causal claims. For further discussion, see Woodward (2003, pp. 113–115).

[2] For Price's later shift away from a reductive analysis of causation toward an anthropological interpretation of agency theory, see Price (2017, pp. 75–79).

is introduced by MP through the idea of *freely bringing about* an event, which MP take to reflect a distinctive agential experience rather than a causal notion; however, they do not provide a general definition of free acts beyond this characterisation. The notion of a free agent, by contrast, is employed to mark the scope of possible manipulation, understood broadly as what some agent could bring about under appropriate conditions.[3]

However, not all causal relations appear to fall within the scope of agency theory. Many widely accepted causal claims concern events that no agent could manipulate. It is commonly held, for example, that the 1989 San Francisco earthquake was caused by friction between continental plates. However, no agent could have brought about the earthquake by intervening on the plates themselves. Such cases place pressure on agency-based analyses of causation, since the relevant causal relations seem to hold independently of any possible manipulations by agents. A common objection to agency theory, illustrated by cases of this kind, is that they appear to restrict causation to relations that fall within the scope of agency. If causation is analysed in terms of what agents can bring about, it becomes unclear how causal relations involving events that agents cannot manipulate are to be understood. This concern gives rise to what is commonly described as the problem of unmanipulable causes. In response, agency theorists seek to extend causal claims beyond directly manipulable cases, typically by appealing to analogies between cases in which agents can manipulate events and cases in which they cannot.

2.2 Interventionist Theory

Interventionist theory retains the manipulationist intuition while abandoning a reductive appeal to agency. Rather than analysing causation in terms of what agents do, interventionist theory characterises causal relations by reference to interventions, as a more abstract and fine-grained version of manipulation. This shift is motivated in part by the difficulties faced by agency-based accounts in dealing with unmanipulable causes.

Building on the work of Spirtes et al. (1993), Judea Pearl (2000) and others, James Woodward (2003) develops an interventionist account of causation and explanation. On this view, causal claims are assessed in terms of hypothetical interventions defined by their role within a model of relations among variables, rather than by actions that might be performed by agents. Interventions need not correspond to physically performable manipulations; what is required is that a well-defined intervention can be specified within a structural model, even if it cannot in practice be carried out.

In Woodward's account, however, not every hypothetical change to a variable counts as an intervention. An intervention variable must satisfy a set of conditions governing its causal relations to the target variable and the outcome. Together, these conditions ensure that any resulting change in the outcome can be attributed to that target variable alone. In particular, the intervention variable must break the causal dependence of X on its prior causes, so that, when the intervention is in place, the value of X is determined solely by the intervention (Woodward 2003, p. 98; Woodward and Hitchcock 2003, pp. 12–13).

Within a structural model, such interventions are represented by replacing the structural equation for X with an assignment that fixes its value, which introduces a "new mechanism" in Judea Pearl's sense (2000). This representation presupposes *modularity*, that is, the assumption that the structural equations represent relatively independent mechanisms such that one equation may be altered without requiring simultaneous changes in others (Woodward 2003, p. 48). In this framework, causal relations are identified not merely by correlations, but by their stability under an appropriate range of interventions. Invariance thus functions as a condition for representing a dependency as causal, and structural equations, once established, are taken to represent stable causal relations across relevant interventions.

In manipulable cases, the conditions on intervention variables can in principle be physically realised. An intervention that satisfies them can be carried out. In unmanipulable cases, by contrast, no physical process may be available that satisfies these conditions. Their satisfaction is instead to be assessed within the structural model on the basis of background theoretical knowledge and observational evidence (Woodward 2003, pp. 128–131). In such cases, a well-defined intervention can still be specified within a structural model. It is precisely this class of cases that the interventionist framework seeks to accommodate. This framework thus allows causal claims to be extended beyond cases of direct manipulation. The question, however, is how this extension is to be understood and justified. The next section addresses this issue by examining the role of analogical reasoning in extending causal claims beyond manipulable cases.

3 Responses to the Problem of Unmanipulable Causes

This section considers how manipulationist theories address the problem of unmanipulable causes. It examines the responses offered by agency theory and interventionist

[3] Although MP do not give a clear definition of this notion, we may refer to Menzies's generalisation of "within my control" in his theory of scientific laws when formulating a characterisation of a free agent, which allows us to expand the domain of causation as fully as possible. For further discussion, see Menzies (1993, p. 208).

theory, focusing on how each seeks to extend causal claims beyond cases of direct manipulation. The concern here is not with evaluating these responses, but with making explicit the form of the extension they employ. As will become clear, both responses involve an extension that can be understood as a form of analogical reasoning, even where this dependence is not explicitly acknowledged.

3.1 MP on the Agency Account

In response to the problem of unmanipulable causes, MP do not abandon the agency account altogether. Instead, they propose a weakened version of the account designed to extend causal claims beyond cases of direct manipulation. The key move in this proposal is an explicit appeal to analogical reasoning. Rather than requiring that causal relations be grounded in an actual or possible exercise of agency, the account allows causal relations to be recognised by reference to similarities between manipulable and unmanipulable situations. The role of analogy is built directly into the formulation of the weakened version of the agency account, as the following definition makes clear:

> The agency account states that a pair of events are causally related just in case the situation involving them possesses intrinsic features that either support a means-end relation between the events as is, or are identical with (or closely similar to) those of another situation involving an analogous pair of means-end related events (Menzies and Price 1993, p. 197).

