

Turning Negative Causation Back to Positive

Peter Fazekas

School of Philosophy, Psychology and Language Sciences,
The University of Edinburgh;
Institute for Philosophical Research, Hungarian Academy of Sciences
email: P.Fazekas@sms.ed.ac.uk

George Kampis

Department of History and Philosophy of Science,
Eötvös University
email: gk@hps.elte.hu

— draft —

Abstract

In contemporary literature, the fact that there is negative causation is the primary motivation for rejecting the physical connection view, and arguing for alternative accounts of causation. In this paper we insist that such a conclusion is too fast. We present two frameworks, which help the proponent of the physical connection view to resist the anti-connectionist conclusion.

According to the first framework, there are positive causal claims, which co-refer with at least some negative causal claims. According to the second framework, negative causal claims are generated from mapping and comparing different scenarios, which can fully be accounted for in purely positive terms. Since the positive causal claims evoked by both frameworks pose no obvious difficulties for the physical connection view, these frameworks make it possible for the connectionists to accommodate negative causal claims into their theory.

Once these strategies are available, the connectionists become able to render all the arguments starting from the observation that there are negative causal claims in our causal discourse inconclusive with regard to the viability of the physical connection view.

Turning Negative Causation Back to Positive

‘[O]missions and so forth are not part of the real driving force of nature. Every causal situation develops as it does as a result of the presence of positive factors alone.’ (Armstrong, 1999, p. 177)

1. Introduction

Causation, much like intentionality, supports intrinsic and derived notions. It should be clear that texts owe their intentionality to the persons who use them—without such authors or readers, it is just paint on paper. Similarly, certain causal accounts, as we shall claim in this paper, derive their meaning from other causal accounts.

The parallel between causal and intentional accounts goes even further. It does not take an intentional realist to admit, as we did above, that the meaning of texts derives from humans. The point where the two main positions, that of the intentional realist and the interpretationist differ is what they think about the intentionality of humans, not of texts. Similarly, in causality, we will identify cases where a causal account derives from a more primary account. Nonetheless, our analysis remains neutral about what the ultimate nature of causation might be, just as a reductive analysis of the intentionality of texts remains neutral about the ultimate nature of intentionality.

Negative causation is often used in accounts that describe seemingly ordinary causal relations that most people accept and perceive as causal. Not pressing the stop button kills our enemy who has fallen into the meat grinder. If anything is causal, it would be absurd to deny the causal effect of *not pressing* in this example, and this judgement is reflected in the ordinary moral, legal, and now philosophical accounts of the situation. A person can get convicted for not pressing the button if she had the ability to do so, and common sense coincides with contemporary philosophical analysis on this issue.

However, having the ability (and opportunity) is not all—in fact it is not sufficient. The person also needs an understanding (i.e. knowledge) of how the system works. Not knowing how buttons and other controls operate, there can be no way to stop the grinder, and there is no issue of responsibility either: if you don’t *already* know what can cause what in this scenario, all you can do is just to try things out in random (depicted as a funny series of erratic events in less heroic movies) or (at a higher degree of mechanical ignorance, such as in very young age) just watch in horror. Not

pressing the (right) button under these latter circumstances does not constitute killing—as common sense and legal practice tells. How about philosophical analysis?

Our purpose in this paper is to reconstruct negative causation in a way that utilises the above common-sense logic in a more precise way, and reflects certain observations about generating negative causal claims. In contemporary literature, the fact that there is negative causation is the primary motivation for rejecting a certain kind of causal account, namely the physical connection view.¹ Our fundamental aim in this paper is to show that once one recognises and respects the actual process of formulating negative causal claims, one becomes able to reconcile negative causation with the physical connection view.

2. What is at stake in the debate over negative causation?

Negative causation is the case where an *absence* plays the role of either the cause (*omission*: not-C causes E) or the effect (*prevention*: C causes not-E). Negative causation is undoubtedly present in our everyday causal talk. People often make causal claims like: ‘I failed on the exam due to my lack of knowledge’ meaning that not having the necessary knowledge (not-C) caused one’s failing on an exam (E), or ‘we couldn’t finish the tennis match because it started to rain heavily’ meaning that the downpour (C) caused not finishing the match (not-E). Negative causation also populates scientific theorising (the absence of vitamin-C causes scurvy, the absence of insulin causes diabetes, etc.), and features in legal judgements (the driver not paying attention to the road signs caused the accident).

We have no intention to deny any of these observations. On the contrary, we acknowledge that negative causation legitimately exists in causal talk. However, even if the presence of negative claims in causal discourse is admittedly beyond dispute, there is little consensus about the status of negative causal claims.

For example, Philip Dowe argues that negative causation, rather than being genuine, is only ‘quasi’ causation (Dowe, 2000, 2001, 2004). Dowe starts his analysis by identifying two conflicting intuitions he claims we all share. On the one hand, we often have an intuition of difference when comparing cases of negative causation (e.g. that it is the father’s negligence—the absence of his attention—what causes the son’s accident) and cases of genuine causation. On the other hand, in relation with certain cases of negative causation (e.g. chopping off the head causes death) we have

¹ According to this account, the causal connection between the relata of causation is due to a physical connection such as, for example, energy flow. Versions of the physical connection view differ in what they think about the nature of this physical connection. Some claim that it consists in a transference of energy-momentum (Fair, 1979). Others argue that it is in fact a process, a continuous qualitative presence transmitting a mark (Salmon, 1984, 1994, 1998; Dowe, 2000). Still others characterise the physical connection in question as an intrinsic tie (Mackie, 1980).

strong intuitions of genuine causation. Next, Dowe provides his so-called counterfactual account which (as opposed to the strategy of the proponents of negative causation who typically build on the second intuition while neglecting the first one) is able to deal with both intuitions: it respects the first and explains away the second. According to Dowe, negative causal claims are in fact counterfactual claims about genuine causation—instead of expressing actual genuine causation they express the mere possibility of genuine causation.

