



Backwards Causation in Social Institutions

Kenneth Silver¹

Received: 16 November 2021 / Accepted: 20 August 2022
© The Author(s) 2022

Abstract

Whereas many philosophers take backwards causation to be impossible, the few who maintain its possibility either take it to be absent from the actual world or else confined to theoretical physics. Here, however, I argue that backwards causation is not only actual, but common, though occurring in the context of our social institutions. After juxtaposing my cases with a few others in the literature and arguing that we should take seriously the reality of causal cases in these contexts, I consider several objections. These objections involve whether the cases should be reinterpreted, whether they are properly within the institution, whether they involve necessitation or else Cambridge changes, whether and how they involve changing the past, and whether this should call us to question institutional reality. I end by suggesting that it is a virtue of our institutions that they allow backwards causation, that this is a kind of technology that they are often built to incorporate.

1 Introduction

Causation is typically taken to be a two-place, irreflexive, asymmetric, and transitive relation of metaphysical dependence between two events, where the cause temporally precedes the effect. Though this is the standard understanding of causation today, in the literature nearly every aspect of this conception has been challenged. (Is it *really* a relation between events, rather than facts? Are we *sure* that it is transitive?) What is less frequently challenged, however, is the claim that causes precede their effects.

✉ Kenneth Silver
kennethmsilver@gmail.com

¹ Trinity Business School, New Building, Trinity College, Dublin 2, Ireland

For a case of *backwards* causation, then, the effect temporally precedes the cause. And the debate over backwards causation tends to concern whether it is possible or even conceptually coherent. The notion that causes precede their effects was part of the definition of causation for Hume (1748/2000),¹ and it was among the paradigmatic synthetic a priori truths for Kant (1783/2004). More recently, Flew (1954, 1956, 1956–7), Black (1956), Mellor (1998), and Ben-Yami (2007, 2010) have argued against the possibility of backwards causation. Meanwhile, while there are proponents who argue that backwards causation is at least *possible* (see Dummett, 1954, 1964; Tooley, 1997; Roache, 2009; Garrett, 2014, 2015), no cases are found.

If there are any actual cases, we are typically told that they will be in the far-flung realm of theoretical physics. There, perhaps tachyons travelling faster than light can backwardly cause (Faye, 2018), or perhaps positrons are just electrons that move backwards through time (Feynman, 1949). There may be other models within quantum mechanics that permits it besides (Corry, 2015; Dowe, 1997). What seems clear from these authors is that even if backwards causation is possible, we are not likely to see it on our way to the grocery store.

Given the focus in the causation literature on the physical causation of billiard balls and the like, I think this bewilderment over the possibility of backwards causation is unsurprising. It is hard to imagine medium-sized physical objects actually moving in reverse. And given how unfamiliar and unintuitive the realm of theoretical physics can be, it is also unsurprising that we would rest our hopes of backwards causation there. However, I do not think that we must look so far for possible cases of backwards causation. In the world that we inhabit—the social world—apparent cases of backwards causation are all around us.

Here, I want to highlight and make the case for the ubiquity of backwards causation in the context of our social institutions. In the first section, I will present several apparent cases of backwards causation and situate them within the social world. After considering what few precedents exist for discussing these kinds of cases, I will discuss what we should say to ameliorate the sense that there is something off about them, given that they arise within institutions of our own creation. In the second section, I will consider a series of other objections to the cases. These appeal to how the cases are interpreted, the features of causation, the kinds of properties at issue, and the commitments these cases appear to involve. I conclude by affirming that these are genuine cases of backwards causation and by briefly suggesting that this is a feature of our social reality to be celebrated, not a bug to be explained away.

2 The Cases

Consider the following three cases:

¹ Understood as an antirealist about causation, there is no great surprise that causation cannot be backwards. Technically, the relation of causation is never actually instantiated; instead, it is our way of labeling occurrences of constant conjunction, with the cause as the first conjunct.

Case 1: Retroactive Enrollment

Kevin made a mistake in graduate school. One semester, say in year X, he registered for PHIL755 rather than PHIL756. Both classes are open-ended, dissertation writing classes that one registers in as a formality, but he had already ‘taken’ PHIL755, and PHIL756 was next in the sequence. The mistake was not discovered until a year later ($X + 1$) when it was time to make sure his affairs were in order to graduate. After reviewing his transcript, recognizing his error, and becoming hysterical, he shared his mistake with the department administrator. Graciously, she assumed the bureaucratic headache of fixing it. Calls were made, emails exchanged, and eventually Kevin was *retroactively* enrolled in the course. By the end of the ordeal, actions taken by Kevin’s administrator in year $X + 1$ caused him to have technically been registered for and taken PHIL756 in year X.

Case 2: Forced Forfeit

The Tigers are a wrestling team that competes for and wins the National Championship in year X. Unfortunately, however, they cheated by using rigged weight scales. When this is uncovered in year $X + 1$, the league commissioner decides to strip the team of their title in the wake of the scandal. It is written down that the team forfeited their last match, and the title went to their final opponents. By the end of the ordeal, the announcement of the commissioner in year $X + 1$ causes the team to forfeit the match that occurred in year X.

Case 3: Annulment

Billy and Suzy decide that after being friends for many years they must be meant for each other, so they get married (say, on day X). Unfortunately, they quickly realize that they should *not* be married; they’re both just too competitive. As it happens, they are also both Catholic, and both hope to remarry someone else someday. Luckily, they petition the church successfully and are granted an annulment of their marriage (on day $X + 100$). In the eyes of God, the marriage never occurred. So, actions of the diocese on day $X + 100$ cause the erasure of the marriage of Billy and Suzy on day X.