The core idea can be restated as follows:

> *An event A causes a distinct event B in a situation S if and only if:*

Either:

> *S* possesses intrinsic features *I* that support a means-end relation between *A* and *B*.

or:

> *S* possesses intrinsic features *I* that are identical with (or closely similar to) those *I** of another situation *S**, where *I** support a means-end relation between *A** and *B**.

The weakened account thus permits causal claims in unmanipulable cases by extending them from manipulable cases on the basis of similarity in intrinsic features. Although the notion of intrinsic features plays a central role in this extension, MP offer no explicit definition, allowing its meaning to emerge implicitly from the surrounding analysis. We may define this concept as follows:[4]

> An intrinsic feature *I* is one that an object possesses in virtue of itself alone, and not in virtue of its relations, environment, or companions. In particular, intrinsic features are not themselves causal or means-end relations.[5]

A similar lack of explicit definition applies to the notion of a means–end relation. It appears to correspond to this:

> A means-end relation holds between events *A* and *B* if and only if freely bringing about *A* would be an effective means by which a free agent could bring about *B*.

MP (1993, p. 197) explicitly maintain that "when an agent can bring about one event as a means to bringing about another, this is true in virtue of certain basic intrinsic features of the situation involved, these features being essentially non-causal though not necessarily physical in character." This claim establishes a crucial commitment: the means–end relations in manipulable cases are grounded in intrinsic features of the situation. However, MP do not specify the nature of this grounding. For the purposes of present analysis, it is useful to reconstruct this grounding as a structural relation *K* holding among intrinsic features.[6] *K* is not itself an intrinsic feature, but a relation among such features in virtue of which a means–end relation obtains.

On this reconstruction, the structure of MP's account can be made explicit. In a manipulable situation *S**, the means–end relation between *A** and *B** obtains in virtue of their intrinsic features standing in *K*. In unmanipulable situations *S*, no genuine means–end relation can obtain, since no agent could freely bring about *A*. Nevertheless, on this reconstruction, the intrinsic features *I* of *S* can be treated as standing in relation *K*, even though this relation cannot be exploited in action. The extension of causal claims then proceeds by analogy. If the intrinsic features *I* of an unmanipulable situation *S* are identical with, or closely similar to, the intrinsic features *I** of a manipulable situation *S** in respect of *K*, which, on the present reconstruction, captures the structural features in virtue of which means–end relations obtain, then

[4] In the context, dispositions are not to be classified as intrinsic features in their own right, but they supervene on intrinsic features that ground the relevant analogical inferences.

[5] It follows from the requirement of not being relations. For a canonical discussion of intrinsic properties, see Langton and Lewis (1998, pp. 17–24).

[6] This reconstruction goes beyond the letter of MP's text but serves to make explicit what their account appears to presuppose.

the causal relation between A and B in S can be extended from the manipulable case. The role of analogy is thus to license the projection of this underlying structure from cases in which it is grounded in agency to cases in which such grounding cannot be directly realised.

The 1989 San Francisco earthquake provides a clear illustration of how the weakened agency account is intended to handle unmanipulable causes. Although the earthquake is commonly taken to have been caused by friction between continental plates, no agent could have brought it about by intervening on the plates themselves. Nevertheless, the causal claim may still be extended by analogy. In scientific practice, seismologists employ physical models in which agents can manipulate elements representing tectonic plates and thereby bring about effects analogous to earthquakes, such as fractures in the model's surface. For example, moving certain components of the model may result in a crack forming. These manipulable cases involve intrinsic features that stand in relation K, in virtue of which a means–end relation is supported within the model. If the real geological situation possesses intrinsic features that are similar in the relevant respect K, then, according to the weakened agency account the causal claim is extended from the model to the real-world case.

The foregoing analysis indicates that similarity between intrinsic features—here reconstructed more precisely as similarity in respect of K—plays a central role in the extension of causal claims from manipulable to unmanipulable cases. However, the nature of this similarity remains underspecified, and it is unclear why it should be sufficient to warrant such an extension. The next section investigates this difficulty by examining Woodward's critical discussion of MP's proposal.

3.2 Woodward on the Interventionist Account

A central question arises when the interventionist framework is extended beyond direct manipulation: on what basis are the relevant causal claims evaluated? On Woodward's account, such claims are assessed by means of counterfactuals, that is, claims about what would happen to Y if X were changed by a well-defined intervention. The truth of such counterfactuals is determined by a causal model, by fixing the value of X in the relevant structural equation and holding the remaining equations unchanged (Woodward 2003, pp. 15, 122, 339).[7]

This raises a prior question that Woodward does not explicitly address: where do these structural equations come from in cases in which intervention is not available? His framework explains how counterfactuals are evaluated given a model, but it does not explain how such models are epistemically justified in contexts beyond direct manipulation.

In settings where intervention is available, part of the answer is relatively clear. By varying variables under controlled conditions, one can determine whether the relevant dependencies remain stable across changes, thereby supporting their representation as structural equations. Intervention thus plays a distinctive epistemic role: it not only reveals dependencies between variables but also provides evidence for their stability under change.

