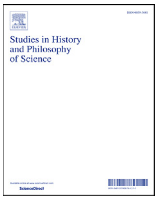
Studies in History and Philosophy of Science 89 (2021) 74–83

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa



Intervening on time derivatives

Toby Friend

Department of Philosophy, University Bristol, Cotham House, Bristol, BS6 6JL, UK

ARTICLE INFO

Keywords:
Interventionism
Testing interventions
Control interventions
Interventionist theory of causation
Time derivatives

ABSTRACT

Interventionism analyses causal influence in terms of correlations of changes under a distribution of interventions. But the correspondence between correlated changes and causal influence is not obvious. I probe its plausibility with a problem-case involving variables related as time derivative (velocity) to integral (position), such that the latter variable must change given an intervention on the former unless dependencies are introduced among the testing and controlling interventions. Under the orthodox criteria such interventions will fail to be appropriate for causal analysis. I consider various alternatives, including permitting control interventions to be chancy, restricting the available models and mitigating variation of off-path variables. None of these work. I then present a fourth suggestion which modifies the interventionist criteria in order to permit interventions which can influence other variables than just their own targets. The correspondence between correlated changes and causal influence can thereby saved when dependencies are introduced among such interventions. This modification and the required dependencies, I argue, are perfectly in line with practice and may also assist in a wider class of cases.

1. Introduction

Interventionism is based on the plausible-sounding idea that ‘causal relationships are relationships that are potentially exploitable for the purposes of manipulation and control’ (Woodward, 2009, p. 234). More contentiously, interventionists take causal relationships to be *definitively* so: where an effect is identified as some potential outcome, its causes are just those very mechanisms, levers, handles, devices or variables by which that outcome can be brought about (Woodward, 2016a). As defenders of the view have pointed out, plausibility for this view comes from an appreciation of the way causal information is empirically accessible (i.e. via experimental intervention), practically useful (for control and manipulation of the environment) and conceptually acquired by organisms as they learn to exploit and navigate their environment (Woodward, 2003, pp. 25–38).

By far the most advanced Interventionist *theory* of causation has been developed by Woodward (2003, 2008, 2015) and put to work by many others for various philosophical purposes (e.g. Weslake, forthcoming; Shapiro & Sober, 2007; Frisch, 2014; Wilson, 2014; Ross, 2020). At its basis is the idea (to be presented more precisely below) that causal relations relate variables such that if one were to intervene on the cause variable by changing (‘wiggling’) it in some way, the effect variable would also (or be more likely to) change. Therefore, the central notion of a causal relationship being one exploitable for the purposes of manipulation or control at the root of the interventionist intuition is captured by the theory in terms of a *correspondence of changes* in the causal relata.

The inference between causal relationships and correspondence of changes so central to interventionism suggests that where the causal profile of a variable and its time derivative come apart the theory will struggle to provide that result.[1] Here I aim to probe just this issue by comparing the respective influence of a car's velocity and position on a speed gun and radar. I will also later give examples to show that the issue is not restricted only to variable-pairs related by differentiation.

Here's the plan. In §2 I introduce the ‘Orthodox Interventionist Theory’ in more detail. In §3 I will introduce two examples in which a car travels respectively towards a speed gun and a radar making explicit the differences in causal influence which interventionist theory should acknowledge between the car's velocity and its position on the speed gun and radar's readings. In §4 I will suggest that when all the variables of interest in the example are included in a causal model, the interventionist criteria of the orthodox theory will be violated under any attempt to intervene on velocity while controlling for position. This means the theory will not achieve entirely the correct causal verdicts in either case. I'll then consider three unsuccessful responses on behalf of the

E-mail address: toby.friend@bristol.ac.uk.

[1] This observation has already been made by Weber (2016) in the context of Pearl's interventionist (2000) *do*-calculus. Anderson (2020) objects that Weber's concerns are misguided, although I suspect these objections will not be sustained (see footnote 8). Nevertheless, I focus here on the issue specifically for Woodward's interventionist theory, which in contrast with Pearl's theory is an attempt to give an *analysis* of causation in interventionist terms.

https://doi.org/10.1016/j.shpsa.2021.07.005
Received 23 February 2021; Received in revised form 6 July 2021
Available online 10 August 2021

interventionist. The first (§5) questions the claim that the interventionist criteria are invalidated in the ways suggested by proposing the existence of 'chancy control' interventions. The second (§6) proposes to restrict causal models to those in which position and velocity do not both feature, thereby negating the need to provide a control intervention on position. The third (§7) proposes under an 'Extended Interventionist Theory' to permit such models but to mitigate the varying of position alongside velocity rather than requiring it to be controlled for. None of these responses are sufficient, or so I claim. They either fail to overcome some of the initial worries, and so continue to imply that velocity is not a cause of the speed gun reading, or they make the further error of claiming that position is a cause of the speed gun reading. Only a fourth response, presented in §8 and which proposes a new modification of the interventionist criteria, avoids all these troubles. As I will point out, this 'Modified Interventionist Theory' does so not by avoiding the basic intuition that causation is a correlation of changes under intervention but rather by *emphasising* the sufficiency of this point, and thereby relaxing other conditions in the interventionist criteria such as the independence of testing and control interventions. In proposing the modified theory I'll draw attention to its wider applicability beyond problem cases involving time derivatives. §9 then discusses further corollaries and limitations, specifically concerning (respectively) constitutively related variables and causal exclusion. §10 concludes.

2. Orthodox interventionist theory

Orthodox Interventionist Theory (OIT), as I will mean by it, is a view about what relations of causal influence among variables are. That is, it is an analysis of causal influence rather than a prescription for causal discovery or prediction. In its uncompressed form, defended at length in Woodward (2003), causal relations are divided between 'direct' and 'contributing' relations, both of which are to be co-defined with (rather than reduced to) interventions, in the following way.

> **M.** A necessary and sufficient condition for X to be a (type-level) *direct cause* of Y with respect to a variable set $\mathbf{V}$ is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables Z_i in $\mathbf{V}$. A necessary and sufficient condition for X to be a (type-level) *contributing cause* of Y with respect to variable set $\mathbf{V}$ is that (i) there be a directed path from X to Y such that each link in this path is a direct causal relationship […], and that (ii) there be some intervention on X that will change Y when all other variables in $\mathbf{V}$ that are not on this path are fixed at some value. (2003, 59)

Accordingly, X is a cause of Y in a model if and only if X is either a direct or contributing cause of Y in that model. Evidently, **M** confines itself only to characterising causal influence within a model. But interventionists will take it that a relationship of causal influence exists between two variables *simpliciter* if and only if there exists some model in which there is a relationship of causal influence between the variables.

As can be seen from the formulation of **M**, an intervention on some variable X is always *with respect to another variable* Y, and so relative to whatever causal relationship is being tested for. More specifically, to be appropriate for establishing causal relationships, OIT requires that interventions satisfy **IV**.

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

If an attempted intervention satisfies these criteria we may call it 'successful'; if not, 'failed'. I'll refer to the variable an intervention is performed on with respect to another, the intervention's 'target variable'.

As should be clear from **M**, interventions come in two broad varieties. *Testing* interventions are those whose target variables are those whose causal influence on the effect is to be assessed by changing their value. *Control* interventions are those whose target variables are those whose causal influence on the effect is to be controlled for by being held fixed at some value. Nothing in the interventionist criteria marks a distinction between the two, but it will crucial to bear the difference in mind for what follows.