Contrary to this, Jonathan Schaffer argues that negative causation is indeed genuine causation (Schaffer, 2000, 2004). He draws attention to how strong the genuinist intuition is: negative causation is supported by all the central conceptual connotations of causation, it features in paradigm cases of causation, it is required by the most useful theoretical applications of causation, and is recognised in scientific practise (cf. Schaffer, 2004, p. 203). Moreover, Schaffer argues that Dowe's main motivation for thinking that negative causation is only quasi-causation, namely the intuition of difference, is in fact a misdescription of our intuition and implicitly assumes the physical connection view of causation.

This latter point is especially problematic, since what is at stake in this debate is whether certain accounts of causation are to be preferred over other accounts. That is, the problem of negative causation owes its importance to the consequences of its existence. As Schaffer puts it, negative causation refutes the program of the physical connection account. He says:

‘The physical connection view of causation may seem plausible if one concentrates on colliding billiard balls, or other cases of connection. But negative causation reveals the view to be a hasty generalization.’ (Schaffer, 2004, p. 204)

Schaffer argues along the following line of thought. Given that negative causation is genuine causation the physical connection view necessarily fails to provide a general account of causation since in the case of omissions and preventions causes and effects are not physically connected: there can be no physical connection between the *absence* of something and the occurrence of something else.²

Moreover, since there are theories of causation, which do just fine in incorporating cases of negative causation, the conclusion that these theories are to be preferred over the physical connection account naturally arises. As Schaffer says:

² Dowe (2004), by claiming that negative causation is not genuine causation tries to reject this line of thought and defend the physical connection account.

‘So the situation is that negative causation poses a problem for only some theories. What should one conclude from that? I would have thought that the logical conclusion would be that negative causation provides an argument in favor of the theories that can handle it.’ (Schaffer, 2004, p. 212)

That is, negative causation is utilised as a motivation for criticising the physical connection account of causation and favouring others. Schaffer himself argues for the contrastive account (Schaffer, 2005), whereas, for example Carl Craver (2007)—on the very same grounds, i.e. building on the observation that there is negative causation—provides an argument for the interventionist (Woodward, 2003) account.

To recap, consider the role negative causation plays in the argument against the physical connection view. The claim is that cases of negative causation cannot be accounted for in terms of physical connections. The debate over the failure of the physical connection view turns on whether negative causation is genuine causation or not. That is, the argument from negative causation to the failure of the physical connection view (AfNC) has the following structure:

(AfNC) – Argument from Negative Causation

- (P1) There are negative causal claims (where either the cause or the effect (or both) is an absence) in causal discourse.
- (P2) Absences are not connected physically to either occurrences or other absences.
- (P3) Negative causal claims express cases of genuine causation.
- (C) The physical connection view fails to provide a general account of genuine causation.

In what follows, our paper is concerned with what possibilities a connectionist might have in order to render this argument inconclusive.³ We explore two different frameworks which could help the connectionist to overcome the obstacle of negative causation.

The first framework is motivated by the observation that negative causal claims often seem to co-refer with certain positive causal claims. That is, the first framework accepts that negative causal claims express genuine causation, but claims that the relations negative causal claims pick out can also be picked out by positive causal claims, which pose no difficulty for the physical connection

³ Of course, one such possibility, i.e. Dowe’s strategy, is already on the table. Here we are interested in what possibilities remain for those, who accept Schaffer’s objections to Dowe’s counterfactual account.

account. If this is the case, then one cannot infer from negative causation to the conclusion that the physical connection view is unable to provide a general account of genuine causation.

The second framework is motivated by an observation about how negative causal claims are formulated. Within this framework, it can be shown that negative causal claims in fact do not express genuine causation. Rather, they are derived from mapping and comparing different scenarios such that whereas the different scenarios themselves involve genuine causal sequences the negative claim picks the cause and effect from different causal sequences. Since all scenarios involved can be described in terms of positive terminology posing no difficulty for the connectionist, within this framework the physical connection view is able to explain the idea of negative causal claims expressing genuine causation away, which in turn, by rejecting (P3) stops the anti-connectionist argument.

In other words, both frameworks seem to offer a way for the connectionist to account for the relations picked out by negative causal claims solely in terms of physical connections. Once such a strategy is available for the connectionist, she becomes able to render all the arguments starting from the observation that there are negative causal claims in our causal discourse inconclusive with regard to the viability of the physical connection view.

3. Introducing the co-reference framework

Consider electron-holes and the void. Both are absences (electron-holes are absences of electrons, the void is the absence of everything), still both occur in causal claims as causes—they are classical examples of negative causation. However, on closer reflection they reveal something important which might give ammunition for those who try to reject the anti-connectionist argument.

Take electron-holes first. Electron-holes are positively charged quasi-particles—literally they are absences of electrons on the valence band. Physics often makes causal claims about electron-holes. The very fact that they are treated as quasi-particles reflects this. The quasi-particle talk attributes similar causal efficiency to electron-holes as ordinary particles possess, which makes them a clear case of negative causation (cf. Schaffer, 2004, pp. 202-203).

Closer reflection reveals that the electron-hole term stands for a phenomenon where an electron (due to gaining enough energy) escapes its bond and moves from the valence band to the conduction band. The electron-hole talk credits certain properties to the absence of the electron and treats it thereby as a quasi-particle. These properties, however, are in fact the properties of the band from which the electron escaped. That is, a positive description of the phenomenon reveals itself: a

band with such properties is causing what happens in scenarios usually described by the electron-hole talk.