In contexts beyond direct manipulation, however, this grounding is unavailable. No intervention can be performed and no direct test of the relevant structural assumptions can be carried out. Yet if counterfactual claims are to have determinate content, structural equations must still be in place. Woodward treats the relevant causal structure as given, as part of background knowledge, without explaining how it is justified in such cases (Woodward 2003, pp. 128–131).

The claim advanced here is that, in such contexts, these structural equations are not constructed within the target domain itself, but are instead drawn from or informed by other contexts in which the relevant dependencies have been established and assessed. In other words, the relevant causal structure is established elsewhere and then brought into the target system.

This structure is particularly clear in the moon–tide case discussed by Woodward. The case is intended to show that causal claims can be evaluated even when no physically possible intervention satisfies the required conditions (Woodward 2003, pp. 128–131). However, this evaluation depends on two interconnected steps. First, at the level of model construction, tidal motion is represented as depending on gravitational attraction, which in turn depends on the position of the moon. This structure is not established through direct intervention on the moon–tide system. It is instead drawn from Newtonian mechanics as developed and applied in other contexts, including terrestrial experiments and a range of gravitational phenomena in which the relevant dependencies have been supported under conditions more amenable

[7] The interventionist appeal to counterfactuals draws on a counterfactual framework familiar from accounts of causation, most prominently Lewis (1973; 1979). By contrast with Lewis's possible-worlds semantics, interventionist theory employs counterfactual reasoning within structural causal models to extend causal claims beyond direct manipulation (Woodward 2003). For Woodward's discussion of similarities and differences between Lewisian and interventionist counterfactuals, see Woodward (2023).

to control. The structural equations are therefore introduced into the target system rather than derived from it.

Second, at the level of counterfactual evaluation, an intervention on a variable is represented by replacing its structural equation with a fixed assignment while holding the remaining equations unchanged. This step presupposes modularity: the other structural equations of the model are assumed to remain stable. In contexts where intervention is available, this assumption can in principle be tested. In the moon–tide case, no such test is available, and the assumption is applied without independent confirmation within the target domain. Taken together, these two steps show that the structural equations and their associated assumptions are not independently established within the target system. They are carried over from elsewhere. The question is how this transfer is to be justified.

In the absence of direct intervention, such transfer depends on some judgment about the similarity between the source contexts in which the structural equations were established and the target domain to which they are applied. The issue is not whether this judgment is ultimately correct, but that it is required: even if it is mistaken or insufficient in a given case, it remains a necessary condition for the extension to proceed.

However, not every similarity judgment is sufficient to license such a transfer. Some constraint is required to distinguish similarities that are relevant to the preservation of the causal dependencies from those that are not. The interventionist framework itself does not provide such a constraint. In the absence of direct intervention, the extension therefore depends on a judgment about which similarities matter. It is this kind of judgment that can be understood as analogical reasoning.

This point can be clarified by considering what follows if no such constraint is imposed. Suppose that any form of similarity, regardless of its connection to the relevant causal dependencies, were sufficient to license the projection of a model. Consider a case in which water supports the counterfactual that, if consumed, it quenches thirst. On the basis of superficial similarity alone, such as transparency and colourlessness, the same counterfactual could be projected onto other liquids that do not share the properties responsible for the original causal dependence. Such a projection is clearly unwarranted. The superficial similarity between the substances provides no reason to suppose that the features responsible for the original causal dependence are preserved. This shows that some constraint on similarity is required. Without it, the transfer of causal structure becomes unconstrained, and the framework cannot distinguish warranted from unwarranted applications.

The conclusion is therefore that the extension of interventionist reasoning beyond direct manipulation presupposes analogical reasoning as an essential component. It provides the basis on which structural equations and their associated assumptions are carried over across contexts. The question of under what conditions such analogical judgments are warranted is taken up in the following section.

4 Analogical Warrant and Its Limits

The previous section argued that both the weakened agency account and interventionist theory extend causal claims to unmanipulable cases, in ways that can be understood as analogical reasoning. In each case, causal claims concerning unmanipulable phenomena depend on treating these cases as sufficiently similar to paradigmatic manipulable situations. This section examines whether such extensions are adequately warranted. It asks under what conditions similarities between source and target domains can support the extension of causal claims beyond direct manipulation and evaluates how far these conditions are satisfied by the two theories. As will be argued, both accounts face important limitations at the level of analogical warrant.

4.1 What Is Required for Analogical Warrant?

Appeals to analogy are often used to extend causal claims from cases in which manipulation is possible to cases in which it is not. Such extensions raise an immediate epistemic question: under what conditions can an appeal to analogy provide warrant for extending causal claims? This section proposes a condition on analogical warrant and argues that as it stands, it is not met by either of the two accounts under examination. The condition introduced here will serve as the evaluative standard for the analyses that follow.

A familiar starting point is the observation that not all similarities are relevant to the conclusion drawn. Two cases may share certain features, such as observable properties, behavioural patterns, or formal descriptions, but such similarities alone do not justify extending a claim from one case to another. For this reason, only some similarities are relevant to the conclusion supported by an analogy (Hesse 1963; Bartha 2010). Paul Bartha's articulation model makes this more precise: a good analogical argument requires a prior association in the source domain—a connection, logical, causal, or statistical, between the known similarities and the further similarity projected to hold in the target domain. What matters is not the overall degree of similarity between the two domains, but whether this connection has been adequately articulated. However, this requirement leaves open a further question: what justifies treating a given connection as relevant for the projection?