In sum, OIT ties the causal status of a relationship between two variables to the possible results of interventions, so defined; as Woodward remarks, 'no causal difference without a difference in manipulability relations, and no difference in manipulability relations without a causal difference' (Ibid. 61). The explicitly causal requirements in **IV** mean that the interventionist approach to causation is unavoidably non-reductive. But the causal claims in **IV** are nevertheless different to those defined in **M** in that they are not model-relative. So, an intervention I on X with respect to Y can therefore fail to be successful if *there is no model whatsoever* for which the intervention I, treated as a variable in the model, is a cause of X (i.e. a failure of I1). Similarly, it can fail even if *there is only one model* in which I causes Y via a route that doesn't go through X (a failure of I3). It is also a straightforward consequence of OIT that there can be no direct causal relationships between the variables A and B in a model in which some off-path variable(s) $C, \ldots$ cannot be held fixed. Moreover, if two variables are not directly causally related but cannot be held fixed independently from each other, then there can be no contributing causal relations in that model which go through those two variables.

OIT has many merits. Not least are the fact that it bears close affinity and inspiration from actual experimental practice while simultaneously avoiding the apparent pitfalls of analysis in terms of statistical dependence. Statistical treatments of causation typically rely on conditions (causal Markov and faithfulness) for which it is at best contentious that they hold universally. As a consequence statistical approaches are typically restricted to defining algorithms for causal discovery and prediction rather than analysing what causal relations are (e.g. Spirtes, Glymour, & Scheines, 2000; although see also; Schurz & Gebharter, 2016). By contrast, OIT apparently avoids commitment to the problematic conditions (see, e.g., Woodward, 2003, pp. 64–5, 108)—though of course it is consistent with them. What enables the theory to do so is that it defines causal relations in terms of a comparison between a single instance of the un-intervened-on values of variables in the model with a (potentially hypothetical) 'one-shot' array of testing and control interventions which disrupt the extant causal relations.[2]

This attributed benefit of OIT is also what motivates a concern. By moving from a measure of statistical dependencies to an analysis in terms of changes under intervention, the interventionist theory relies on a correspondence between causal facts and facts about hypothetical changes in the variables of a *single system*. But are further causal influences of changes really relevant to the causal status of a pair of

[2] Statistical procedures may utilise interventions, of course, but they are compelled to treat them just as further variables in the model to be conditioned on.

(potentially static) variables? The cases to be presented in the next section aims to probe this issue.

3. Two problem-cases

The velocity of a car in the direction of an observer can be measured with a speed gun. The gun sends an a radiowave at cms^{-1} with frequency f in the direction of the car which is then reflected and received back by the gun at a modified frequency f'. This modified frequency depends directly on the (time-indexed) velocity V of the car in the direction of the speed gun according to the relationship

$$V = \frac{f' - f}{f}\frac{c}{2}. \tag{1}$$

The speed gun then outputs a reading corresponding to that velocity by comparing the frequencies sent and received. For simplicity's sake, let's assume the reading is a binary variable R whose values (*above*, *below*) correspond to whether it measures the car to have a speed above the legal speed limit or below it. Let's also assume the car is set in motion at V from a position a and is constrained by a straight road to move only along the direction between itself and the speed gun, so that the reading reflects the car's actual speed (see Fig. 1).

I take it that the velocity V of the car in the direction of the speed gun is a causal influence of the reading R on the speed gun. If justification is needed, we may consider the following points. First, the speed gun is a device explicitly manufactured for *measuring speed*, a trivial function of velocity, where the measurement is clearly not achieved by virtue of V and R being the effects of common causes, R causing V, or some logical or constitutive relationship between the variables. Second, R is counterfactually dependent on V; a difference in whether the speed gun gives a reading *above* or *below* is counterfactually dependent on whether the car's velocity in the direction of the speed gun is in fact above the speed limit or below it.[3]

By contrast, consider the relevance to R of the car's *position* X along the axis between the car and speed gun. This, I will also take it, is *not* a causal relationship. First, the speed gun does not measure position. It would measure position if the device were a radar instead, by calculating the time for a signal sent out to be reflected and received back (see below). But a speed gun doesn't do this. Second, the reading R is not counterfactually dependent on X. A difference in whether the car is in position x or y (see Fig. 1) will not affect the reading, and this lack of dependency is clearly not a consequence of causal redundancy. Of course, the speed gun is influenced by *changes* in position, since a speed gun exploits the fact that the car will change position as it reflects the radiowave—the same phenomenon as witnessed in the Doppler effect.[4] We'll later see that this tight relationship between change in position and velocity is precisely what makes trouble for interventionism, since it means that X and V are not independently manipulable (more of that later!). But this should not detract from the point that X itself, i.e. a measure for how far the car is away from the speed gun, is not an influence of R, since the fractional change in position involved in taking the speed gun reading is irrelevant to how far away the change occurs.

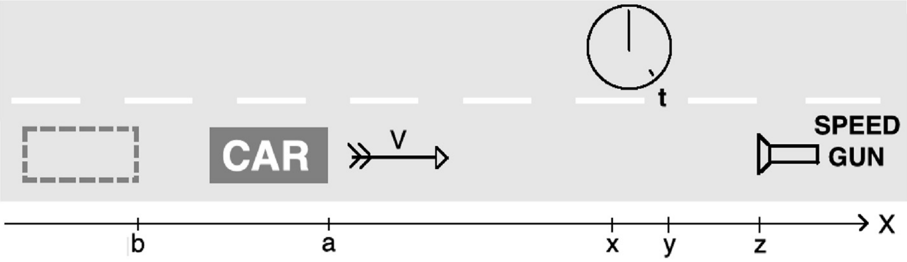


Fig. 1. Speed gun measuring the speed of an approaching car at t

To further emphasise the lack of causal influence of X on R, we can consider a different circumstance in which position clearly *is* relevant to a device's reading. For instance, we might modify the speed gun to work also like a radar, which calculates the car's velocity in the above way *and* also its position X according to,

$$X = z - c\Delta t/2, \tag{2}$$

where z is the location of the speed gun and Δt is the time taken for the radiowave to be emitted and returned. The device, let's stipulate, then provides a reading R of either of two values (*threat*, *no-threat*) depending on whether the value of the quotient $V/(z-X)$ is above or below some threshold (the chance of ‘threat’ is increased if the car is nearby or is moving fast).

Let's refer to each of these cases respectively as ‘SPEED GUN’ and ‘RADAR’. In RADAR the position of the car *is* an influence on R, since R is a function of position. In SPEED GUN it is not, because R is not a function of position. Perhaps in either case, we might say that the fact the car is positioned on the road at all is a cause of the fact that there is a reading R at all, but I take it that this is not causal influence in the manner that interventionists typically aim to capture, and is rather a condition of the set-up.[5]

With this contrast in mind, the causal profiles of V and X in SPEED GUN are clearly different, since one is and the other is not a causal influence of R (see the causal graph in Fig. 2a).[6] But there is a sense too in which the causal profiles of X and V are different in RADAR. This is because even though both V and X are causes of R, they are causal influences *along different routes*: X doesn't influence R by influencing V and V doesn't influence R by influencing X; they both influence R directly (Fig. 2b). If there's any doubt about that, one can simply reflect on the fact that the only difference in X's influence on R between the two cases comes from a difference in the devices' mechanisms involved in SPEED GUN and RADAR; nothing to do with the car's behaviour changes from one case to the other.