‘The hole is an alternate description of a band with one missing electron, and we either say that the hole has wavevector $-\mathbf{k}_e$ or that the band with one missing electron has total wavevector $-\mathbf{k}_e$.’ (Kittel, 2005, p. 194)

That is, the term ‘electron-hole’ (the negative term referring to the absence of an electron) stands for a complex state of affairs which itself can be accounted for using only positive terminology.⁴ Since this positive description is a quite complicated one compared to the version employing the ‘electron-hole’ term, the negative term might be called a *shorthand*: it stands for a phenomenon which, though could be described by using positive terms only, this description would be a much more detailed (and complicated) one than what one can get by employing the negative term. That is, the negative terminology simplifies the story to be told.

The same with the deadly void. David Lewis’ famous example shows how the void is able to kill by sucking the air from the lungs, boiling the blood, draining the warmth from the body and inflating enclosures in the body until they burst (cf. Lewis, 2004, p. 276).

Just like in the case of electron-holes, closer reflection reveals that the negative causal claim ‘the void kills’ is a shorthand for a much longer story (which involves only positive terminology) telling how the air departs from the lungs, the blood boils, the body freezes and bursts. For example, the air departs from the lungs because the molecules in the lung, due to their kinetic energy, collide with each other and with the inner surface of the lung, thus building up a certain amount of pressure, which—unless one is placed into equal or bigger outside pressure—forces the mouth to open and lets the molecules escape. Similarly, the whole body bursts because the molecules inside exert a force, which expands the inner cavities until the tissues get damaged.

What these observations point out is the following. It is possible to refer to the very same causal relation⁵ by two quite different causal claims—one using some negative terms and one using only positive terms. That is, there is a positive causal claim picking out the very same phenomenon, which is picked out by the negative causal claim. For example, the negative claim ‘electron-holes

⁴ Notice that the positive term for ‘the band with one missing electron’ is the actual number of electrons on the band (which is often denoted as $N-1$, thereby emphasising that there could be N electron on the band).

⁵ Here, and throughout the paper, we use the term ‘causal relation’ as being neutral to what the correct account of causation might be.

attract negatively charged particles' refers to the very same causal relation as the positive claim 'a region with N-1 electrons attract negatively charged particles from a region with X electrons'⁶.

Notice that whereas the negative causal claim seemingly impedes a physical connection account (for it is hard to see how there could be a physical connection between something not being there—e.g. a missing electron—and anything happening—e.g. the attraction of a negatively charged particle) the positive causal claim readily invites such an account. The N-1 electrons in region A together exert a given amount of repulsive force on a negatively charged particle, which is also repulsed by the X electrons in region B, and since the latter force is bigger than the former one, the negatively charged particle in question moves toward region A.

These observations give rise to what we call the *co-reference framework*. This framework acknowledges that at least some instances of negative causal claims are such that the causal relation picked out by them can also be picked out by a positive causal claim. Since these positive causal claims do not pose any obvious problem for the physical connection view, the anti-connectionist needs to do more than just pointing out that there are negative causal claims in our causal discourse. That is, the co-reference framework draws attention to that for the argument from negative causation to the failure of the physical connection view to go through an extra premise needs to be inserted into the argument claiming that no positive causal claim picks out the causal relation picked out by the negative causal claim in question. The anti-connectionist argument supplemented with this further premise looks like this:

(AfNC)*

- (P1) There are negative causal claims (where either the cause or the effect (or both) is an absence) in causal discourse.
- (P2) Absences are not connected physically to either occurrences or other absences.
- (P3) Negative causal claims express cases of genuine causation.
- (P4) *No positive causal claim picks out the same causal relation picked out by a negative causal claim.*
- (C) The physical connection view fails to provide a general account of genuine causation.

This new version of the anti-connectionist argument opens up a new possibility for the connectionist to stop the argument. She can deny (P4) by pointing out that there are cases where there are positive causal claims co-referring with the negative claims in question. If so, the connectionist

⁶ Where N-1 and X are the actual numbers of electrons, and, say, X is bigger than N-1.

argument might proceed, then the anti-connectionist needs to show that there are cases where (P4) actually holds, for only those cases would suffice to drive the anti-connectionist argument home. That is, the burden of proof is on the anti-connectionist side.

4. The co-reference framework: pros and cons

At this point, it might be tempting for the connectionist to say something like ‘what *really* does *the* causal job, then, is the band itself with N-1 electrons on it’. That is, to argue that negative causal claims are in fact just re-descriptions of certain causal relations, which could otherwise be picked out *more accurately* by a positive description.

Philip Dowe (2001) has something similar in mind. When motivating what he calls the intuition of difference (that upon reflection we recognise certain instances of negative causation as being not genuine cases of causation) he argues that instead of what the negative claim ‘the father caused the accident by failing to guard the child’ suggests, what clearly was the cause of the accident was ‘the child’s running onto the road’ (cf. Dowe, 2001, pp. 217-218).

However, as Jonathan Schaffer (2004) drew attention to it, there are at least two problems with this logic. First, claiming that the positive descriptions reveal the ‘real’ causal connections whereas the negative descriptions don’t begs the question since it presupposes the priority of the physical connection view. Schaffer argues that what pumps one’s intuitions in this case is that one implicitly assumes that the physical connection view is correct, which in turn disqualifies the father’s negligence as a real cause.