A stronger version of this requirement appears in Wolfgang Pietsch (2019), who proposes that a predictive analogical inference holds if and only if the differences between source and target are causally irrelevant to the phenomenon in question, where causal irrelevance is understood in terms of difference-making. This condition captures a genuine insight: where the relevant causal dependencies are independently understood, identifying which differences are irrelevant can provide a principled basis for analogical inference. However, it faces a fundamental difficulty in the present context. Establishing causal irrelevance in terms of difference-making presupposes access to independent information about which factors are causally relevant in the domain under consideration. In cases involving unmanipulable variables, such information is typically unavailable. As a result, the irrelevance condition cannot be applied in an epistemically independent way, since it relies on the very causal knowledge that the analogical inference is meant to support.

This suggests that what is needed is a condition that does not presuppose causal knowledge of the target domain, but nonetheless imposes a substantive constraint on analogical inference. The condition proposed here is a requirement of mechanism sharing. For an analogy to provide warrant for extending a causal claim, there must be independent grounds for treating the source and target domains as sharing relevant aspects of the mechanism responsible for producing the outcome in the source domain. By "independent grounds", I mean considerations that are not generated by the analogical inference itself, but instead stem from background knowledge, theoretical understanding, or empirical evidence concerning the processes responsible for the outcome. They are therefore independent of the analogical projection, though not necessarily independent of all causal concepts. Here, "causal concepts" refers to general theoretical knowledge about mechanisms and causal processes, or to causal relations established in source domains, rather than to prior knowledge that the projected causal relation already obtains in the target domain. These independent grounds determine which similarities count as relevant for the purposes of analogical projection.

This requirement rules out projections based solely on superficial or formal similarities, without support from features connected to the production of the outcome. It does not assume that the relevant mechanisms are already established in the target domain, but only that there are independent grounds for treating relevant aspects of those mechanisms as shared across domains. Here, relevance is not understood in terms of prior knowledge of causal structure or difference-making, but in terms of independently grounded connections to the mechanism producing the outcome.

This condition occupies a middle position between existing accounts. Unlike approaches that appeal to causal irrelevance in terms of difference-making, it does not require prior knowledge that the projected causal relation already obtains in the target domain. At the same time, it is more restrictive than accounts that rely solely on articulated associations in the source domain, since it requires that such associations be supported by independent grounds connecting them to the mechanism responsible for the outcome.

4.2 The Limits of Analogical Extension in MP

MP attempt to extend causal claims from manipulable to unmanipulable situations by appealing to a principle of analogical reasoning. According to their weakened agency account, even when agents cannot manipulate an event at will, causal claims can still be made by drawing an analogy with manipulable cases. The key idea is that certain intrinsic features present in manipulable situations also obtain, or are closely similar, in unmanipulable ones. These shared features are taken to support the extension of causal claims beyond the domain of manipulation.

In manipulable situations, such intrinsic features I^* are taken to support a means–end relation between A^* and B^* in virtue of standing in a structural relation K. It is on this basis that A^* is regarded as causing B^*. The analogical strategy then proceeds by extending this causal claim to unmanipulable situations. The similarity relation that is meant to ground this extension—captured here by K—does the crucial work. By appeal to this similarity, MP hold that where the intrinsic features I of an unmanipulable situation S are identical with, or closely similar to, the intrinsic features I^* of a manipulable situation S^* in respect of K, the causal relation between A and B in S can be extended from the manipulable case.

When this strategy is assessed against the standard of analogical warrant set out in Section 4.1, a difficulty emerges. According to that standard, an analogy can justify extending a causal claim only if there are independent grounds for treating the source and target domains as sharing aspects of the underlying mechanism responsible for producing the outcome in the source domain. It is not enough that two situations resemble each other in some respect; the shared features must bear on those mechanisms. MP's appeal to analogy does not meet this requirement. Although K is required to support the extension of the causal claim, MP provide no positive account of what K consists in, beyond the insistence that it must be non-causal. This requirement can be understood in light of the reductionist ambition of the account, which seeks to analyse causation without presupposing causal notions. As a result, it remains unclear whether similarity in respect of K provides independent

grounds, since no account is given of how *K* connects to the features in virtue of which the means–end relation obtains. Without such an account, the appeal to similarity cannot explain why resemblance in respect of *K* should support extending the causal claim to unmanipulable cases.

This problem is sharpened by Woodward's criticisms. Woodward raises two related objections. First, he argues that MP cannot explain how intrinsic features *I* and *I** are relevantly similar without invoking causal concepts. In cases such as scientific models of earthquakes, resemblance appears to matter only insofar as the same causal processes are operative in both the model and the real system (Woodward 2003, p. 125). If similarity in respect of *K* is understood in this way, then the extension of causation rests on causal similarity, and the reductionist ambition of MP's account collapses. The appeal to analogy no longer provides a basis for extending causal claims in the way required by their account.

Second, Woodward considers an alternative interpretation. Perhaps MP intend to appeal to a certain kind of relation among intrinsic, non-causal features that agents exploit when bringing about *B** by bringing about *A** in manipulable cases, and that can also obtain in unmanipulable situations. Woodward's suggestion is that such a relation may be intrinsic and exist independently of agents, and that causation in unmanipulable cases would not be constitutively tied to agency (Woodward 2003, pp. 125–126). In that case, the distinctive motivation of the agency-based account is undermined.