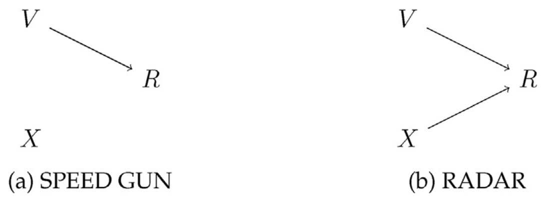
The contrast in the causal profiles of position and velocity in both SPEED GUN and RADAR is something which the interventionist theory of causation should hope to make sense of. As I'll now explain, however, the variables' close relationship raises a number of problems.

[3] Of course, this dependency may itself be dependent on the absence of back-up potential causes, as there would be in cases of pre-emption.

[4] Thanks to an anonymous referee for encouraging me to clarify this.

[5] If one still struggles to get on board with the idea that X is not a cause of R in SPEED GUN, then one should at least grant that X is only a cause by virtue of its relationship with V. Contrast this with the influence it has in RADAR due to the new detection mechanism and readout-algorithm in the device. It is this distinction in causal influence which extant interventionist theories struggle to capture.

[6] Since V and X are (implicitly) indexed to the same time, I take it that neither is a cause of the other, rather they are logicaly related; see §4.



(a) SPEED GUN

(b) RADAR

Fig. 2. (a) SPEED GUN; (b) RADAR.

4. Can we intervene on time derivatives?

Velocity and position may have different causal profiles, but they enter into the following logical relationship.

$$X = X_0 + \int_0^t V(X)\, dt \tag{3}$$

This tells us that position of an object in the X-direction at some time t is the sum of its initial position X_0 (which might, e.g., be set to a or b, in Fig. 1) and the integral of the X-component of velocity $V(X)$ over the period 0 (when the car is at X_0) to t. A similar relationship holds for other components of velocity as well so that, in general, velocity is the time derivative and position the integral. From hereon I'll simplify by assuming all velocities are in the X-direction, i.e. $V(X) = V$.

What kind of relationship does Eq. (3) imply? In the first place, I take it to imply a *logical* relationship between V and X. Though we may, perhaps, find philosophical, or even scientific, reason to believe in intrinsic velocities, the ordinary notion of velocity, the one that features in Eq. (1), expresses a relationship we should not expect to be even possibly broken. However, it is not enough to render the two variables completely indistinct. In particular, the relationship is not one of supervenience, since for any object o with specific position x at some time t and specific velocity v at t it will typically be the case that o could have had v and not x at t, and that o could have had x and not v at t. It is metaphysically possible, for instance, that the car in Fig. 1 could have any velocity at any location along the road.

Nevertheless, there is a supervenience relationship implied by Eq. (3). It implies that V at t supervenes on the values of X in the *vanishingly small neighbourhood* of position at t. This will include the exact position x at t but it must also include positions at other times too. This means that the velocity V of an object o at some time t *necessarily depends on* the positions X of o in the neighbourhood of t, so that any causal influence on V at t must also influence position in the region of t.

This consequence of the relationship between V and X is enough to reveal a number of potential issues for the interventionist treatment of SPEED GUN and, especially, RADAR. To begin with, it is enough to show in either case that any testing intervention I_t on V with respect to R (or indeed any variable) must have a 'collateral influence' on X. Consequently, any further control intervention I_c on X can only be an *addition* to X's causal influences. We can represent this graphically in the usual way with variables and interventions as nodes connected by directed edges representing relations of causal influence, as in Fig. 3. Such a circumstance is reminiscent of so-called 'soft' interventions which render their target variables dependent on their original influences in the model *and* the intervention (Eberhardt, 2014; Eberhardt & Scheines, 2007). Soft interventions therefore fail condition I2 of IV, and a control intervention on X which fails to remove all the causal influences on its target will fail it too.[7]

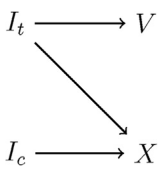


Fig. 3. Graph showing the collateral influence alongside a control intervention on X of a testing intervention on V.

More issues arise in the specific case of RADAR. First, due to X's influence on R via an alternate route than via V, I_t's collateral influence on X gives rise to a failure of I3 (see the thickened edges in Fig. 4). Second, any simultaneous control intervention I_c on X must hold X fixed and so *compensate* for the influence of I_t. But as I'll demonstrate, this seems to require the introduction of problematic dependencies.

Consider how the compensation would be practically achieved. Say the car's driver intends to test the influence of the car's velocity V on the speed gun reading R and chooses to do so by increasing the pressure put on the gas pedal. And say their chosen method of compensation for the collateral influence on X is to adjust the starting position of the car to be further away from the speed gun (e.g. from position a to b in Fig. 1). The graph in Fig. 5 indicates how the driver could make sure position would remain the same in this case despite a change in velocity. Given some pre-decided setting of V to v the driver will need to calculate which exact position to start from so that X remains at the value x when the reading is taken in both scenarios. Alternatively, the driver could first decide to start from position b then work out the pressure on the gas pedal required to get the car to x at t. A third option would involve the driver writing down a list of six pairs of starting positions and gas-pedal pressures each of which ensures the car will reach x at t and then selecting which pair to realise on the basis of the roll of a die. With each strategy, the driver is establishing a *causal dependency* between V and X. In the first strategy, the pre-decided change in velocity influences how position is to be maintained; in the second strategy, a pre-decided intervention on position influences how much velocity is to be influenced; in the third strategy, a further variable Z (the outcome of the die) decides both the appropriate intervention on velocity and intervention on position. These strategies correspond to the respective graphs a – c in Fig. 6.

The various strategies considered are only some of the countless plausible ways in order to ensure that the control intervention on X keeps it at the same value when a testing intervention V occurs. But the lesson is also entirely general: some dependency relation between interventions will need to be established.[8] Assuming the dependency will have to be causal, the graphs in Fig. 6 seem to exhaust the options in this regard. But in RADAR, where X does influence R, each of these established dependencies violate one or other of the interventionist criteria. If I_t influences I_c, as in Fig. 6a, then it violates I3 by influencing the effect variable (R) not only via a direct influence on X, but also now through I_c. Alternatively, if I_c influences I_t, as in Fig. 6b, then *it* violates I3 by influencing the effect variable via a route that doesn't go through *its* target. Finally, if both interventions are influenced by a common cause Z, as in Fig. 6c, then they both violate I4, since they will be statistically correlated.

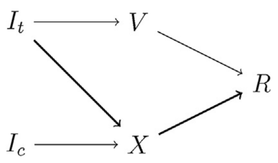


Fig. 4. Graph showing influence of R of a testing intervention on V with respect to R via X (displayed with thickened arrows).

[7] The latter intervention may not technically count as soft, since X is exogenous in the causal model and the intervention only fails to remove the causal influence of *another intervention*, viz. the testing intervention on V. It does seems plausible, however, that if we included whatever testing intervention I_t on V in an expanded model, the intervention on X would have to be soft in the technical sense.

[8] Thus, Weber's (2016) concern is arguably vindicated against Anderson (2020). Whether or not we take interventions on variables and their time derivatives to involve two different interventions, they will nevertheless be dependent interventions.