Second, there is no such thing as ‘the’ cause; one can only talk about ‘a’ cause (given certain conditions). What is *a* cause depends on what condition is brought into the foreground while the others are kept in the background. In this sense, different conditions can all be *partial* causes. Schaffer argues that both the father’s negligence and the son’s running onto the road are partial causes (cf. Schaffer, 2004, p. 210).

Note, that the co-reference framework provides enough resources to answer both of these challenges. If negative causal claims co-refer with certain positive causal claims in a way that the negative claims shorten and simplify the story to be told then one does not need to rely on an implicit assumption of the physical connection view in order to judge the positive descriptions as superior. One can stay entirely neutral about whether the physical connection view is correct or not, still one’s intuitions would be pumped toward preferring the positive description by accuracy-considerations. The fact that typical examples of negative causal talk deploying such negative terms

as electron-holes or the void simplify the situation (i.e. provide causal claims which are less complicated than those deploying only positive terms) supports the argument that negative causal claims are just short—but less accurate—ways of picking out causal relations. By providing a shorthand, a negative causal claim makes it possible to grasp a certain causal relation even without knowing much of the details about it. If so, then, as the argument goes, the co-referring positive causal claim provides a probably much more complicated but, on the other hand, more accurate way of grasping the very same causal relation. That is, the argument that positive descriptions reveal the ‘real’ causal relation is supported in this sense of being more accurate.

As of the second problem: note that Schaffer’s point that both the cause as picked out by the negative claim and the cause as picked out by the positive claim are partial causes cannot be applied in the co-reference framework. Claiming that, for example, the presence of the void and the inner pressure higher than the outer pressure are both causes (partial causes) of the death is simply wrong. Since the negative description and the positive description both refer to the same causal relation it cannot be the case that the void is a partial cause along with the cause picked out by the ‘inner pressure higher than the outer pressure’ description. Once the positive story is told, there is no further role for the void—the void is nothing over and above than what have already been picked out by the positive description.

That is, it seems that within the co-reference framework there are reasons to give in to the temptation of seeing the positive causal claim as the one that picks out *the real* causal relation.⁷

The connectionist is not done, however. For there is a way for the anti-connectionist to retort. All the anti-connectionist needs to emphasise is that (P4) as in (AfNC)* is inaccurate. It is not enough just to have a positive causal claim co-referring with every negative causal claim—for the connectionist to be able to stop the argument from negative causation it is also necessary that the positive causal claims in question are such that they pose no problem for the physical connection view, i.e. can be accounted for in terms of physical connections. Inserting this requirement into (AfNC)* results in the following formulation:

(AfNC)**

- (P1) There are negative causal claims (where either the cause or the effect (or both) is an absence) in causal discourse.
- (P2) Absences are not connected physically to either occurrences or other absences.

⁷ Of course, again, only in the sense of being more accurate. Since, according to the co-reference framework, the negative and the corresponding positive descriptions co-refer, this framework does not support the further claim that negative causation is not genuine causation. For possible stronger conclusions, see Section 7.

- (P3) Negative causal claims express cases of genuine causation.
- (P4) No positive causal claim—*which can be accounted for in terms of the physical connection view*—picks out the same causal relation picked out by a negative causal claim.
- (C) The physical connection view fails to provide a general account of genuine causation.

Now, as a next step, all the anti-connectionist have to do is to come forth with one single example, where even if there is a positive causal claim co-referring with the negative one, it cannot be accounted for in terms of the physical connection view.⁸

As it is actually the case, there are quite good candidates for this purpose. Consider, for instance, the classical example of the gardener and the plant. According to the negative causal claim, ‘the gardener not watering the plant caused the plant’s death’. The anti-connectionist could argue here that even if a positive description was given, i.e. ‘the gardener napping caused the plant’s death’, the physical connection view would still fail, since there would be no physical connection between the gardener’s napping and the plant’s death.

Of course, the situation is a bit more complicated. The connectionist could dig her heels in and object that the failure of seeing the physical connection in the positive story might be a result of a failure of picking out the correct positive description co-referring with the negative one. That is, the positive description cited by the anti-connectionist might not co-refer with the original negative claim. Just to give one illustration, the connectionist could point out that ‘the gardener not watering the plant’ description tells us something about the gardener, the water and the plant, whereas ‘the gardener napping’ description has nothing to say about the water and the plant. That is, it seems that the two descriptions do not co-refer.

However, even if the connectionist wants to maintain that once the correct co-referring positive description (the one that really co-refers with the negative claim, and which might be rather complicated) is given the physical connection becomes obvious, this would not much of a help in this situation. For in cases like the gardener and the plant whether it is possible to give a positive causal claim, which is apt for being accounted for in terms of physical connections, is not straightforward at all. On the face of it, such cases do pose a problem for the physical connection view. As a result, the burden of proof is, again, on the connectionist side.

⁸ The anti-connectionist can even acknowledge that within the co-reference framework the positive causal claims pick out their referents more accurately, and thus are more real. In this case, the anti-connectionist could argue that the debate over negative causation ultimately boils down to whether physical connectivity might be caught out in the positive stories lurking behind negative causal claims.

That is, though the co-reference framework provides natural treatment for instances of negative causation like those involving electron-holes, it seems to be unable to offer a full characterisation covering all cases.

5. Introducing the mapping framework

As it happens, relying on co-reference is not the only option available for the connectionist to block the argument from negative causation. Another framework reveals itself if one concentrates on how negative causal claims get formulated.