Now, before beginning to discuss these cases and what they suggest in earnest, it is worthwhile to take a second to, well, beg the reader not to get off the boat too soon. I think these cases present genuine instances of backwards causation, but it would not be an uncommon reaction for a reader to balk strongly at the suggestion. It is not lost on me that there is something a little too ‘cute’ about the cases. There is a strong temptation to think that they can be immediately explained away. For all that I say below, perhaps they can be. But what I will do here to get ahead of this strong reaction is to point out that there may be something helpful in this exercise. I will not implore the reader to suspend judgment until the case has been made and defended—I doubt this is possible—but I think it survives a surprising number of immediate worries. And responding to each of the worries, as we will see, deepens

our understanding of what goes on in these cases. So, if nothing else, this justifies looking beyond knee-jerk reactions.

Focusing on these cases, then, the first thing to notice is that there is nothing special about them. They are common in our social lives, even banal.² They are not miraculous natural occurrences that only the brightest physicists could discover; they occur as a matter of routine. Thus, if they do involve backwards causation, then backwards causation will turn out to be not merely possible, but actual, though occurring in the context of social institutions.

In the literature, the rough idea of these kinds of cases is not without precedents. Peijnenburg (2006), for example, offers a view of acting in ways that retroactively set features of our past conduct. The character with which we engaged in some particular action can depend, she maintains, upon our subsequent behavior. For example, whether an act genuinely reveals your bravery depends upon the pattern of your conduct after the act. This may allow us to have some say about our past. However, Peijnenburg's understanding of her cases is closer to changing the *meaning* or maybe the nature of concrete events in the past, not changing which events in fact occurred.³

In contrast, Torrenco (2018) has recently also recognized how social institutions create the space for what he calls 'retroactive enactments', but where these do involve changing the events that occurred in the past. Torrenco himself is primarily concerned with how their possibility creates a problem to be handled (in his case, by accepting a version of the B-theory of time) as opposed to recognizing it as something to be explored. Given this, he fails to appreciate the position of these cases in the literature on backwards causation, their ubiquity, the various concerns they raise, and indeed the advantage of them to be highlighted below. What Torrenco captures well, though, is the reliance of cases like these on institutional reality.

Confronting the above cases, this reliance on social institutions is apparent, and it admittedly does make it feel as if something shifty is going on. Who's registered when and how can change and what it counts for is so malleable, we might think, that it is somehow illegitimate to consider it on a par with the kind of causation that we are talking about when we are talking about billiard balls. After all, we invented these notions and change them around at our whim. These cases feel invented because they are invented; they are a product of institutions, conventions, and practices that we developed. The question, though, is whether the mere fact that they involve institutions we created undermines their claims to causation. And I think it does not.

Start with idea that of so-called institutional facts, or facts that involve institutions or require certain institutional elements to obtain. This has roots in Anscombe (1958) and Searle (1995), and many have come to accept institutional facts as

² This is against Dummett (1964: 358) and in agreement with Roache (2015) in maintaining that there is nothing peculiar about being agents who take ourselves to be able to affect the past. However, important for Dummett and Roach is that they are considering agents who take themselves to be capable of performing actions that cause the past to be *as it was already*, whereas the cases here involve changing the past. This introduces its own problems, which will be discussed below.

³ See van Putten (2006) for criticism of the piece along these lines.

legitimate facts. Now, there are many questions about their nature: How they are related to non-institutional facts? To brute facts? To what extent do they depend on our minds? How are they established? There is no question, however, that an acceptance of these facts has been institutionalized at least within social ontology. If we grant this much, we should also grant that there can be institutional events, where these may be events that occur in the context of institutions, or the event-tokens that are the subject of institutional facts.

Once we have institutional events, we can also easily recognize cases of uncontroversial causation between institutional events. Your pawn-movement caused my bishop's retreat. Tara's registering for the class caused the professor to email her the syllabus. These cases of apparent causation within institutions easily satisfy whatever test we arrange to judge causation. My bishop's retreat counterfactually depends on your pawn-movement, say, and Tara's registering for the class raises the probability of the professor's emailing her the syllabus. Though these are in the context of institutions, they are paradigmatic instances of causation. Our attributions of causation in these cases are not metaphoric, but literal, and it would seem to be a kind of scientism to claim that these are not instances of *real* causation.

(Our paradigmatic case of real causation is the causation between billiard balls, but we can notice that of course billiard balls themselves are institutional objects. Moreover, there is arguably even backwards causation in the context of billiards! Whereas Dummett (1954: 29–30) calls on us too appreciate how complex it would be to try to model shots of billiard balls in reverse, we would not even try to appeal to complex versions of those laws when explaining what happens when we scratch. When that happens, the rules dictate that we bring the condition of a player *back to an earlier state*. My scratching sets me back causally, undoing my earlier successful shot.⁴)

So, if we are willing to take causation in the context of institutional reality seriously, then we are in a position to model the apparent backward causation in these cases. Take Retroactive Enrollment. Originally, in year X, there is an event (a) 'Kevin's taking PHIL755'. Taking a class is an extended event with which most of us are familiar. Though this class has no classroom or set assignments, it does occur over a span of time and is located (such as it can be) at the university.⁵ Though this

⁴ The point is meant to drive home this idea of the ubiquity of these cases and how they may even affect our core understanding of causation, but I do not deny that there may be other ways to interpret this particular case. I think it is well-understood as a setback, where setbacks are characterized causally in terms of a player's being made to inhabit an *earlier* position. However, it may be interpreted as simply *the next* state to occur given the one before it, where that state is worse-off for the player as a punishment. That said, there may yet be even more radical interpretations. Perhaps causation in the context of games like billiards is a candidate for causation understood as fundamentally teleological (*à la* Hawthorne & Nolan, 2006). On that interpretation, each move is causally understood in terms of how it advances the player towards a goal (winning the game). In that case, the result of a scratch would be viewed also as a kind of backwards causation, but where this is interpreted as being made *farther* from the goal at hand.