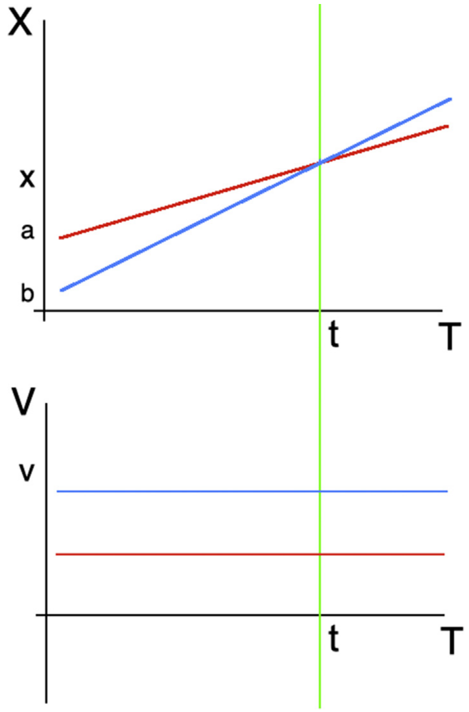


Fig. 5. Position-time (above) and velocity-time (below) graphs showing two possible trajectories of the car from locations a and b.

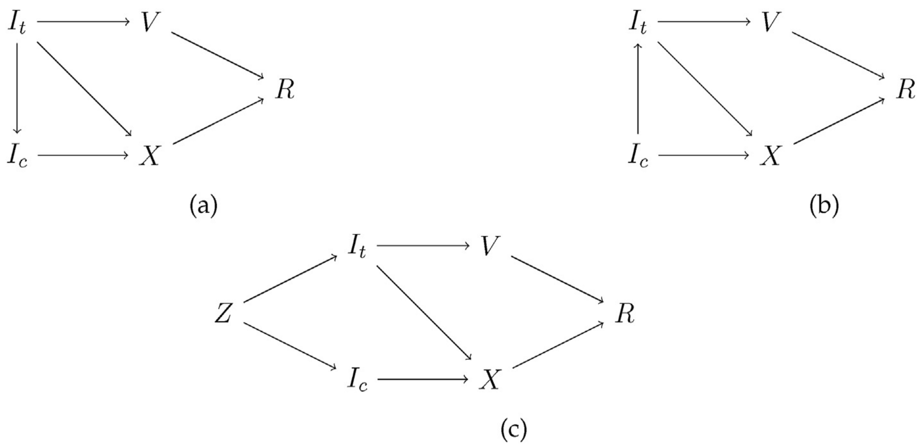


Fig. 6. Three ways testing and control interventions can be rendered dependent. Either (a) testing intervention influences control intervention; (b) control intervention influences testing intervention; or (c) a further variable influences both interventions.

In sum, the logical relationship between velocity and position mean that when there are testing interventions performed on the former with respect to some effect variable, control interventions on the latter will fail to remove all its other causal influences. Whether or not this means the control technically counts as a 'soft' intervention it is clear that the intervention will violate I2 for any model. Moreover, when position is itself a causal influence of the effect variable (as in the RADAR scenario) any practical strategy for ensuring that it is controlled for will lead to violations of either I3 or I4 for any model. Given that OIT aims to be an analysis of causal influence, it is compelled to accommodate the causal profiles holding in both SPEED GUN and RADAR. Specifically, it must be able to explain in interventionist terms why V is a cause of R but X need not be, or may be a cause along a different route. But by the foregoing reasoning it appears that the only strategies for intervention lead to violations of its criteria for intervention. At the very least this means that OIT cannot get the right causal verdict with respect to V's influence on R. For OIT only validates the existence of a relationship of (direct or contributing) causal influence if there is an intervention on V with respect to R which satisfies IV while control interventions, which must also satisfy IV, hold fixed off-path variables (including X). So how should the interventionist react? I'll consider four different responses.

5. Response #1: chancy control

One reaction to the imputed practically necessary dependencies between the interventions I_t and I_c in order for the latter to hold X fixed is that they are not in fact essential. The foregoing reasoning highlights the need for dependencies between interventions I_c and I_t if the compensation for I_t's collateral influence on X is to be *ensured* by I_c. But the correlation might alternatively simply be a matter of chance, with I_c managing to compensate for I_t's influence on X by just so happening to have the appropriate causal influence.

I don't think we should deny such chancy occurrences. But there are problems with relying on them in an interventionist setting. First, it's hard to see what could justify the claim that I_c *holds X fixed* in such a case. Typically, interventions which hold others fixed are taken to be 'controls' (Woodward, 2003, p. 64) which 'lock' the variable to a particular value (Eberhardt, 2014, p. 736). This suggests that control interventions do so robustly under perturbations in whatever other interventions are being performed. But I_c's ability to cancel out the collateral influence of I_t on X would not be robust if it were chancy, and one could be forgiven for thinking the idea of 'chancy control' is simply incoherent.

Second, this lack of robustness shows exactly why one would never perform interventions which 'hold variables fixed' by such methods in practice. If one really wanted to test the causal relationships among the variables, either one would implement a mechanism which ensures that the one intervention cancels the other—in the manner of Fig. 6a–c[9]—or one would take enough data that one can reliably *condition* on the case in which X retains the same value before and after the interventions are performed.[10] As we have seen, however, the former option renders the variables only *dependently* fixed. But the latter option undermines the whole interventionist enterprise. As already mentioned in §1, the interventionist theory of causation is an attempt to go beyond statistical techniques. Woodward, for instance, has repeatedly emphasised the

[9] Frisch (2014, 87) gives an vivid example of this kind of technique at the LHC.

[10] This seems to be the technique Eberhardt and Scheines (2007) have in mind.

contrast between statistical (or ‘non-experimental’) and interventionist approaches to causation (Woodward, 2003, pp. 31–32; Woodward & Hitchcock, 2003a,b, 14; Woodward, 2016b). The former are, he advises, implausible analyses of causation. Yet if interventions were simply treated as further randomised variables to be conditioned on in the usual statistical way (via Markov and faithfulness conditions) the interventionist account would have offered no improvement.

Third, interventionism is celebrated for its close affinity with scientific practice. But as we've seen, if we want to test the influence (e.g.) of V on R we might well consciously invoke a mechanism whereby we compensate for the collateral influence on X of interventions on V by working out what starting positions the car would need to have. That is, we make sure that the variables are dependent *exactly in order to* test for relations of causal influence. So although there might be wiggle room for interventionists to maintain that the interventionist criteria are not technically violated by the need to ‘hold X fixed’, it is hard to see why an interpretation of this requirement which avoids such violation should find its way into the theory's conditions on causal influence. Given that practice will actively pursue techniques which render interventions dependent in order to test for causal relationships, we should surely prefer an interventionist theory of causation which reflects why these practices work.

Fourth, and finally, the response from chancy compensation by control interventions is not anyway going to address the unavoidable fact that I2 will be violated. Nor will it address the possibility of I3 or I4 being violated in circumstances like that provided in RADAR, where I_t is an influence of the effect variable R via an alternative route than through its target variable. Therefore, it's worth looking elsewhere for a better response.

6. Response #2: model restriction

If chancy control is incoherent, or can at least be ignored, then the speed gun examples involve variables which effectively fail to be *independently fixable*. Although nothing in the interventionist definitions **M** or **IV** entails it, the plausibility of the framework itself is often taken to be applicable only where models satisfy a criterion of ‘independent fixability’ (or sometimes ‘independent manipulability’).