Consider, as a test case, John catching the billiard ball before it reaches the pocket. One might feel tempted to say that ‘John’s catching the ball causes not scoring’. By saying this, one claims that catching the ball (C) causes not scoring (not-E), which makes it a clear example of prevention, i.e. of negative causation. But how could one know that if John had not caught the ball, the ball would have fallen into the pocket? True, in all those cases where John catches the ball, the ball ends up in John’s hand, therefore it does not fall into the pocket. However, neither bounces the ball back from the wall right next to the pocket (or occurs in the opposite corner of the table or in John’s mug, etc.)—one simply cannot know whether John’s catching the ball causes not scoring or, say, not missing the shot, without knowing what happens when John does not interfere.

It seems that the claim ‘John’s catching the ball causes not scoring’ cannot be formulated solely on the basis of observing John catching the ball. For one to conclude the negative claim (catching the ball causes not scoring), one needs to know that without John’s hand interfering, the ball would have fallen into the pocket—instead of e.g. popping out of space-time at one point and re-manifesting itself in John’s hand. That is, over and above the observational fact that John catches the ball, what one also needs to know is that without John’s action the ball’s journey would have ended up inside the pocket.

Therefore, in order to be able to formulate the negative causal claim, one must *map and compare two scenarios*. The first scenario consists of the cue ball hitting the red ball, the red ball rolling towards the pocket, and the red ball falling into the pocket. The second scenario involves the cue ball hitting the red ball, the red ball rolling towards the pocket, John catching the red ball, the red

ball being in John's hand.⁹ It is only the act of mapping and comparing, which presents John's catching the ball as the cause of the ball not being in the pocket (i.e. not scoring).

In fact, it is only the act of mapping and comparing these two scenarios, which brings John's catching and the pocket together so that they can be parts of the same causal claim. Originally, they occur in distinct causal sequences. On the one hand, in that sequence where John catches the ball, the ball never gets to the vicinity of the pocket, so the fact whether scoring or not scoring should have happened cannot be determined. On the other hand, when the ball falls into the pocket John's hand never interferes, his catching never occurs.

That is, the negative claim in question is such that it connects an event¹⁰ present in one of the scenarios (John catching the ball) with the negation of an event present in the other scenario (the ball not falling into the pocket, i.e. not scoring). The events connected by the negative causal claim, thus, are not part of a single causal sequence and, therefore, are not related of a single causal relation. In other words, the negative causal claim 'John's catching the ball causes not scoring' does not express a case of genuine causation.

Moreover, the scenarios on the basis of which one is able to formulate the negative claim are grasped via positive descriptions. The two causal sequences ('cue ball hitting the red ball, the red ball rolling towards the pocket, the red ball falling into the pocket' and 'cue ball hitting the red ball, the red ball rolling towards the pocket, John catching the red ball, the red ball being in John's hand') do not seem to involve absences at all, and thus are apt for being picked out by positive causal claims. Neither are there palpable disconnections between the causes and the effects of these causal sequences as characterised by positive claims, so accounting for them does not pose any obvious problem for the physical connection view.

This is the *mapping framework*. According to it, negative causal claims are derived, or generated from one's knowledge of distinct scenarios via mapping and comparing, where the individual scenarios themselves can be characterised by purely positive terminology posing no apparent problem for the physical connection view. Therefore, the mapping framework allows the connectionist to stop the argument from negative causation: if negative causal claims do not pick out genuine

⁹ Of course, we usually formulate negative causal claims without actually observing both scenarios. Rather, based on our previous knowledge of similar situations, we presuppose what would happen in the other scenario if it actually happened. Still, the negative claim itself is generated from two distinct scenarios. These scenarios consist in causal chains or sequences with causes and effects occurring one after another. Note that by relying on such causal sequences our argument does not beg the question against negative causation. We remain neutral about what can play the role of a cause or an event here. In this sense, an absence 'occurring' at one stage of a causal sequence would result in a case of negative causation.

¹⁰ One can replace 'events' here with one's favourite candidate for being the relate of causation—we would like to remain neutral about what exactly the relate of causation are throughout the paper.

causal relations but occur only in causal discourse as a result of mapping and comparing positive causal claims referring to different scenarios, which themselves are compatible with the physical connection view, then the mere observation that there are negative causal claims in our causal talk does not entail the failure of the physical connection view.

6. Using mapping iteratively and together with co-reference

As the anti-connectionist might want to pinpoint, however, it is not immediately obvious that the mapping framework is able to account for all cases of negative causation in a way, which is helpful for the connectionist. Consider, for example, the case of the gardener and the plant, the very problem, which disqualified the co-reference framework as a general solution.

On the face of it, the negative claim ‘the gardener not watering the plant causes the plant’s death’ is generated from the mapping and comparing of two scenarios, which might be best characterised as consisting of the following causal sequences: (S1) the gardener not watering, plant dying; and (S2) the gardener watering, plant living. Similarly to the argument of the previous section, knowing that the plant dies when the gardener does not water it is insufficient for concluding the negative causal claim, since it might have been the case that the plant would have died even if the gardener had watered it. So one needs (S2) as well, which makes it a case covered by the mapping framework.

However, the problem now is that one of the causal sequences, namely (S1), involves absences: the gardener *not* watering, and the plant dying, i.e. *not* living. In fact, these are the very same absences, which are connected by the original negative causal claim. Since they occur in the same causal sequence, it seems that contrary to what the mapping framework states, the negative causal claim generated do pick out a genuine causal relation. Moreover, (S1) does pose an obvious problem for the physical connection view, since not watering and not living seem to be physically disconnected. Therefore, as the anti-connectionist might want to argue, the mapping framework cannot help the connectionist to evade the conclusion of the argument from negative causation.