⁵ It is an interesting question precisely where this event occurs or what objects instantiate it. We cannot address this sufficiently here, but I refer readers to a recent exchange on an adjacent discussion on the materiality of establishments (Hindriks, 2020; Korman, 2020). Given that exchange, there is a fleeting thought that if Korman is right, and establishments like universities as objects are abstract, then perhaps some events involving them are abstract as well. Perhaps Kevin's taking PHIL755 or PHIL756 are events

is an event that raises certain questions, it is also a familiar event, one we are likely to countenance within our social reality.

In year $X + 1$, we saw that there is a second event (b) ‘The administrator’s phone call’. The suggestion is that this event is a cause of a third event, (c) ‘Kevin’s taking PHIL756’, which occurs back in year X . Once the administrator has made the phone call, there is no more event (a) of ‘Kevin’s taking PHIL756’ in year X .

Now, (c) would not have occurred if (b) had not occurred. (b) also raises the probability of (c)’s occurrence. (Assume it is no sure thing that whoever is on the other end of the phone will augment the computer system, but the administrator’s phone call makes it highly likely.) When (b) is spelled out suitably, it will also provide what seems to be a satisfying, diachronic explanation for (c). In short, (b) will count as a cause of (c) on many accounts of causation widely discussed, though it will be backwards in time.

More specific events could be pointed to in the other cases to make much the same point. And more specific accounts of causation could be drawn out to deliver this verdict. But the point seems clear. Or, rather, I suspect that the source of a reader’s skepticism is not to be found in the account of causation chosen or even how the events in these cases are individuated. Given that, I want to move on to what I take to be likely objections.

3 Objections

Given the recent rise of social ontology, hopefully these cases (and what they purport to show) cannot be easily dismissed. Nevertheless, one may think that there must be a quick explanation of how the cases so described do not depict backwards causation. Below are a number of such attempted explanations or challenges with responses, ending with a reflection on what this means for institutional reality.

3.1 The Cases Can and Should Be Reinterpreted to Not Involve Backwards Causation

First, someone might argue against how I have interpreted the cases. In each case, I said that an event at a time caused the occurrence of another event in the past. However, given how these institutions are all a matter of how we treat things, we may attempt to reinterpret the cases as follows: Events unfold in a certain way; then, at a later time, we decide to act from then on as if they had unfolded a different way.

Footnote 5 (continued)

without embodiments, and perhaps it would be more plausible to readers that backwards causation can occur with physical events in the present causing abstract past events. At least, such causation would not involve moving around physical *stuff* in the past. I float this as an idea, but I will not pursue it further, since I am inclined to think that establishments are material, and Kevin’s taking PHIL756 occurs where he is over the year, though it may be ontically indeterminate which segment of his conduct over that year constitutes his participation in that class.

In the first case, for instance, I have suggested that the actions of the administrator caused Kevin to have been registered for a different class. Instead, we might say that the actions of the administrator make it such that from that day forward we consider Kevin to have been registered for a different class than the one he was in fact registered for in the past. And similar re-interpretations may be given for the other cases.

It would not be hard to reinterpret Retroactive Enrollment in this way, especially if we assumed that the rules of the university were quite strict. Sure, perhaps the administrator *can* pull some strings to massage Kevin's transcript to look right for graduation, and perhaps she should do this both for Kevin and in the interest of the university. But we may nevertheless maintain that it is illegitimate, that Kevin did not in fact take PHIL756, that he never took it, especially not back in year X. We may think that taking institutional reality seriously means being willing to not just throw up our hands and say that it's all made up; it means acknowledging that Kevin really did not take PHIL576 in year X as per the university rules.

In response, I will just grant that the case certainly can be reinterpreted that way if this happens to be the way that this university in particular actually works. But does it *have* to work that way? It seems just as likely that the university has, by its own rules, given authority and discretion to its department administrators regarding enrollment, both past and present.⁶ Perhaps these mistakes are so common and pedagogically inconsequential that the university's rules explicitly allow for being de-enrolled and re-enrolled in these meaningless dissertation-writing classes at any point in the student's time in graduate school. In any case, it is at least possible that the university works this way, thus blocking the problematic reinterpretation.

This kind of reinterpretation is even more apparent in the case of Annulment. Although I have spelled out the case as one where a couple that is actually made to never have been married, it has been pointed out to me quite fairly that this is not how annulment is properly understood in the Catholic faith. Instead, we may say that marriages can only be annulled when there are grounds to say that the marriage was never valid to begin with (perhaps because consent was not appropriately given). We uncover that there was no valid marriage; we do not *make* the marriage to be invalid.

If this is the right interpretation of Annulments in Catholicism, then I agree that this will not be a case of backwards causation. But this is not to say that we cannot have an institutional practice very similar to annulment that is interpreted in my way. We could imagine a society where re-marriages are illegal, marriages can end in either divorces or annulments, and the difference between them is as follows: whereas divorces make it such that couples *are no longer* married, annulments make it such that couples *never were* married. (Perhaps there are cultural reasons for pursuing one rather than the other.) So, that there are clever ways of reinterpreting my cases is insufficient to show that institutional reality is incapable of fostering backwards causation. I think that it certainly is, and in the conclusion I will mention a reason for why we should want it to be.

⁶ There are certainly procedures we are all aware of concerning how a grade can be changed after the fact, and the conditions licensing them are as codified as possible and institutionally real.