> **IF.** All subsets of variables in the model are independently fixable; i.e. it is metaphysically possible for all variables in any subset to be set to any combination of their individually possible values by independent interventions. (cf. Yang, 2013, 330; Woodward, 2015, 316; Weslake, forthcoming, 15)

Referring specifically to the interventionist characterisation of direct causation (‘DC’), Weslake describes **IF** as a ‘natural generalisation […] standardly assumed to hold in causal models, which can be motivated by the idea that for any set of variables appearing together in a model it must be possible to non-trivially test whether DC holds’ (forthcoming, 15). Yang similarly points out that while **IF** may not be strictly entailed by the interventionist theory, the theory ‘should be supplemented with a further condition such that distinct causes can be represented in a variable set only if the variables representing the causes are independently manipulable, i.e. the value of other variables is independent of any intervention performed on one of the variables in the set’ (2013, 330).

When considering the causal profiles of two variables which aren't independently fixable, the moral Weslake and Yang draw is therefore to preclude one of the variables from the model. Woodward (2015) is sympathetic with this idea, suggesting that **IF** should be granted when dependencies among variables are exclusively causal. And while his more formal definition of direct causation does not make explicit reference to the independence of interventions, Woodward's informal introduction of the relation seems to build the requirement of **IF** right into it.

> The basic idea is that X is a direct cause of Y if and only if the influence of X on Y is not mediated by any other variables in the system of interest **V** in the following sense: there is a possible manipulation of X that would change the value of Y (or the probability distribution of Y) when all other variables in **V** are held fixed at some set of values *in a way that is independent of the change in* X. (2003, 42, my emphasis)

Now, it is unclear in SPEED GUN and RADAR whether V and X technically count as failing the criterion of independent fixibility. But if the chancy control of X is incoherent, or must at least be ignored for practical purposes, then the variables in some sense fail to be *effectively independently fixable*. This might give some hope that restricting the available models which contain either V or X to ones which don't contain the other will eradicate the issues presented by the example.

If we do restrict the models this way, under the pretext of independent fixability then we do seem to avoid the incompatibility with the interventionist criteria described in §4. For if X no longer appears in the model alongside V then it needn't be controlled for when performing a testing intervention on V with respect to R. Consequently, there is no danger of a failure of I2. Moreover, since no causal model will involve *both* X and V (due to their effective dependent fixability) there can be no concern that I_t will influence R along an alternate route than via its target V. After all, it can't be fairly assumed that the causal influence of pressing harder on the gas pedal via the car's velocity V counts as a violation of I3 without our taking it that there is a model in which V is included alongside X and R (and that X is a cause of R which doesn't go through V). But if models involving both V and X are blocked, there can't be any such models. Finally, if X needn't be controlled for, there is no danger of a dependence between I_t and a control intervention on X.

Despite avoiding the previous issues the response of restricting the available models creates the new problem that it doesn't get the causal profiles right. We might expect it to get the right causal profile of V. That's because there seems to be an intervention I_t on V (e.g. pressing harder on the gas pedal) which changes R, and with X out of the way I_t does not need to be explicitly controlled for (assuming there are no further variables introduced). However, by analogous reasoning we are committed to the result that X is also a causal influence of R. For the logical connection between V and X means that very same intervention I_t which changes V (pressing harder on the gas pedal) can also change X. And since I_t is an intervention which results in a change to R in a model which includes X and R but not V, the interventionist theory will deem X a cause of R. But that is not the way we should want to assess X’s influence on R. We have, of course, considered a case in which X is a cause of R, viz. RADAR. But the prohibition of V and X from the same model effectively gets the right result in this case by the wrong reasoning—for the causal relationship between X and R shouldn't be revealed by *precisely the same intervention* as it is for V’s influence on R. More significantly, in SPEED GUN, X is decidedly *not* a cause of R. So keeping V out of the model will get the wrong result in this particular case.

Notice that the interventionist can't protest that the I_t intervention on X violates I3 because it goes through V as well. For as we just pointed out, it can't fairly be assumed that the causal influence of pressing harder on the gas pedal counts as a violation of I3 without our taking it that there is a model in which V is included alongside X and R. And if there are no models in which it can be demonstrated that pressing on the gas pedal causes influences X while also causing R through a different route, then we have no reason to think that it can't be employed as an intervention on X with respect to R. Perhaps this assumption could be brought into question, but this would then reflect back on the positive result that V is a cause of R which was justified by the same reasoning in the previous paragraph: if we can validate V’s influence on R, then the same procedure validates X’s influence on R, whether or not there is such influence.

Restricting the available models by prohibiting X and V from being in the same models is not, therefore, a good option. Although interventionist criteria might be satisfied if we do so, OIT simply gets the wrong causal results. However, the response was motivated by the observation that V and X are in some sense not independently fixable. So perhaps we should consider the alternative treatment of depdendently fixable variables demonstrated by Woodward's more recent extension of the interventionist framework.

7. Response #3: mitigated variation

We just explored the consequences of applying OIT to the speed gun examples under the assumption that V and X are not (effectively) independently fixable. According to the claims of Weslake, Yang and (at times) Woodward, such combinations of variables should not be allowed in the same model. By prohibiting such models OIT did not suffer the original violations of interventionist criteria, but these issues were replaced with an inability to get the right causal results. However, Woodward has also offered an *extension* of the interventionist framework specifically in order to *accommodate* variables which are not independently fixable. So perhaps the right response to SPEED GUN and RADAR may be to apply this alternative framework instead.

This 'Extended Interventionist Theory' (EIT), as I'll refer to it, works by updating the definition for causation with an alternative (**M***) and the criteria for intervention I3 and I4 with alternatives (I3* and I4*). The updated theory mitigates the varying of particular off-path variables under intervention which are related to the target variable or putative effect-variable by some special class of relations L. Hence, not all off-path variables need to be held fixed in determining relations of direct and contributing causation. The new conditions can be stated as follows.

> **M*.** A necessary and sufficient condition for X to be a (type-level) direct cause of Y with respect to a variable set $\mathbf{V}$ is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables Z_i in $\mathbf{V}$ *except those related to either X or Y by L*. A necshy; essary and sufficient condition for X to be a (type-level) contributing cause of Y with respect to variable set $\mathbf{V}$ is that (i) there be a directed path from X to Y such that each link in this path is a direct causal relationship [...], and that (ii) there be some intervention on X that will change Y when all other variables in $\mathbf{V}$ that are not on this path are fixed at some value *except those related to either X or Y by L.*
>
> **IV*.** I is an intervention variable for X with respect to Y if and only if, I1; I2;
> I3*. Any directed path from I to Y goes through X *or some variable related with X or Y by L*. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y, if any, that are built into the $I - X - Y$ connection itself *or are related with X or Y by L.*
> I4*. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X *except if Z is related to X by L.*[11]

Obviously, much hangs on what goes into L. The relations theorists typically have in mind are ones of supervenience. The idea is that this will be helpful in avoiding causal exclusion worries by establishing, for instance, that variables concerning the mental can causally influence some physical variable despite being incapable of being intervened on without simultaneously intervening on some other physical variable on which the mental variable supervenes and which is causally sufficient for the physical effect. As already remarked on, it is doubtful that the relationship between V and X are related by supervenience. Nevertheless, the relationship is a logical one, and one might well ponder whether EIT might be interpreted to accommodate such relationships. Let us therefore expand the class of relationships L determining which off-path need not be held fixed so that it includes relationships of time derivatives to integrals. Does this enable EIT to succeed where OIT has failed?

Unlike in the previously considered response, under EIT X and V can both be in the same model. But also, due to the relationship L holding between V and X, a testing intervention I_t on V with respect to R again needn't be accompanied by a control intervention I_c on X. Hence again there is no risk of a failure of the interventionist criteria due to a dependency between I_c and I_t, or because I_c would fail to be arrow-breaking. Moreover, although in RADAR there is a failure of the original I3 criterion due to I_t causally influencing R via a route that doesn't go through V, there is no violation of I3*, since the alternate route goes through a variable (X) related to V by L.