Both objections point to the same underlying difficulty. MP need a notion of similarity *K* that is non-causal and strong enough to support the extension of causal claims. Woodward's first objection shows that if *K* is characterised in causal terms, the account becomes circular. MP can accept this point. However, Woodward's argument does not show that a non-causal characterisation of *K* is impossible. It shows only that the tempting move to cash out *K* as a shared causal process cannot work. The deeper problem lies elsewhere. Although MP insist that *K* is non-causal, they do not explain what such a non-causal similarity would be, nor why it should support extending causation. Agency can be built into their weakened definition of causation[8], but the crucial intermediate step specifying what similarity in respect of *K* amounts to is left undeveloped, let alone an account of why it provides independent grounds for treating the similarity as tracking the underlying mechanism. As a result, the appeal to analogy fails to meet the standard of analogical warrant set out in Section 4.1. One might attempt to repair the view by offering a more substantive proposal for *K*, for example by appealing to some structural relation holding among features that remain stable across manipulable and unmanipulable cases, but MP do not develop such a proposal. Accordingly, their appeal to analogy provides no warrant for extending causal claims beyond the domain of manipulation, and the problem of unmanipulable causation that the account is meant to address remains unresolved.

[8] On the weakened agency account, the definition of causation for unmanipulable situations already involves agency, insofar as it appeals to what a free agent would freely bring about in a suitable manipulable analogue.

4.3 The Limits of Analogical Extension in Woodward

Woodward's interventionist framework aims to extend causal analysis beyond cases of direct manipulation, allowing causal claims about unmanipulable phenomena without reducing causation to actual human agency. As argued in the previous section, this extension beyond direct manipulation amounts to a form of analogical reasoning. It proceeds at two levels: first, in the construction of causal models; and second, in the evaluation of counterfactuals within those models.

The question, then, is whether this analogical extension satisfies the condition for analogical warrant articulated in Section 4.1. On that condition, an analogy provides warrant for extending a causal claim only if there are independent grounds for treating the source and target domains as sharing aspects of the underlying mechanism responsible for producing the outcome in the source domain. The issue is therefore not merely whether the two domains exhibit similarities, but whether those similarities are supported by considerations bearing on the mechanisms that produce the outcome.

The interventionist extension appears, at least initially, to satisfy the requirement of mechanism sharing. In many cases, similar types of processes are taken to underlie both manipulable and unmanipulable phenomena. This suggests that the source and target domains share aspects of the mechanism. However, closer examination shows that this appearance depends on two assumptions, neither of which satisfies the condition articulated in Section 4.1.

The first concerns how this apparent sharing of processes is established at the level of model construction. The structural models used to represent unmanipulable systems are not derived directly from those systems but are instead adapted from modelling practices developed in manipulable contexts. The attribution of a shared mechanism therefore already involves an analogical step: the model is taken to apply to the target domain because of its success in structurally similar, experimentally tractable cases. This transfer carries with it the assumptions built into the modelling framework, most notably modularity.

This difficulty becomes clearer once the status of modularity itself is examined more closely. As discussed above,

modularity is not an independently established feature of causal systems, but a structural assumption that enables the representation of interventions as local modifications. It arises from the way causal mechanisms are represented as autonomous within the model. Its application to unmanipulable domains is therefore not supported by independent evidence about the mechanisms at work in those domains. Rather, it is imported from the source domain as part of the modelling framework itself. The appeal to a shared mechanism rests on an assumption whose validity in the target domain is precisely what is at issue.

It is precisely this feature that has been the target of sustained criticism. Nancy Cartwright argues that the kind of modular organisation presupposed by interventionist models cannot simply be taken for granted in realistic systems (Cartwright 2001, pp. 70–73; Cartwright 2007, pp. 80–96). In many mechanical and social contexts, causal relations are often tightly coupled, such that intervening on one component typically disrupts others. Her objection is not merely that modularity fails in practice, but that the very idea of representing interventions as precise, localised modifications rests on an assumption that lacks general applicability.

Related concerns have been raised in econometrics. James Heckman (2005, p. 44) emphasises that constraining one equation in a system affects the entire set of internal variables and may alter the parameters of other equations, thereby violating parameter stability. Building on this point, James Heckman and Vytlacil (2007, p. 4843) argue that policy interventions cannot typically be implemented in isolation, since they generate a range of behavioural and equilibrium responses. Economic systems, on their view, exhibit a high degree of interdependence and are therefore nonmodular in the relevant sense. As a result, interventions directed at one causal relation tend to induce adjustments throughout the system, undermining the assumption that causal mechanisms can be modified independently. These critiques reinforce the point that modularity is not a generally established feature of the systems to which interventionist models are applied.

A natural response, developed by Pearl in the second edition of *Causality*, is to emphasise that the autonomy required for interventionist reasoning should be understood at the level of symbolic manipulation rather than physical intervention (Pearl 2009, pp. 374–375). Understood in this way, modularity is a feature of representation rather than of the underlying system. However, this reply does not fully resolve the difficulty. Even if modularity is understood symbolically, the use of interventionist counterfactuals still presupposes that it is meaningful to specify what it would be to alter one mechanism while holding the relevant background fixed. In domains where causal influence is distributed across highly interconnected pathways such as social, psychological or large-scale physical systems, it is unclear whether such a specification can be given. The problem, therefore, is not merely empirical. Rather, the difficulty is that the application of this assumption lacks independent grounds in precisely those contexts in which it is most needed.