To be clear here, though, none of this is to say that our cases can be represented as either having backwards causation or not, where the apparent backwards causation is just a trivial feature of how we are representing the cases. These institutional practices *really do*, as a matter of institutional reality, work one way or the other. Of course, how we set up this institutional reality will depend on how we choose to represent the situation. After all, our institutions largely involve systems of representation. So, the fact that the cases only have backwards causation because we are representing them that way does not invalidate the cases; it vindicates them. If we have set up the institution to allow for this, then saying that this is not *really* backwards causation but a feature of how we are representing things is to fail to sufficiently take institutional reality seriously.

3.2 These Cases Do Not Feature Backwards Causation Within an Institution

A second objection begins with the sense that the administrator in Retroactive Enrollment is somehow *outside* of the system in question. Perhaps she in some way determines facts about the system, and she certainly works for the university, but we may think that events involving her conduct do not quite occur at the level of social reality relevant. Kevin's taking PHIL575 might partially cause his taking PHIL576, and his taking PHIL576 might partially cause his being awarded his diploma, but actions of the administrator do not seem to directly bear on the causal workings of the degree progression. We could make similar claims about the conduct of the league commissioner from Forced Forfeit or the diocese in Annulment. If the actions of these individuals do not have the proper institutional status, then we might not be talking about causation between institutional events, let alone backwards causation. Instead, we could perhaps say that facts about the institution depend on or are grounded in facts about what these individuals do, but where this is some non-causal determination relation.

Given the cases above, this concern does seem warranted. However, a number of responses are available. Starting with the most concessive response, I will just say that if these cases do not convince, then one discussed below should. It lacks this feature of agents acting apparently from outside of the system in question. As we will see, the case of Re-Do Checkers involves (or could involve) backwards causation fully within the game at hand. Though I do like the above cases, there are in principle other cases of apparent backwards causation within institutions that suffice.

To be less concessive, I would also challenge the characterization of the actions of these agents as not properly within the systems in question. Part of our answer to the last objection involved recognizing how the administrator may need to be institutionally empowered to make these kinds of decisions. Part of the rules of the institution that defines the degree make room for agents like the administrator to do things like switch enrollment around for good reason. Similarly, only the league commissioner, in accordance with league guidelines, is empowered to strip a team of a title. So, while there is something exogenous about the agents in the cases above in that

they represent interventions⁷ in the order of things, they are empowered by and a part of the system that they affect.

We may still wonder, however, about whether what occurs in these cases is properly thought of as *causation*. Why not think that these cases feature some kind of backwards determination, but not backwards causation? This is worrying if we take it that what is genuinely metaphysically puzzling concerns the idea of backwards causation in particular.

On this score, the first thing to note is that what is happening in these cases seems significant even if there is some good reason not to think of it as ‘causation’. The cases involve influencing an event in an institution’s past. There is at least backwards determination, where this is still a two-place, irreflexive, asymmetric, and perhaps even transitive relation of metaphysical dependence between two events. Yes, the literature focuses on backwards causation, but the same scholars arguing against it would surely have as much of a problem with a past event’s depending on, being partially grounded in, or being explained by some event in the present. (We would be just as surprised to hear that the event of a window’s shattering yesterday was dependent upon a rock thrown today.) What is troubling is backwards diachronic dependence as such, whether or not we deem it causation.

Even if we thought the relation in question was better thought of as one of grounding, we could still argue that the cases involve backwards causation by taking on board a view on which grounding is a kind of causation (Wilson, 2018). But the broader point is that the relation at issue is at least relevantly like causation. It’s a relation of what Schaffer calls ‘directed dependency’ (2016: 83), as it backs explanations. And what matters here is that the literature does not seem to accept instances of events in the present metaphysically explaining past events.

With all that said, I do think that the cases are cases of causation. We are not only talking about metaphysical dependence, but a relation of dependence between events at different times. Scholars do disagree on these points, but most take grounding to be a relation between facts and causation to be a relation between events. Moreover, as discussed above, it does seem like the relation obtaining in the cases satisfies numerous accounts of causation in the literature. The events in the past counterfactually depend on the events in the present. The present event raises the probability of the past event. In fact, apart from accepting a realist conception of causation that ties causation to something like energy transmission,⁸ for example, I am not sure what plausible accounts of causation would specifically rule out the cases as involving causation.

Opponents, then, would need to show that there is a separately plausible account of causation according to which the relation obtaining in the above cases does not

⁷ That they can be framed as interventions is one reason to think that the language of causation is appropriate.

⁸ Note that someone who goes in for causation in the context of institutions is likely to separately reject an account of causation in these terms, given how it in a sense prioritizes physical causation. So, failing separate arguments for why causation is physical and does not actually occur within institutions, opponents will need an account of causation that proponents of institutional causation can accept to demonstrate why backwards causation is impossible by their lights.

count as causation. Moreover, they would need to show that the right way to characterize this relationship avoids any kind of problematic metaphysical dependency moving from the present to the past.

3.3 Causes Do Not Necessitate Effects, But These Events Do Involve Necessitation

Someone may object that the relation between the events mentioned in these cases cannot be causation because these later events seem to necessitate the earlier events, and causes do not necessitate their effects. Following Hume's dictum, we may think that these events cannot genuinely be distinct if one entails the other. If the administrator changes a transcript today, then it follows that Kevin took that class last year. If the commissioner made the announcement of the verdict today, then the team forfeited in the past. These events that follow do so with such certainty that it may seem that they *must* follow—they are in some way entailed by the current events.

To respond to this concern, I suggest that we first focus on *other* events in the vicinity and notice that they are not similarly necessitating. For instance, we could consider the event of the administrator's phone call in Retroactive Enrollment, not the changing of the transcript. Or, in Forced Forfeit, we could focus on the event of the discovery of the wrestling team's cheating. The team certainly wouldn't have had to forfeit if not for the discovery, so the forfeit counterfactually depends on the discovery. The discovery also raises the probability of the forfeiting.⁹ So, on a number of tests for causation, we are led to say that the discovery is a cause of the forfeiting. But the discovery does not necessitate the forfeit. That depends on the commissioner's announcement, and that is influenced, but not necessitated, by the discovery. Even if we remain worried about the particular cases above, then, this worry will not hold for many nearby instances of backwards causation.