Though the connectionist seems to be in trouble here, in fact she has two different answers ready at her hands. She can either rely on the mapping or on the co-reference framework in order to drive her point home. Relying on the mapping framework would mean drawing attention to that (S1) as formulated above strictly speaking is not a scenario that one observes, but rather itself is generated from mapping and comparing concrete causal sequences, three in this case. So for example, (S1) can be derived via mapping and comparing (S1a) where the gardener takes a nap and falls

asleep, (S1b) where the water is in the bottle and remains there, and (S1c) where the concentration of water becomes lower inside the plant and the plant withers. None of these three causal sequences involve absences or evoke obvious disconnections—that is, they are apt for serving the connectionist’s purpose.

Similarly, the connectionist could rely on the co-reference framework and argue that the negative description of (S1) co-refers with, or to be more precise, is a shorthand for the disjunct of the positive claims describing (S1a), (S1b), (S1c). In this latter case, the original issue with the co-reference framework (namely that there is an obvious disconnection between the events picked out by the three positive claims) is not a problem anymore, since it is the mapping and comparing of (S1a), (S1b), (S1c) and (S2) which generates the negative claim, and thus a physical connection between the events occurring in the different sequences is not a requirement.

Some might favour the mapping, others perhaps the co-reference framework here.¹¹ In fact, it doesn’t really matter whether one considers (S1a)-(S1c) as the mapping base of (S1) or as a disjunct, which co-refers with (S1); what does matter, is that (S1a)-(S1c) and (S2) pose no obvious problem for the physical connection view and they together can serve as the basis of generating the negative causal claim ‘the gardener not watering the plant causes the plant’s death’.

In short, we claim that by using the mapping framework iteratively and together with the co-reference framework it becomes possible for the connectionist to account for the occurrence of negative causal claims in causal discourse in terms of the resources provided by these frameworks and the physical connection view. In other words, no alternative account of causation is required to cover cases of negative causation.

That is, the full answer the connectionist can give to the argument from negative causation is the following. For every negative causal claim either there is a single positive causal claim co-referring with the negative claim, or the negative claim is generated from mapping and comparing different scenarios ultimately characterisable by positive causal claims. None of the positive claims in play here evoke obvious disconnections, i.e. unless evidence to the contrary is brought forward they can be treated as compatible with the physical connection view. Thus, the mere fact that there are negative causal claims in causal discourse does not disqualify the physical connection view.

On the basis of all this, the connectionist can argue that (AfNC)** is still not the most accurate version of the argument from negative causation. The anti-connectionist argument, in order to

¹¹ As we see it, rather than using either the mapping or the co-reference framework straightforwardly, one formulates (S1) by determining which are the relevant factors in the scenario consisting in the three causal sequences (S1a), (S1b) and (S1c) compared to the other scenario (S2). See Section 7 for more details about how relevance claims are generated via applying both the mapping and the co-reference frameworks together.

be able to reach its desired conclusion needs to exclude the applicability of the mapping framework as well. With this amendment the argument from negative causation has the following structure:

(AfNC)***

- (P1) There are negative causal claims (where either the cause or the effect (or both) is an absence) in causal discourse.
- (P2) Absences are not connected physically to either occurrences or other absences.
- (P3) Negative causal claims *are not generated via mapping and comparing different scenarios, which themselves can be accounted for in terms of the physical connection view, but rather* express cases of genuine causation.
- (P4) No positive causal claim—which can be accounted for in terms of the physical connection view—picks out the same causal relation picked out by a negative causal claim.
- (C) The physical connection view fails to provide a general account of genuine causation.

Given this analysis of the argument from negative causation to the conclusion that the physical connection view fails, the connectionist can stop the argument by claiming that though it is a valid argument, in fact it is unsound: (P3) and (P4) together are never true. If certain negative causal claims do express a case of genuine causation then there is always a co-referring positive causal claim picking out the same relation, and when there is no single positive claim co-referring with the original negative claim, then the negative claim is always derived from the mapping and comparing of further positive claims describing different scenarios. And since the positive claims in question pose no obvious difficulty for the physical connection view, its failure as a general account of genuine causation does not follow from the observation that there are negative causal claims in causal discourse.

7. Derived causal notions

The major claim of this paper is that the negative causation based argument against the physical connection view is inconclusive. Besides this, however, a further important issue emerges from what has been said so far. As the mapping framework tells us, the notion of negative causation (at least in certain cases, i.e. when one considers the non-(P3) horn of the connectionist reply) is a derived notion. That is, since negative causal claims are derived from the mapping and comparing of

positive claims, those who are subscribed to a particular causal account compatible with the positive claims in question, could argue that from their account, given the mapping framework, the notion of negative causation can be derived. It is not only negative causation, though, which turns out to be derived in this sense. Other popular causal notions share the same fate.

Consider relevance claims. In the case of the void, for example, even if one acknowledges that the negative claim ‘the void kills’ is a shorthand co-referring with the positive claim ‘inner pressure significantly higher than outer pressure kills’, one could still argue that the crucial factor here is the presence of the void. Inside pressure results in the air departing the lungs, the body bursting etc. only if there is *nothing* outside, only if there is *nothing* that could exert a force to resist the inner pressure. The void kills by not exerting counter-pressure from the outside of the body. That is, one could argue that referring to the void reflects our intuitions better in picking out the cause, because it the presence of the void, which is the unexpected event, which is normally not the case. Bodies usually do not burst, because their inner pressure is compensated by atmospheric pressure. The fact that the inner pressure is *uncompensated* is the relevant factor which (when compared to normal cases) stands out as ‘abnormal’. That is, one might conclude, in causing the effect in question (death) what is relevant, is the lack of everything, i.e. the *presence* of the void.