When it comes to causal assessment, again things also look good for the theory's assessment of the causal profile of V. An intervention I_t on V with respect to R (e.g. pressing firmly on the gas pedal) might fairly be made which establishes that V is a cause of R without having to control for X. But once again, as with the previous response, this same intervention will also count as in an intervention on X with respect to R.

After all, the intervention causes a change in X which doesn't influence R via any route other than through X, except through variables related to X by L. So, there need be no control for V by parallel reasoning. But if an intervention (i) counts as an intervention on both V and X with respect to R, (ii) brings about changes in V, R and X, and (iii) validates the causal relationship between V and R, then it *must* validate the causal relationship between X and R as well. So, by incorporating the variables into the same model, but being more permissive with the criteria for intervention, EIT also fails to allow for the result that X is not a cause of R. In SPEED GUN it will, therefore, get the wrong result; in RADAR, it will get the right result for the wrong reason.

The result for the more permissive EIT which mitigates the variation of some off-path variables, then, is ultimately much the same as the strategy in OIT of restricting the class of available models. In either case the problem is that pressing harder on the gas pedal works as an intervention on both V and X. Since in neither case is it required that the other logically related variable be kept at the same value under the intervention, both variables count as causes of R.

8. Response #4: modified interventionist theory

Let's take stock. We began this enquiry by observing that interventionism analyses the relationship of causal influence between potentially static variables in terms of possible changes is them. This prompted the concern that problems might emerge for the framework where time derivatives are involved. And we saw more precisely how this could occur in §3 and §4 where we considered the SPEED GUN and RADAR examples. In either case interventions on a car's velocity seem to have an inevitable collateral influence on the car's position, hence controlling for position seemed to violate a number of OIT's criteria for intervention.

Over the past three sections, we've considered three different responses to this issue all of which aim to avoid the need for establishing dependencies between testing and control interventions. In the first response (§5) we considered whether problems would dissolve if the control intervention was chancy. In the second (§6), we considered prohibiting the off-path variable from being included in the relevant models altogether. In the third response (§7) we considered allowing the off-path variables to stay in the model but nevertheless fail to be controlled for. All the responses failed in some way or other. The first did not address all the relevant issues and also seemed to rely on the implausible idea that control could be established by chance. The second and third lead to an incorrect assessment of the causal profile of the off-path variable.

The lesson seems to be that we should stick with the idea that X needs controlling for in assessing whether V is a cause of R and, moreover to do

[11] Woodward never formally formulates his extended criteria, but they can be pieced together from what he says (e.g. pp.334). See also Baumgartner & Gebharter (2016, pp. 745–6) Harinen (2018, p. 46) Prychitko (2019, p. 5). Notably, none of these authors include time derivatives or integrals among the permitted 'unfixed' variables.

so in a robust, non-chancy way. In other words, the interventionist should just embrace the fact that X should be controlled for by establishing dependencies among the variables' interventions. In this final response I want to argue for a modification to the interventionist criteria that will do just this.

To get us started, notice that while the problem cases were initially generated via a concern about interventionism's potential conflation of correlations of changes in variables with causal relationships, it is the existence of causal routes from testing and control interventions involving *no* changes which generated the actual inconsistencies. The violation by the testing intervention I_t of I2 and I3 due to its influence on X is due to the control intervention on X failing to break I_t's collateral influence. But we nevertheless allowed that a control intervention on X (e.g. adjusting the starting position of the car) could nevertheless prevent X from changing value. Even if X is a cause of R along a route that doesn't go through V (as in RADAR) the lack of change in X's value effectively puts it out of any causal enquiry. In fact, the dependency between testing and control interventions required in order to enable X to be robustly controlled for doesn't bring about any more changes in variables other than the target of the testing intervention V and the effect R. The conflict with OIT, therefore, came only from the presence of the alternative routes, not any further changes along them.

This suggests that the correct response for the interventionist to the problem case is to reconfirm their original motivating intuition, that it really is just corresponding changes in variables that matter to the determination of causal relationships regardless of whether the interventions required to establish this exhibit dependencies. Consider, then, the following alternative set of criteria for the sort of interventions which are applicable when assessing the causal relations in a model.

> **IV**. I** is an intervention variable *set*, where each intervention I_i in the set has some target X_i with respect to a single variable Y if and only if,
> I1**. Every I_i in **I** causes its target X_i;
> I2**. Every I_i in **I** either acts as a switch for all the other variables that cause its target X_i *or acts in addition to those variables*,
> I3**. the variables along any path from I_i to Y whose values change under **I** goes through the target of a testing intervention in **I**.
> I4**. each I_i is only statistically dependent on at most the target of one other intervention in **I** which changes under **I**.

In conjunction, **M** and **IV**** constitute a *Modified Interventionist Theory* (MIT). The key difference between **IV**** and **IV** is that it characterises the appropriate interventions for a causal enquiry in terms of a *set of* interventions distributed across different targets. This is necessary in order for it to be able to specify which changes in variables are relevant. For each intervention acting alone might bring about many changes through its target or via collateral influence on other variables. But it's what happens when an appropriate set of interventions act together which matters for causal analysis. Recall that according to **M** any relationship of direct or contributing causation is established through an array of interventions where one and one alone has a target which changes its value but where there may also be others which have targets whose values must be held fixed (although, of course, other variables in the model may change which are not subject to any intervention). The former kind of intervention we've been calling 'testing interventions', the latter 'control interventions'. **IV**** makes crucial reference to this distinction. In combination **M** and **IV**** imply that the testing interventions can be dependent or depended on (either by having common causes, causing or being caused by) any other intervention so long as its own target is the only variable which changes among all the interventions' targets.

Let's now compare **IV** and **IV**** more carefully.

I1** is essentially I1 cast in terms of the entire set of interventions necessary to establish causal relationships.

I2 has been amended to I2** to make reference to the full set of interventions but also to permit the case where the intervention's target variable (X) is to be controlled for by an intervention due to the collateral influence of a further intervention. That will permit control interventions on the likes of X in SPEED GUN and RADAR despite there being a simultaneous testing intervention on V which has X as a collateral influence. But it will also permit testing or control interventions on X when the intervention on V is a control intervention (since the pressure on the gas pedal remains an influence of position even when held constant). The result of the amendment is to make I2** seem somewhat superfluous within the interventionist criteria. That is perhaps something we would anyway suspect given an accommodation of soft interventions which have been supposed to 'play largely the same role in inference and causal interpretation as arrow-breaking interventions' (Woodward, 2015, p. 321). I keep it in for clarity's sake.

I3** can serve the same purpose as I3 in cases where collateral influence among interventions is not an issue. Assuming an intervention only causally influences its own target, then it satisfies I3 trivially. But I3** also permits interventions to be an influence of the effect through other routes than through their own targets so long as, under some value, the values of all the variables on each of those routes only change if they begin with the target of a testing intervention. So, in the case where I_i is itself a testing intervention, then it can bring about changes along routes that begin with its own target X but not changes along any routes that begin with the targets of a control intervention which it may have collateral influence on. In the case where I_i is a control intervention, then I3** permits it to bring about changes along routes that begin with other interventions' targets, so long as those are the targets of testing interventions.