Still, what of the commissioner's announcement? One thing to say is that it is the most plausible causal link between the discovery and the forfeiting. The discovery causes the commissioner's announcement, and the discovery would moreover count as a cause of the forfeiting given the transitivity of causation if we understood the announcement to cause the forfeiting. This is no direct argument for causation as the relation between the announcement and the forfeiting, but it is a story that hangs together.

We can acknowledge, though, that the announcement does *seem* to necessitate the forfeiting, and this is concerning. Frankly, I am not entirely sure how we should answer this—whether we should take them to necessitate these events while still causing them or not. Luckily, there is a way of discharging our burden in these cases, and it is by seeing how similar they are to the odd cases of genuinely proximal or even simultaneous causation.

If I push a ball five feet and it moves five feet, then this is a paradigm instance of causation, but I cannot push a ball five feet *without* its moving five feet. That effect is necessitated. Of course, my *effort* does not necessitate the ball's motion, and my

⁹ See Cusbert (2018) on how past events can be chancy in a way allowing for backwards causation.

pushing (understood only as the exertion of effort) also might not necessitate the motion, but that was not the event specified. The event specified was my pushing the ball five feet, and that seems like an event with effects. There are sure to be moves to make here or bullets to bite, but the larger point is that the cases of backwards causes and effects above look closely related to each other in just the same way as proximate or simultaneous causes/effects. I grant that it is hard to think of these cases of backwards causation as ones where the causes are ‘simultaneously with’ or even proximate to their effects. After all, the priest annulled a marriage from many months ago. Nevertheless, these effects in the past are *causally* proximate with their causes. They are the very next thing caused, and so it is no surprise that their causation seems worryingly guaranteed.

3.4 The ‘Changes’ Made to the Past Are Not Genuine Changes, Only Cambridge Changes

Another objection to consider is that the events of the present are not genuinely causing *events* of the past; instead, things in the present are making ‘Cambridge’ changes to the past (*à la* Geach, 1969). Suppose I may make a marvelous cake at t_1 . Then later, at t_2 , we decide it was the best cake ever made. In the parlance of the literature, what we have done is not actually changed the past; we have simply made ‘Cambridge changes’, bestowing the Cambridge property of being labelled the best cake ever made onto my cake, but in no material way changing the cake. This is still plenty mysterious, but we will not say that it is a matter of backwards causation. So, the opponent may say that the commissioner is making a Cambridge change to the team’s performance by labelling it a forfeit. This again will be a change in some nominal sense, but not genuine causation.

In our cases, however, properties are not merely ascribed to things in the past. Those properties are properly instantiated by those things in the past; they instantiate them *in the past*; and they are *made to do so* by events in the present. If we are willing to say all of this, then it is hard to see what is missing that is present in genuine cases of causation.

For example, the wrestling team is made to instantiate the action property of forfeiting. It instantiates that property when it is competing, and it is made to instantiate that property by action properties instantiated in the present by the commissioner. But causation, when taken to be a relation between events and where events are construed along the lines of property instantiations/exemplifications (*à la* Kim, 1976), just is one property instance leading to another. We might object on the basis of *how* these property instances lead from one to the other (this was the basis of the last objection), but there does not seem to be any clear objection in terms of the property instances themselves.

Another point to make concerning this worry is that the changes in question are not extrinsic or relational in the way that Cambridge changes often are. My cake is the best cake because nobody succeeded in baking a better one. So, it has this property in virtue of its relation to the other cakes. The same cannot be said for our cases. It is not that what the team gains some additional (weightless) property

given its relation to other future events. The nature of what is done is fundamentally altered, from winning a tournament to forfeiting it. Kevin's taking of a particular class is not altered; he is made to have taken an entirely different class. A different event occurred.

Further, although Helm (1975) shows that 'Cambridge events' are not efficacious, these events in the past certainly are. They have effects and can even have effects that are themselves in the past (*contra* Torrenco *op. cit.*:248–9). It might be, for instance, that if a team forfeits their final match of the national tournament, then they must also forfeit their first match of the next season. So, even if the verdict is issued a year later, the forfeiture in the tournament is not the only thing caused. The later forfeit is *also* caused, and it can only be caused by the earlier forfeit (if, for instance, the commissioner has no grounds to directly make the team forfeit their first match of the following year). This makes it all the more apparent that the changes made are no mere Cambridge changes.

3.5 These Cases Involve Changing the Past, But the Past Cannot Be Changed

My response to this last objection, however, paves the way for another concern. As I have understood backwards causation, it seems to necessarily involve changing the past.¹⁰ However, quite a few philosophers have argued that events in the past cannot be changed. (See, *inter alia*, Smith, 1997, 2015; Iacona, 2016; Edward, 2015; Baron, 2017; Andreoletti & Torrenco, 2019.) The thought is that there is something incoherent about changing the past, as it would involve the past's being one way *and then* being made to have been different than it was. That does sound odd. If this reduces to saying that an event at one time has a property and lacks that property, then the cases at hand seem to commit us to a contradiction.

One way of responding to this concern would be to accept that backwards causation does involve changing the past, and then finding ways to meet those arguing against it head on. (For support along these lines, see, *inter alia*, Oddie, 1990; Goddu, 2003, 2011; Loss, 2015; Barlassina & Del Prete, 2015; Torrenco, *op. cit.*). I am inclined to think that this is the right approach, although seeing it through to satisfaction goes beyond the scope of this paper. What's interesting for us here, though, is not that backwards causation as I have characterized it *might* involve changing the past. Instead, what's interesting is that my understanding of backwards causation does not *require* changing the past.