The second assumption concerns the role of empirical support in unmanipulable contexts. Even where direct manipulation is not possible, it may be argued that models can be tested against observable and repeatable phenomena, thereby providing indirect support for their causal claims. In many geological cases, for example, patterns of seismic activity, stress accumulation, and fault behaviour can be observed over time, and models can be evaluated in light of such data. However, this form of support has clear limitations. Observed correlations and regularities underdetermine the underlying causal structure, since they are compatible with multiple causal explanations. What is missing is the intervention-based variation that, in manipulable contexts, allows such alternatives to be discriminated.

This limitation becomes more severe in domains where even such forms of empirical support are unavailable. However, as Woodward also acknowledges, the difficulty in such cases is not merely empirical. In large-scale systems, such as when the system under consideration is the universe as a whole, the interventionist notion no longer has a clear and natural application: there is no external source of intervention, and the distinction between the system and what intervenes on it breaks down. As a result, it becomes unclear how to formulate the relevant causal claims and their conditions of applicability in interventionist terms (Woodward 2009, pp. 256–258). From the perspective developed here, this conceptual limitation has a further epistemic consequence. In domains that are not only unmanipulable but also non-repeatable and only partially observable, the available evidence underdetermines the underlying causal structure and does not establish which underlying mechanisms are responsible for the outcome. As a result, even if interventionist models can be formulated, they do not provide independent grounds for treating the source and target domains as sharing relevant aspects of the mechanism. The appeal to empirical adequacy therefore fails to supply the kind of warrant required for extending causal claims to such cases.

Taken together, these considerations show that the appearance of mechanism sharing is sustained by two forms of support that fall short of the requirement articulated in Section 4.1. The first relies on the transfer of a modelling framework that already presupposes modularity, without independent grounds for attributing that structure to the target system. The second relies on observational adequacy in contexts where the available data do not suffice to identify the underlying mechanisms in a way that is comparable to

manipulable cases. In neither case is there independent support for treating the source and target domains as sharing the relevant aspects of the mechanism, and the extension therefore depends on similarity judgments that are not internally constrained by the framework.

Accordingly, the analogical extension at the core of the interventionist framework does not, by itself, meet the condition for analogical warrant. While the framework provides a coherent way of representing and reasoning about unmanipulable phenomena, its extension from manipulable contexts depends on assumptions whose applicability to the target domain is not independently established. Where independent grounds for such mechanism sharing can be provided, for example, through well-supported claims about modularity or evidence bearing on key aspects of the mechanism, the condition for analogical warrant may be satisfied. In their absence, however, the extension of causal claims in such cases remains insufficiently warranted.

5 Laws of Nature as a Possible Source of Warrant

This section examines whether appeals to laws of nature can supply the warrant required to extend causal claims beyond manipulable contexts. It considers two prominent strategies, the agency-based appeal to law-conformity and the interventionist appeal to law-governed counterfactuals, and argues that neither satisfies the condition for analogical warrant.

5.1 Law-Conformity in the Agency Theory (MP)

As argued earlier, MP's response to the problem of unmanipulable causation relies on an appeal to analogical reasoning. Causal claims concerning unmanipulable situations are licensed by their resemblance to paradigmatic manipulable cases. However, their account leaves the relevant similarity relation *K* largely unspecified. In order to preserve their reductionist ambition, MP require that *K* be characterised in non-causal terms. However, satisfying this requirement is not sufficient for an appeal to analogy to be adequately warranted. In addition, *K* must satisfy the condition set out in Section 4.1. In particular, it must provide independent grounds for treating manipulable and unmanipulable situations as sharing relevant aspects of the mechanism responsible for producing the outcome. This requires that *K* be characterised in a way that does not presuppose that the projected causal relation already obtains in the target domain, while also being supported by independent grounds connecting it to features involved in producing the outcome in the source domain and plausibly present in the target domain.

One natural thought, given MP's requirement that *K* be non-causal, is that *K* might consist in conformity to the laws of nature. Since both manipulable and unmanipulable phenomena are ordinarily taken to fall under the same laws, law-conformity initially appears to offer a common, non-causal basis for extending causal claims, and thus presents itself as a promising candidate under MP's constraints. This suggestion is further motivated by Peter Menzies's account of laws, which is continuous with the agency-based approach defended by MP. On Menzies's view, laws of nature are regularities that are resilient under actual and hypothetical experimentation, where experimentally possibility is defined in terms of what suitably equipped agents could freely bring about (Menzies 1993, pp. 206–210).

To assess this proposal, however, we must ask what it is for two situations to conform to the same law in a way that is relevant to causation. In a weak sense, everything conforms to the laws of nature insofar as nothing violates them. Understood in this way, law-conformity is trivial and cannot distinguish causal from non-causal relations. A stronger construal would require that the intrinsic features in question instantiate the same law in the same way, occupying the same role within the relevant law-governed structure.