Similarly, I4** can obviously serve the same purpose as I4 in the case where interventions are not dependent. But I4** also permits interventions to be rendered dependent on other variables either via direct causal relations or because they are the effects of a common cause. Crucially, however, such dependencies are only allowed when at most one route to the effect (either through the intervention's own target or the further correlated variable) has variables all of whose values change.

The amendments to I3 and I4 can be motivated by example. Consider the case discussed in Woodward (2003, 110–1) and Pearl (2000) in which an incubated child at risk for retrolental fibroplasia is injected with vitamin E. In order to administer the injection, the oxygen-saturated incubator has to be opened affecting the air-pressure and oxygen in the child's environment. But it is known that oxygen-saturation and air-pressure are independent influences on health. If the injection is to count as an intervention I_t which reveals whether vitamin E is a direct causal influence on health H, the air-pressure and oxygen saturation A would therefore need to be controlled for. According to OIT, this would require an independent controlling intervention on A, such as the influence of a further experimenter who briefly pressurises and oxygenates the room where the incubators are at the same time that the first experimenter administers the vitamin. But given that it would be impractical to have the room permanently pressurised and oxygenated, and impractical to hope for a chancy correspondence in pressurisation and oxygenation of the room with the administration of the vitamin, the two experimenters will likely arrange to act together.

The conspiracy of the two experimenters immediately makes for certain violations of the traditional criteria I3 and I4.[12] First, the intervention on E will violate I3, since I_t would be connected to H via A. Moreover, the conspiracy between the experimenters would mean that the interventions themselves would not be independent. Either one experimenter tells the other when to act, in which case there will likely be violations of I3, or they both arrange to act when some further variable tells them to, in which case there will be a violation of I4. Even so, I take it that the conspiracy of the experimenters is an entirely legitimate way to test for the influence of E on H. The modified interventionist criteria help us see why. For although there may be multiple routes to the effect, only

[12] I2 is also violated because the testing intervention on E will influence A.

ones which go through the target E of the testing intervention are such that all their variables change. The dependencies established by the experimenters' consciously acting together don't matter so long as there are only changes in the right places.

The incubator case is structurally similar to RADAR, and the **IV**** criteria make sense of the advised interventions on the latter in the same way. Moreover, these criteria help us establish the causal verdicts we wanted in both RADAR and SPEED GUN. First, let's look at V's influence on R. According to MIT, V counts as a cause of R in a model which includes X since there is a (testing) intervention I_t on V with respect to R (e.g. changing the pressure on the gas pedal) which changes V and R while another (control) intervention I_c (e.g. adjusting the starting position) holds fixed X. As we saw in §4, I_t will have a collateral influence on X and so violate I2. But it won't violate I2**, which tolerates such collateral influence. Also, we saw how I_t will violate I3 and possibly also I4 in the specific case of RADAR. One reason for the violation I3 was because I_t will influence R via a route that doesn't go through V. But I_t won't thereby violate I3**, since the alternate route is not one along which all the variables' values change; specifically, X's values don't change. The concern that I_t was bound to violate either I3 in a further way or else violate I4 was due to the fact that I_t and I_c must be correlated in order to ensure robust control of X. But this won't lead to a violation of either I3** or I4** since again there is only one route along which all the variables change values, and that's the one leading from the target of the testing intervention I_t.

This is enough to reveal success of the modified theory where OIT and the response from chancy control failed. To this extent MIT is on at least an equal footing with the model-restricting response and the mitigated-variation response. But MIT also improves on these too, since it can also make sense of the causal relationship between X's influence on R. The issue for the model-restricting and mitigated-variation responses was that in both cases the same action which counted as a testing intervention I_t on V with respect to R, e.g. changing the pressure on the gas pedal, also counted as a testing intervention on X with respect to R. That means there was no possibility of a distinction between the causal verdicts between X and V. But under MIT this does not follow. If I_t is a testing intervention on X then it must change the value of X. But if I_t were also to bring about a change in V (as we might expect) then it will violate I3**.[13] That is because there will be a route through to the effect along which all the variables change which doesn't begin with the target of the testing intervention on X, viz. the route through V. To prevent V changing, we should, according to the modified theory, implement a control intervention on it, e.g. severing the connection between the gas pedal and the engine. It may then be hard to conceive of how changing the pressure on the gas pedal could still bring about a change in X, but supposing it were able to do so then we would presumably we would *not* thereby expect a change in R.

Of course, in RADAR, X is a cause of R. And so there should be interventions on X with respect to R which do show this. For example, starting position might be adjusted while pressure on the gas pedal is kept constant. In this case, R would signal the influence. Not so for under the same kinds of intervention in SPEED GUN. MIT accommodates this in a way that OIT can't, since it permits the control on the gas pedal to have collateral influence on position.

The move to MIT therefore seems to be the only response considered to the speed gun examples which can validate the causal results. It does so by removing the requirement that interventions not have direct influence on the same variables as each other and focuses instead on where the changes are as a consequence of them. As long as dependencies can be established among the testing and control interventions then it will still be possible to establish only a change among all the interventions' target variables of only the target of a single testing intervention. That's enough, it seems, for robust causal verdicts.

The acceptability of this strategy raises a question as to the importance of Independent Fixability (**IF**). Due to vagueness in the terms employed, it's perhaps not clear that **IF** is technically violated by the permitted dependencies among interventions. The difficulty in saying either way concerns whether or not ‘chancy control’ is an acceptable form of genuine control. Beyond this, **IF** may be required, as Yang suggests, for reasons of conceptual validation of variables' distinctness. Nevertheless, for causal theorising purposes it seems that a weaker criterion of ‘respective fixability’ might by more suitable.

> **RF.** All subsets of variables in the model are respectively fixable if it is metaphysically possible for them to be set to any combination of their individually possible values by (dependent or independent) intervention(s).

A requirement of **RF** is sufficient to preclude interventionism wrongly making famously undesirable conclusions, such as that saying hello a cause of saying hello loudly (cf Kim, 1973), but it stops short of asserting the potentially overly constraining requirements of **IF**.

9. Corollaries of the modification

The circumstances arising from the SPEED GUN and RADAR examples are enough, I believe, to have motivated a modification to the interventionist theory. Moreover, I hope to have shown (by reference to the incubator case) that the suggested modifications in MIT (specifically of I3 and I4) are independently justifiable. However, the issue posed by SPEED GUN, RADAR and the incubator case is only one among a number which the interventionist faces. In particular, there has been plenty of recent debate surrounding the interventionist treatment of causal exclusion (Baumgartner, 2009, 2010, 2013; Yang, 2013; Woodward, 2015; Weslake, forthcoming) and part-whole relations (Baumgartner & Casini, 2017; Baumgartner, Casini, & Krickel, 2020; Baumgartner & Gebharter, 2016; Krickel, 2018). One might wonder, therefore, whether MIT sheds new light on these other problem-cases.

In the case of causal exclusion, the worry stems from the fact that supervening variables may also causally influence effects which their supervenience base influences. Since the former cannot be changed without a change in the latter, it is impossible under orthodox interventionist theory to establish the required control interventions to test for the supervening variable's influence. MIT is unlikely to help here. The modification proposed specifically exploits the *non*-supervenience of variables on their time derivatives. Hence, for example, although the respective interventions will be dependent, it is still possible to hold position fixed while changing velocity. For this reason, MIT will likely not be straightforwardly relevant to a solution to the causal exclusion problem.