To see this, consider a fourth case:

¹⁰ Notice that it is this feature that also explains why we don't seem to face the bilking worry that has preoccupied much of the literature. That objection to backwards causation concerns *present* events being caused by *future* events, and the worry is that we are possibly in a position to interfere to stop the future purported cause from occurring. All of our cases so far, however, have not been present events being caused by events that have not yet occurred; they concern past events being caused in the present.

Case 4: Re-Do Checkers

Suppose you are playing Max, a five-year-old, in checkers. Well, almost checkers. Max wants to play a game just like checkers, but he wants it to be the case that after your every move, he has the option of ordering a ‘re-do’. A re-do takes your move back and his latest move back, and then he is able to choose a different move instead. He also demands the right to a ‘super re-do’, or the ability to at any time rewind the game back to any state, including the very beginning position, and he wants to retain that right for the whole game. (This is only fair, he argues, because after applying a re-do the game is again in a position where you have just moved, and so he is again entitled to a re-do, leading to an arbitrary succession of re-dos. Separately, Max is an obnoxiously bright child.) You don’t really mind, so you fully agree to these rules. In any case, the game proceeds and, quite impressively, Max plays the game and wins without using a re-do. Good for him.

Here, every move depends upon how Max feels about it after its occurrence. Though Max does not ever use a re-do, any move counterfactually depends on his choosing not to use the re-do option at some point later. The whole game in fact depends upon his choice not to avail himself of a re-do after the last move.¹¹

I want to say that the game as it proceeds is at least partially backwardly caused. Each move depends for its occurrence in the game full-stop upon events later on. Even though Max never actually uses the re-do option, there is still backwards causation in this case (causation by omission). So, the past is never actually changed, though it is also not fixed.¹²

¹¹ Although this case is different such that present events *do* depend on future ones, notice that it’s still the case that nobody has the power to bilk. [So, this is compatible with Garrett (2017)’s argument that backwards causation is not compatible with bilking.] Neither of you has the power to intervene (within the game) to prevent a future choice of using or foregoing a re-do from partially causing the current move. In fact, no one in the world has that power. Even if a friend came in and turned over the table, effectively ending the game, the game would still not technically be finished. (This is clear from *Jumanji*.) If we stop playing, it is not that those past moves are settled and not caused by future events; instead, those past moves are never fully determined, because it is constitutive of the game that the moves depend on future choices.

¹² For a case in the actual world, consider long-form improvisational comedy. Actors frequently accidentally do something odd, and thereafter fabricate a story from their character’s past that explains the odd conduct. More mundanely, scenes frequently begin in the present by making assertions to demonstrate a past connection between actors. Once these things are said, they are taken at face value as a part of the past in the world of the fiction. These are present instances that affect the characters’ pasts, but it does not *change* that past, because no determinate past existed. Instead, with each new bit of information, they *nail down* or make more determinate the nature of their past, and other improvisors immediately accept it as truth (at least, when they are doing it well). That these cases involve precisifying the nature of the past instead of changing the past make them seem much closer to the cases discussed in Peijnenburg (*op. cit.*). Notice also that this is not our pretending that backwards causation exists. We are only pretending things about your character. The backwards causation is actual; it’s just a feature of the acceptable conduct for character generation in improv.

3.6 Still, These Cases Cannot Involve Changing the Past, Because the ‘Past Events’ Have No Other Effects

Consider the original event of the Tigers’ winning the championship. This event had consequences—celebrations were had, bragging was done. Take the specific event of the championship trophy being put on display at the home of the Tigers’. If we imagine a case of time travel where the past is changed, the rest of the past is typically *re-written*. If the Tigers were genuinely made to forfeit the championship *in the past*, there would not have been parties, nor bragging, and the trophy would not have been put on display. Yet, after the commissioner’s decision is announced, these events *still* occurred. This appears to undermine the claim that the commissioner’s announcement really does alter the past. The Tigers are stripped of their title, but they had been the champions until that point. What else could have caused the trophy’s being placed in their case?

To answer this challenge, consider a final case:

Case 5: The Barn Thief

Imagine a child (Barney), born in a barn in rural North Carolina, who goes on to invent time travel. Suppose that no one has been in that barn since Barney was born, though passersby often see the barn and remark on its rustic aesthetic. Assuming the barn will have become decrepit, and concerned for preserving it for posterity, Barney decides to travel back in time and steal the whole barn just after his birth. To avoid too many changes to the past, however, he builds a fake barn façade in its stead. Afterwards, passersby often see the façade and remark on its rustic aesthetic.

This case introduces a host of issues involving time travel we cannot resolve here. But there is a feature that will help to answer the objection at hand. In the case, the past has been changed—one object replaces another. For decades, there is a barn façade, not a barn. However, the events that the barn caused are now caused by the façade. The point is that while we would expect that a changed past would have different effects rippling out from the change, further changes are not *necessary* for the past to have been changed. (Perhaps all that is necessary is that further changes were possible.)

In our original cases, we can say that something similar has happened. The past has been changed, though many of the effects emanating out of that past remain the same. And why shouldn’t they? Many of the events of the championship still happened. People were moving in particular ways, scorecards were raised, and people cheered. This all gave the impression that the Tigers won the tournament. After all, it looked *just like* an event of tournament-winning.

My claim is that when the commissioner makes the announcement, it causes the team to have forfeited the tournament. There are other ways for a tournament forfeiture to look—the team could have forfeited by making an announcement and not competing. Or the conspiracy could have been uncovered during the tournament, ending their run. But the thought is that much of what happened is *also* how a forced forfeiture can present itself. The events involving the Tigers can constitute the event

of their winning, but it could also constitute the event of their forfeiting. However, it may be that the events at the time do not constitute a forfeiture until later.