Consider the case of electric charge. Both positively and negatively charged particles are governed by Coulomb's law, yet they figure very differently as bases for causal claims. Mere subsumption under the same law does not determine which intrinsic features are causally efficacious or how they enter means–end relations. To avoid triviality, conformity to law would therefore have to be specified at the level of the particular role played by an intrinsic feature within that law. At this point, however, the proposal faces a dilemma: either this role is characterised in causal terms, undermining the non-causal requirement on *K*, or it remains too coarse-grained to distinguish differences, such as the sign of charge, that correspond to distinct roles within the same law.

A further difficulty concerns the problem of spurious correlations. Agency-based theories are motivated precisely by the thought that law-governed regularities are insufficient to distinguish genuine causal relations from spurious correlations arising from common causes or background conditions. Although Menzies's account of laws addresses this issue in manipulable contexts by appealing to experimental resilience, this resource is unavailable in unmanipulable situations, where no such testing is possible. Moreover, his definition of lawhood is itself restricted to experimentally possible courses of events, and the theory offers no independent account of how such laws may be extended to unmanipulable cases (Menzies 1993, pp. 195–225).

For these reasons, appealing to natural laws fails in supplying the required warrant for MP's analogical strategy. Law-conformity is either too weak to distinguish causally relevant features, or it must be specified in a more fine-grained way that undermines its role as a non-causal basis. Moreover, insofar as it relies on experimental resilience, it depends on resources available only in manipulable contexts. Consequently, this proposal fails to meet the condition on analogical warrant articulated in Section 4.1, and the problem of unmanipulable causation remains unresolved.

5.2 Law-Governed Theory in Interventionist Theory (Woodward)

A natural response to the difficulty identified in the previous sections is to appeal to law-governed theory. If intervention is not available, it may be argued that well-confirmed physical theories can supply the counterfactual structure required for causal judgment. On this view, even in cases such as the relation between the position of the moon and tidal behaviour, causal claims can be evaluated on the basis of theoretical knowledge, without reliance on direct intervention.

This strategy is explicitly endorsed by Woodward. In the moon-tide case, Newtonian gravitational theory is taken to provide reliable information about what would happen under counterfactual circumstances in which the position of the moon differs from its actual value. On this basis, Woodward argues that causal counterfactuals can be evaluated even in the absence of intervention (Woodward 2003, pp. 129–131). This appeal to law-governed theory succeeds in an important respect. It explains how counterfactual claims can be evaluated even where no intervention is possible, and hence why causal claims about unmanipulable systems are not automatically ruled out.

However, this appeal to law-governed theory does not address a more fundamental issue. In manipulable cases, the distinction between causal relations and mere regularities is secured by dependence that tracks changes under intervention rather than by law-governed considerations alone. In unmanipulable contexts, however, where such intervention is unavailable, law-governed theory must bear this burden, but it does not suffice to perform this discriminating function. In the moon-tide case, law-governed theory may tell us what would happen if the moon were in a different position, but it does not explain why the relation between lunar position and tides should count as causal rather than as a mere regularity. More importantly, it does not provide a basis for distinguishing between competing causal structures that are compatible with the same observable regularities.

The issue, therefore, is not merely whether counterfactuals can be evaluated, but whether there are grounds for treating the resulting dependencies as genuinely causal. This point bears directly on the condition set out in Section 4.1, which requires that any extension of causal claims be supported by independent grounds for treating the source and target domains as sharing aspects of the mechanism. However, the appeal to law-governed theory does not provide such independent grounds.

6 Possible Objections

A possible objection appeals to Price's later work, in which he abandons the reductive ambitions of the agency-based account (Price 2017, pp. 75–80). One might suggest that if the requirement that K be non-causal is abandoned, then an appeal to conformity with laws could supply a basis for projection without facing the earlier difficulty. This proposal can be understood as a strategy for addressing the problem of unmanipulable causation within the agency framework.

This strategy faces a dilemma. If laws are construed as fully objective features independent of agents, the proposal departs from the agency framework. It then aligns more closely with a regularity or law-based account of causation. In that case, it is no longer an extension of the agency theory but a replacement. If laws are instead understood along Menzies's lines, in terms of resilience under actual and hypothetical experimentation, the proposal remains internal to the agency framework. On this construal, the earlier difficulty concerning non-causal characterisation may be avoided. However, the appeal to experimental resilience depends on resources available only in manipulable contexts. It relies on the possibility of testing and confirming regularities through intervention. In unmanipulable cases, no such resources are available. The appeal to law-conformity therefore fails to provide independent grounds for extending causal claims beyond manipulable contexts. As a result, it does not secure the projection from manipulable to unmanipulable cases. The underlying difficulty remains unresolved, since neither option provides independent grounds for extending causal claims to unmanipulable cases.

A further objection, directed at the interventionist framework, holds that the constraints required to extend causal claims to unmanipulable cases need not derive from analogical reasoning, but may instead be supplied by independent background theories. On this view, in cases such as seismic activity, our confidence that modularity conditions are satisfied does not rest on analogy with manipulable models, but on what geophysical theory tells us about the separability of the relevant mechanisms. This objection has the advantage of reflecting scientific practice: in extending causal claims beyond direct manipulation, scientists typically rely on theory rather than explicit analogical arguments.