Things may be different in the case of part-whole or ‘constitutive’ causal cases, where supervenience is not guaranteed between a whole and some proper part. Subsequently, MIT might offer some intuitive results not available on other extant accounts. Consider a case (borrowed and adapted from Thomas Blanchard) in which a Jury of ten vote either to convict or acquit a defendant. The law stipulates that a majority of ten votes is required to reach a decision. Let V_i be dichotomous variables standing for the vote of each individual juror (where $i \in \{1-10\}$) and D for the overall decision. I assume that D is *constitutively related* to the juror's individual votes. Now let A be a dichotomous variable representing the anxiety of juror 1. Plausibly, A will be a function both of that juror's own vote and the overall verdict, i.e. $A = f(D, V_1)$ (suppose they feel anxiety about their own decision independently of the overall verdict, but also feel affected by whether their vote matches the overall decision).

A critical feature of such a scenario is that, since V_1 and D are necessarily statistically correlated, a testing intervention on V_1 will not be independent of a control intervention which holds D fixed (imagine that including V_1 there are exactly six votes for acquittal, so that changing V_1 to

[13] Of course, there is also a violation of I3 too. But this is mitigated by the model-restricting response, according to which no model includes both V and X, and by the mitigated variation response, which substitutes I3* for I3.

convict will require changing some other V_i to acquittal). Consequently, OIT will not permit us to test the independent influence of V_1 on A due to a violation of I4. EIT suffers in a different way. By permitting constitutive relations like that between D and each of the V_is to fall into L (see §7), it will permit the interventions, but it will not allow us to test the influence of D and V_1 *individually* on A, thereby failing to reveal the fact that A is influenced in different ways by each variable. Only MIT is able to capture the nuances of the scenario in the way we should want interventionist theory to do so. Specifically, it allows for D to be held fixed by a control intervention which is dependent on our testing intervention on V_1 and as a consequence it allows us to reveal V_1's distinct influence on A.

MIT works in the jury case in a similar way to that in RADAR and the incubator case. In all these cases two closely related variables have a distinct influence of the same effect-variable. As before, the overall lesson is that we need an interventionist theory which can expose this distinct influence and MIT seems to be the only way to do it. The case involving the jury and the juror's anxiety shows that this sort of issue can plausibly crop up when the problem-variables are constitutively (or part-whole) related. It should be pointed out, however, that MIT does not itself pass any novel judgement on whether or not constitutively related variables are causally related and what sort of interventionist treatment, if any, can be given for revealing relations of constitutive relevance.

10. Conclusion

Interventionist theories of causation aim to give an analysis of causal influence among variables in terms of their changes under suitably distributed testing and control interventions. The correspondence between correlated changes and causal influence is not a given and it is foreseeable that problems will emerge when we consider models in which variables are related as time derivatives to their integrals. Two problem-cases involving a car travelling towards a speed gun (or augmented device) showed that there is indeed cause for concern, revealing potential issues for the orthodox criteria for interventions. I considered a number of responses to the case, including permitting control interventions to be chancy, restricting the available models and mitigating the variation of some off-path variables (such as those related as time derivative to its integral or vice versa). All these solutions, I argued, are insufficient to get the causal verdicts right.

The conclusion at this stage might justifiably have been to give up on the interventionist conflation of correlations in changes under intervention with causal relationships. But I have suggested, on the contrary, that the response should be to reassert this basic insight and give up on something else, viz. the assumption that interventions are only causally probative when they are independent. The proposal of a 'Modified Interventionist Theory' which eschews this requirement seems to get the causal verdicts right. And although such dependencies can initially seem out of keeping with the framework, they are reasonable in experimental settings (as we have seen in the incubator and speed gun cases). Interventionists should therefore allow that interventions can be rendered dependent so long as the modified interventionist criteria are satisfied. As far as such an interventionist is concerned, causation just is a correlation of changes under intervention.

Funding

The Metaphysical Unity of Science Project, funded by ERC through the Horizon 2020 Research and Innovation programme (grant no. 771509).

Declaration of competing interest

None.

Acknowledgments

I would like to thank Thomas Blanchard, Totte Harinen, Francesca Bellazzi, Samuel Kimpton-Nye, Vanessa Seifert, Tuomas Tahko and an anonymous reviewer for their helpful comments on various drafts of this paper.

References

Anderson, W. (2020). The compatibility of differential equations and causal models reconsidered. *Erkenntnis, 85*, 317–332.

Baumgartner, M. (2009). Interventionist causal exclusion and non-reductive physicalism. *International Studies in the Philosophy of Science, 23*(2), 161–178.

Baumgartner, M. (2010). Interventionism and epiphenomenalism. *Canadian Journal of Philosophy, 40*, 359–383.

Baumgartner, M. (2013). Rendering interventionism and non-reductive physicalism compatible. *Dialectica, 67*(1), 1–27.

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

Baumgartner, M., Casini, L., & Krickel, B. (2020). Horizontal surgicality and mechanistic constitution. *Erkenntnis, 85*, 417–430.

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

Eberhardt, F. (2014). Direct causes and the trouble with soft interventions. *Erkenntnis, 79*, 755–777.

Eberhardt, F., & Scheines, R. (2007). Interventions and causal inference. *Philosophy of Science, 74*, 981–995.

Frisch, M. (2014). *Causal reasoning in physics*. Cambridge University Press.

Harinen, T. (2018). Mutual manipulability and causal inbetweenness. *Synthese, 195*(1), 35–54.

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

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

Pearl, J. (2000). *Causality*. New York: Cambridge University Press.

Prychitko, E. (2019). The causal situationist account of constitutive relevance. *Synthese*. https://doi.org/10.1007/s11229-019-02170-4

Ross, L. (2020). Multiple realizability from a causal perspective. *Philosophy of Science, 87*(4), 640–662.

Schurz, G., & Gebharter, A. (2016). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese, 193*, 1073–1103.

Shapiro, L., & Sober, E. (2007). Epiphenomenalism: The dos and don'ts. In G. Wolters, & P. Machamer (Eds.), *Thinking about causes: From Greek philosophy to modern physics* (pp. 235–264). University of Pittsburgh Press.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, prediction, and search*. MIT Press.

Weber, M. (2016). On the incompatibility of dynamical biological mechanisms and causal graphs. *Philosophy of Science, 83*, 959–971.

Weslake, B. (forthcoming). Exclusion Excluded. International Studies in the Philosophy of Science.

Wilson, J. (2014). No work for a theory of grounding. *Inquiry, 57*(5–6), 535–579.

Woodward, J. (2003). *Making things happen*. Oxford University Press.

Woodward, J. (2008). Response to strevens. *Philosophy and Phenomenological Research, 77*(1), 193–212.

Woodward, J. (2009). Agency and intervention theories. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *Oxford handbook of causation, chapter 11* (pp. 234–264). Oxford University Press.

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

Woodward, J. (2016a). Causation and manipulability. In E. N. Zalta (Ed.), *The stanford Encyclopedia of philosophy (Winter 2016 ed.)*. Metaphysics Research Lab, Stanford University.

Woodward, J. (2016b). Causation in science. In P. Humphreys (Ed.), *The oxford handbook of philosophy of science*. Oxford University Press.

Woodward, J., & Hitchcock, C. (2003a). Explanatory generalizations Part 1: A counterfactual account. *Noûs, 37*(1), 1–24.

Woodward, J., & Hitchcock, C. (2003b). Explanatory generalizations Part 2: Plumbing explanatory depth. *Noûs, 37*(1), 1–24.

Yang, E. (2013). Eliminativism, interventionism, and the overdetermination argument. *Philosophical Studies, 164*, 321–340.