Moreover, as we saw at the end of Sect. 3.4, it may still be that these changes themselves really do cause further changes in the past. If forfeiting the final match of a tournament compels a team to have to forfeit their first match in the next season as well, then the forfeiture causes another forfeiture, even if that forfeiture *also* at the time looked like a legitimate wrestling match. Or, to move away from forfeitures, perhaps being in PHIL756 caused Kevin to be qualified for financial aid for the last year, and there is thankfully still an open window for Kevin to fill out the paperwork to claim back that aid from last year. That these events can have effects after the cause of them has occurred is clear enough. (Now, Kevin can graduate.) There is no bar, however, to an event's having effects after it occurs but before it is caused. It just might be hard for us to recognize that these effects have occurred.

3.7 If These Would Be Genuine Cases of Backwards Causation, We Should Just Deny Institutional Facts

So far, we have been taking for granted institutional reality and the existence of institutional facts. Accepting this allows us to use the cases to make out the argument for the existence of backwards causation. But perhaps this is too big a pill to swallow. If accepting institutional reality commits us to backwards causation, we may feel more comfortable simply rejecting institutional reality.

The acceptance of institutional facts is not universal. Dörge and Holweger (2021) argue that these apparent facts are really illusions. Moreover, Wahlberg (2021: 5846) argues against institutional reality, and almost exactly for the reason that it could require the acceptance of some kind of backwards dependence, which he takes to be a result that social ontologists would generally reject.¹³ So it is possible that an argument for backwards causation via an acceptance of institutional reality is more likely to erode our confidence in the latter.

Finally, there may still be interesting things to say about the cases even without accepting institutional reality or backwards causation. Though it would not be literally true that Kevin was registered for PHIL756 in year X, it could be *fictionally* true, or true in the fiction. This could be affirmed by anyone inclined towards anti-realism

¹³ In particular, Wahlberg is concerned with Searle's understanding of the creation of institutional facts through declarations. So, for example, on Searle's view, if I declare that my student's paper has achieved the mark of 74, then it is thereby true that the student has a 74 on their paper at that very moment, even if the student is far from me as I declare it. However, since events can only be simultaneous with one another relative to a particular frame of reference, according to the special theory of relativity, Wahlberg shows that this set-up makes it possible for there to be a frame of reference such that (from the perspective of a high-speed observer in that reference frame) the student gains the property before I make the declaration. Wahlberg is not quite talking about backwards causation as we often discuss it, but he uses relativity to show how apparently simultaneous but distant events imply a similar possibility, which he takes as a basis to reject any theory with this implication. As I will go on to say, I do not take this to be a sufficient reason to jettison institutional facts.

about social institutions, social kinds, or artefacts.¹⁴ Moreover, we could still affirm that Kevin's being registered for PHIL756 in year X is *explained* by the administrator's phone call in year X + 1, as some have argued recently in this journal that metaphysical explanation can be non-factive (offering genuine explanations for items that are not true) (Bertrand, 2022) or that certain explanations need not be backed by real relations like causation (Taylor, 2022).

If a reader separately rejects institutional reality, I will not deny that they can offer a coherent story for these cases. And it may be that a distaste for backwards causation could drive one to reject institutional reality. However, I suspect that this will not be the position most thinking about social ontology will find themselves in.

As Dörge & Holweger (*op. cit.*) recognize, an acceptance of institutional facts is the standard view in social ontology. This has become taken for granted, and the debate has largely shifted to consider the extent to which social categories/properties/kinds are mind dependent (see, *inter alia*, Khalidi, 2015; Ásta, 2018; Burman, 2019), whether they should be thought of as emergent (e.g., Silver, 2021), and other questions about how to fit institutional reality into a naturalistic picture. Further, it is not obvious that the specter of backwards causation is likely to dissuade many social ontologists from accepting institutional reality. Even as he is arguing to the contrary, Walberg (*op. cit.*: fn.24) admits that Brian Epstein and Jonathan Schaffer, two prominent social ontologists, are likely to accept institutional reality with some kind of backwards dependence (Epstein, 2015; Schaffer, 2019). A careful examination of social ontology will see a history of scholars regularly willing to allow the social domain to reshape their views on other topics, such as the correct theory of reference (Thomasson, 2003b) or the possibility of co-located objects (as almost any constitution theorist accepts).

If nothing else, it is interesting to show that accepting institutional reality comes with backwards causation.¹⁵ It is worth establishing as a conditional conclusion. Still, given the wide acceptance of institutional reality within social ontology, and a willingness to be open-minded about the metaphysical upshots or theoretical revisions the social world suggests, it is plausible that scholars within social ontology are more likely to accept backwards causation than to reject institutional reality.

¹⁴ See Thomasson (2003a) for an explication of this kind of approach. Thomasson goes on to reject it, as would I.

¹⁵ One additional point to push back on here is the somewhat broad claim that merely accepting institutional reality has this result, as there are different outstanding views about the nature of institutional reality. It seems clear that the conditions for backwards causation have been met when thinking about prominent accounts like those given in Searle (*op. cit.*) or Epstein (*op. cit.*). But perhaps there could be other views that reify institutional facts yet somehow have the resources for denying backwards causation between institutional events. I leave this open as an idea, but I do not think that it undermines my claims. For one thing, it may be impossible to reify institutional events on any plausible theory without admitting cases of backwards causation. It would just take more space than I have here to show this. As a second point, even if there were a view reifying institutional facts without backwards causation, this is not to say that this view is more plausible than others. Since I think appreciating social ontology requires revisions to the understanding of various metaphysical notions, I do not take a view's ability to screen off backwards causation to necessarily furnish an argument for it.