However, this objection does not remove the need for a further step. It leaves open whether the structural assumptions of the theory can be legitimately carried over to the unmanipulable case. Even if a background theory is well supported, applying it to an unmanipulable case requires assuming that its key structural claims, such as modularity, continue to hold in that case. The theory itself does not secure this, since its evidential support is drawn from cases that may differ in relevant respects from the target system. To set those differences aside, one needs independent grounds for thinking that the source and target domains share relevant aspects of the mechanism. Without such grounds, it remains unclear whether the theoretical structure carries over. In this sense, the appeal to background theory does not avoid analogical reasoning but relies on it: extending a theory to a new domain depends on a judgement of relevant similarity. Where this judgement is not supported at the level of mechanisms, the extension lacks warrant.

7 Conclusion

This paper has argued that manipulationist approaches to causation face a structural difficulty in accounting for unmanipulable causal relations. Given their core commitment to relating causation to manipulation, such accounts must explain how causal claims are adequately extended beyond direct manipulation. The problem of unmanipulable causation should therefore be treated as a central challenge rather than a marginal complication for manipulationist approaches to causation. As the preceding analysis has shown, both agency theory and interventionist theory rely on forms of analogical reasoning, though neither, as it stands, satisfies the condition on analogical warrant.

Importantly, the difficulty is not merely theoretical. Many causal claims in scientific and social inquiry concern factors that resist direct manipulation, and extending causal judgments to such cases cannot be treated as methodologically unproblematic. Such extensions require an explicit warrant, which in turn depends on independent grounds for attributing shared aspects of the underlying mechanisms. This helps to explain why interventionist approaches encounter persistent difficulties when applied beyond controlled experimental settings. Where such grounds can be provided, for example through well-supported claims about modularity, evidence bearing on key features of the underlying processes, robustness across heterogeneous evidence, or successful model validation across domains, the condition for analogical warrant may be satisfied. Clarifying how such grounds can be established in unmanipulable cases remains an open task, but recognising this requirement is essential for understanding the scope and limits of manipulationist approaches to causation.

Acknowledgements I am grateful to Alex Broadbent, Nancy Cartwright, and Sarah Wieten for helpful comments and critical feedback on earlier versions of this paper. I am also grateful to the anonymous reviewers for their thoughtful and constructive suggestions, which significantly improved the paper. Finally, I would like to thank Guilin Jiao and Qiuxiang Wen for their support, and C. He for her encouragement during the process. Any remaining errors are my own.

Author Contributions The author solely conducted the research and wrote the manuscript.

Funding This work was supported by the China Scholarship Council (CSC) under Scholarship No. 202306190071.

Data Availability Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations

Consent for Publication Not applicable.

Competing Interests The author has no relevant financial or non-financial interests to disclose.

Consent to Participate Not applicable.

Ethics Approval This study does not involve human participants or animals and therefore does not require ethical approval.

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

Bartha P (2010) By Parallel reasoning: the construction and evaluation of analogical arguments. Oxford University Press

Cartwright N (2001) Modularity: it can—and generally does—fail. In: Galavotti MC, Suppes P, Costantini D (eds) Stochastic Causality. CSLI, pp. 80–96.

Cartwright N (2007) Hunting causes and using them: approaches in philosophy and economics. Cambridge University Press, pp. 80–131.

Collingwood RG (1940) An essay on metaphysics. Clarendon

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

Heckman JJ (2005) The scientific model of causality. Sociol Methodol 35:1–97

Heckman JJ, Vytlacil EJ (2007) Econometric evaluation of social programs, part i: causal models, structural models and econometric policy evaluation. In: Heckman JJ, Leamer EE (eds) Handbook of Econometrics, vol 6B. Elsevier, pp 4779–4874

Hesse MB (1963) Models and analogies in science. University of Notre Dame

Langton R, Lewis D (1998) Defining intrinsic. Philos Stud 92(3):333–345

Lewis D (1979) Counterfactual dependence and time's arrow. Noûs 13(4):455–476

Lewis D (1973) Causation. J Philos 70(17):556–567

Menzies P (1993) Laws of nature, modality and humean supervenience. In: Bacon J, Campbell K, Reinhardt L (eds) Ontology, Causality, and Mind: Essays in Honour of D. M. Armstrong. Cambridge University Press, pp 195–225

Menzies P, Price H (1993) Causation as a secondary quality. Br J Philos Sci 44(2):187–203

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

Pearl J (2009) Causality: models, reasoning, and inference, 2nd edn. Cambridge University Press

Pietsch W (2019) A causal approach to analogy. J Gen Philos Sci 50(4):489–520

Price H (2017) Causation, intervention and agency—woodward on menzies and price. In: Beebee H, Hitchcock C, Price H (eds) Making a Difference: Essays on the Philosophy of Causation. Oxford University Press, pp 73–98

Spirtes P, Glymour C, Scheines R (1993) Causation, prediction, and search. Springer

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

Woodward J (2009) Agency and interventionist theories of causation. In: Beebee H, Hitchcock C, Menzies PC (eds) The Oxford Handbook of Causation. Oxford University Press, pp 234–264

Woodward J (2023) Causation and manipulability. In: Zalta EN, Nodelman U (eds) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University. Summer 2023 edn.

Woodward J, Hitchcock C (2003) Explanatory generalizations, Part I: A Counterfactual Account. Noûs 37(1):1–24

von Wright GH (1971) Explanation and understanding. Routledge & Kegan Paul

von Wright GH (1975) Causality and determinism. Columbia University

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