The mapping and co-reference frameworks can account for how such relevance claims are formulated. As the mapping framework draws attention to it, to be able to conclude on the relevance claim one needs to know two different scenarios: first that when a body is placed in the void the body dies, and second that when a body is placed under atmospheric pressure the body lives. Knowing both of these scenarios is indispensable for formulating the relevance claim—without knowing the first, one wouldn’t know that being placed in the void and dying are somehow related; and without knowing the second, one couldn’t know that the body would not have died if it had been placed under atmospheric pressure. Knowing both scenarios—and mapping and comparing them—is what presents being placed in the void as the relevant cause of dying.

On the face of it, the first scenario ‘when the body is placed in the void the body dies’ involves absences. However, similarly to accounting for (S1) in Section 6, it is possible to account for this scenario in purely positive terms. For example, as the co-reference framework points out, the initial characterisation of the first scenario above (and similarly, that of the second one as well) is only a shorthand for a much more detailed and accurate description.

According to this more accurate description, in the scenario, which originally is characterised as ‘where a body is placed under atmospheric pressure the body lives’, there are objects surrounding the body (say, air with atmospheric pressure) which exert a force on the surface of the body thus

resisting the inner pressure yielding an equilibrium state (body living). In this case, the kinetic energy of the entities inside pushes the surface of the body outwards, whereas the kinetic energy of the entities outside pushes the surface inwards. The outer and the inner pressure—acting in tandem on the surface—result in an equilibrium state where the surface stands stable and inner processes can take place (body living).

Similarly, in the scenario, which originally is characterised as ‘where the body is placed in the void the body dies’, there are no objects surrounding the body, thus no counter-pressure resists the inner pressure, which results in the bursting of the body (death). In this case, the kinetic energy of the entities inside pushes the surface outwards. The surface is constantly pushed outwards, the body bursts, and the inner processes break down (death).¹²

The void (lack of entities outside) is not making its appearance in any of the two distinct scenarios. One is able to formulate the relevance claim (which is also a negative claim) only if one maps the two scenarios onto each other and compares them, thereby picking out the differences between the two scenarios. It is precisely the act of mapping and comparing these different scenarios what shows the presence of the void (i.e. that there is nothing outside of the body) as the *relevant* factor in causing death. That is, the relevance claim saying that it is the presence of the void what is relevant in killing is derived from the comparison of two positive stories: an ‘inner pressure – outer pressure – equilibrium’ story and an ‘inner pressure – expansion – more expansion – burst’ story. No absences are involved in either of the two scenarios compared; neither do they evoke physical disconnections. That is, again, the physical connection view is compatible with these positive causal claims.

Very similarly, contrastive claims can also be seen as derivative. Consider, for example, the claim that ‘being placed in the void rather than under atmospheric pressure causes the body to burst rather than to live’ (cf. Schaffer, 2005). It is quite straightforward how the mapping framework presents such a claim as being derived from mapping and comparing the scenarios where the body is placed in the void and bursts, and where the body is placed under atmospheric pressure and lives. The same line of thought as above applies: one needs both scenarios in order to be able to conclude on the contrastive claim, and both scenarios can be characterised by such positive causal claims, which pose no obvious problem for the physical connection view.

¹² Note that we do not argue here that one needs to know such details in order to be able to formulate the relevance claim. On the contrary, we acknowledge that one can easily conclude that ‘it is the presence of the void what is relevant in causing the death’ on the basis of a much more coarse-grained knowledge (cf. the original characterisation of the scenarios). All that is claimed here is that it is possible to give a more accurate description (a co-referring positive claim in this case) of the scenarios in question.

That is, within the mapping framework, over and above negative causation, such notions as relevant factors and contrasts—which are fundamental notions of novel alternative accounts of causation (cf. Glennan (2010) and Schaffer (2005) respectively)—also turn out to be the results of mapping and comparing different scenarios. This, then, suggests that it is not just negative causation itself that is a derived or generated notion, but popular alternative accounts of causation (such as the contrastive account), and notions like causal relevance, themselves seem also to be derived similarly.

We do not have enough room here to further explore all the consequences of this conclusion, but one observation immediately follows. If contrastive causation and causal relevance are also derived notions, then those being subscribed to an account of causation which is compatible with the positive claims from which these derived notions are generated from could argue that their account enjoys a certain sense of priority over those, which turn out to be derivative.

For example, the connectionist could run the following argument: (i) the physical connection view is compatible with the positive claims in question; (ii) relevance and contrastive claims can be derived from the physical connection view given the mapping and co-reference frameworks; (iii) physical connection claims cannot be derived from the relevance-based or the contrastive accounts given the mapping and the co-reference frameworks; (iv) therefore, the physical connection view is more fundamental than the relevance-based and the contrastive accounts.¹³

8. Conclusion

The fundamental aim of this paper is to show that there is no straightforward route from recognising that negative causal claims populate our causal talk to the conclusion that the physical connection view fails. The paper aims to eliminate the point made by the opponents of the physical connection view, namely that the occurrence of negative causation counts in favour of an alternative view. Contrary to this, the present paper argues that negative causal claims can be accommodated by the connectionist.

The paper offers two strategies to show this. First, one can think of negative causation within the so-called co-reference framework where, for at least some negative causal claims, it is possible to give co-referring positive descriptions which themselves pose no obvious problem for the physical connection account. If this is so, then those who want to argue against the physical connection

¹³ If the connectionist could provide a persuasive case for (iii) that would lend significant support for the strong conclusion of (iv).

view from negative causation need to show at least that the examples they rely on do not fall into this category.

Second, the paper introduces the so-called mapping framework, according to which negative causal claims are generated from mapping and comparing distinct scenarios. Since these scenarios can in principle be accounted for in terms of the physical connection view, explaining negative causation—in the sense of explaining how negative causal claims get formulated—becomes possible for the proponents of the physical connection view.

The co-reference and the mapping frameworks together cover all cases. Since the analysis of any arbitrary case of negative causation within these frameworks seems to be compatible with the physical connection view, it is a safe bet to say that the co-reference and the mapping frameworks help the connectionist to stop the argument from negative causation. At least, the burden of proof is on the anti-connectionist side.

The mapping framework grasps an important fact about formulating negative causal claims, namely that one always needs to know two or more positive scenarios in order to be able to formulate a negative claim. On the other hand, the co-reference framework is not without virtues either. It grasps the fact quite well that some negative causal claims are often utilised as shorthands—they are regular players in e.g. scientific discourse even if the underlying positive scenarios are well known, simply because they provide a short and simple way of putting what happens. In this sense, the mapping framework tells us how negative causal claims are formulated, whereas the co-reference framework has something to say about why we often stick to using negative claims.

The mapping framework has some further interesting consequences. First of all, it provides an easy answer to the problem of too many negative causes. Roughly, the problem is that if one acknowledges that the gardener's not watering the plant is a cause of the plant's death, then it is hard to see why e.g. the Queen's (and therefore anybody's) not watering the plant would not count as a cause of the plant's death (cf. Beebe, 2004). The answer is given by the fundamental tenet of the mapping framework that negative causal claims are formulated on the basis of one's actual knowledge of two scenarios. That is, the mapping framework disqualifies all those persons' not watering the plant as good candidates for being a cause of the plant's death who are not part of the actual scenarios from which the negative causal claim is generated. If one, for example, happens to be in a position that one actually sees the Queen watering a plant, then one would be able to map and compare two scenarios, which would support that the Queen's not watering can be a cause.

Second, the mapping framework is able to incorporate Dowe's counterfactual analysis of negative causation. What Dowe calls the possibility of genuine causation is a causal relation in one

of the scenarios. Consider Dowe's examples of prevention ('the father's grabbing the child prevented the accident') and omission ('the father's failure to grab the child caused the accident'). (Cf. Dowe, 2001, pp. 221-222) According to the mapping framework, both of these negative causal claims are formulated on the basis of mapping and comparing two scenarios: first, a 'father grabs the child, the child walks with the father' scenario and, a second scenario consisting in two parallel causal sequences, a 'father being lost in thought, father wandering next to the road' and a 'child running over the road, accident happening' sequence. The mapping framework points out that the prevention and the omission claims are generated via picking out different events from the two scenarios. The prevention claim seemingly connects the occurrence of an event belonging to the first scenario (father's grabbing the child) with the absence of an event belonging to the second scenario (accident not happening), whereas the omission claim seemingly connects the absence of the event belonging to the first scenario (the father's failure to grab the child) with the occurrence of the event belonging to the second scenario (accident happening). These are two ways of grasping the difference between the same two scenarios mapped onto each other.¹⁴

Acknowledgements

The authors wish to thank for Jonas Christensen and Jesper Kallestrup for their helpful comments on earlier versions of this paper. One of the authors (PF) is grateful to the Eötvös Scholarship from the Hungarian State and the University of Edinburgh College of Humanities and Social Sciences Scholarship for their generous support.

Reference

- Armstrong, D. (1999). The Open Door: Counterfactual versus Singularist Theories of Causation. In H. Sankey (Ed.), *Causation and Laws of Nature* (pp. 175-185). Dordrecht: Kluwer.
- Beebe, H. (2004). Causing and Nothingness. In L. A. Paul, E. J. Hall & J. Collins (Eds.), *Causation and Counterfactuals* (pp. 291-308). Cambridge: MIT Press.
- Craver, C. F. (2007). *Explaining the brain: mechanisms and the mosaic unity of neuroscience*. Oxford: Clarendon Press
- Dowe, P. (2000). *Physical Causation*. Cambridge: Cambridge University Press.
- Dowe, P. (2001). A counterfactual theory of prevention and 'causation' by omission. *Australasian Journal of Philosophy*, 79, 216-226.

¹⁴ Moreover, we think that the mapping framework is able to account for the intuition of difference and the intuition of genuineness in terms of what one knows about the two scenarios mapped. However, since again we do not have enough room here to carry out the necessary analysis, accounting for Dowe's counterfactual theory in terms of the mapping framework in full detail must be the subject of an independent paper.

- Dowe, P. (2004). Causes are physically connected to their effects: why preventers and omissions are not causes. In C. Hitchcock (Ed.), *Contemporary Debates in Philosophy of Science*. Oxford: Blackwell.
- Fair, D. (1979). Causation and the flow of energy. *Erkenntnis*, 14, 219-250.
- Glennan, S. (2010). Mechanisms, Causes, and the Layered Model of the World. *Philosophy and Phenomenological Research*, 81(2), 362-381.
- Lewis, D. (2004). Void and Object. In J. Collins, N. Hall & L. A. Paul (Eds.), *Causation and Counterfactuals* (pp. 277-289). New York: MIT Press.
- Mackie, J. L. (1980). *The cement of the universe: a study of causation*. Oxford: Clarendon Press.
- Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
- Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science*, 67, 297-312.
- Salmon, W. (1998). *Causality and Explanation*. Oxford: Oxford University Press.
- Schaffer, J. (2000). Causation by Disconnection. *Philosophy of Science*, 67(2), 285-300.
- Schaffer, J. (2004). Causes need not be physically connected to their effects: The case for negative causation. In C. Hitchcock (Ed.), *Contemporary Debates in Philosophy of Science* (pp. 197-216). Oxford: Blackwell.
- Schaffer, J. (2005). Contrastive causation. *The Philosophical Review*, 114(3), 297-328.
- Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.