4 Conclusion

Much of our lives occurs in institutions of our own design. And we take there to be many efficacious events that occur in the context of them. At the very least, we can be suspicious that the rules of causation—crafted to capture rocks thrown at windows—will apply without issue to causation within institutions. Why would they? There is already ongoing discussion of how to think about differences in causation at the lower level of quantum physics. Why not at the higher level of institutional reality?

I take the cases 1–4 to present themselves naturally as cases of backwards causation and to suggest that causation within institutions can be backwards. I suspect few readers agree. In fairness, I have not attempted to answer several of the objections that have been given directly to the possibility of backwards causation, nor have I moved forward the conversations on changing the past or time travel. But I have tried to answer some of the specific objections that might occur about the cases. If this will not convince readers that backwards causation is possible in this context, then perhaps it will at least provide fodder for future discussions of causation within institutions. However, I hope that the suggestion will not be brushed aside too quickly, because doing so risks missing just how valuable backwards causation can be.

The case of Re-Do Checkers drives home this value. The child wanted a game to even the playing field between us, and he was able to design one tailored to his preferences. The case is manufactured, sure. But that is not an indictment; it is a recommendation! We often *want* mechanisms for changing things in the past, and we have them. We have embedded such mechanisms into the fabric of our social reality. In this way, backwards causation is not something to be discovered at the fundamental level of physics; it is a technology to be carefully engineered and employed.

Funding Open Access funding provided by the IReL Consortium.

Declarations

Conflict of interest I have no relevant funding details or conflicts of interest to declare.

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

References

- Andreoletti, G., & Torrenco, G. (2019). Time travel and the immutability of the past within B-theoretical models. *Philosophia*, 47(4), 1011–1021.
- Anscombe, G. E. M. (1958). On brute facts. *Analysis*, 18(3), 69–72.
- Ásta. (2018). *Categories we live by*. Oxford University Press.
- Barlassina, L., & Del Prete, F. (2015). The puzzle of the changing past. *Analysis*, 75(1), 59–67.
- Baron, S. (2017). Back to the unchanging past. *Pacific Philosophical Quarterly*, 98(1), 129–147.
- Ben-Yami, H. (2007). The impossibility of backwards causation. *Philosophical Quarterly*, 57(228), 439–455.
- Ben-Yami, H. (2010). Backwards causation is still impossible. *Analysis*, 70(1), 89–92.
- Bertrand, M. (2022). We need non-factive metaphysical explanation. *Erkenntnis*, 87(3), 991–1011.
- Black, M. (1956). Why cannot an effect precede its cause. *Analysis*, 16, 49–58.
- Burman, Å. (2019). Categories we do not know we live by. *Journal of Social Ontology*, 5(2), 235–243.
- Corry, R. (2015). Retrocausal models for EPR. *Studies in the History and Philosophy of Modern Physics*, 49, 1–9.
- Cusbert, J. (2018). Backwards causation and the chancy past. *Mind*, 127(505), 1–33.
- Dörge, F. C., & Holweger, M. (2021). The non-existence of institutional facts. *Synthese*, 199, 4953–4974.
- Dowe, P. (1997). A defense of backwards in time causation models in quantum mechanics. *Synthese*, 112(2), 233–246.
- Dummett, M. (1954). Can an effect precede its cause. *Proceedings of the Aristotelian Society*, 28(Supplement), 27–44.
- Dummett, M. (1964). Bringing about the past. *Philosophical Review*, 73, 338–359.
- Edward, T. R. (2015). Did the past really change in 2012? *Organon f: Medzinárodný Časopis Pre Analytickú Filozofiu*, 22(3), 340–344.
- Epstein, B. (2015). *The ant trap: Rebuilding the foundations of the social sciences*. Oxford University Press.
- Faye, J. (2018). Backwards causation. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. <https://plato.stanford.edu/archives/sum2018/entries/causation-backwards/>
- Feynman, R. (1949). The theory of positrons. *Physical Review*, 76, 749–759.
- Flew, A. (1956–7). Causal disorder again. *Analysis*, 17, 81–86.
- Flew, A. (1954). Can an effect precede its cause? *Proceedings of the Aristotelian Society*, 28(Supplement), 45–62.
- Flew, A. (1956). Effects before their causes—Addenda and corrigenda. *Analysis*, 16, 104–110.
- Garrett, B. (2014). Black on backwards causation. *Thought: A Journal of Philosophy*, 3(3), 230–233.
- Garrett, B. (2015). Some remarks on backwards causation. *Revista Portuguesa De Filosofia*, 71(4), 695–704.
- Garrett, B. (2017). Agents, knowledge and backwards causation. *Analysis*, 77(1), 37–43.
- Geach, P. (1969). *God and the soul*. Routledge.
- Goddu, G. C. (2003). Time travel and changing the past: (Or how to kill yourself and live to tell the tale). *Ratio*, 16(1), 16–32.
- Goddu, G. C. (2011). Avoiding or changing the past? *Pacific Philosophical Quarterly*, 92(1), 11–17.
- Hawthorne, J., & Nolan, D. (2006). What would teleological causation be? In J. Hawthorne (Ed.), *Metaphysical essays* (pp. 265–284). Oxford University Press.
- Helm, P. (1975). Are “Cambridge” changes non-events? *Analysis*, 35(4), 140–144.
- Hindriks, F. A. (2020). Establishments as material rather than immaterial objects. *Australasian Journal of Philosophy*. <https://doi.org/10.1080/00048402.2020.1821724>
- Hume, D. (1748/2000). *An enquiry concerning human understanding*. In T. L. Beauchamp (Ed.). Clarendon Press.
- Iacona, A. (2016). On the puzzle of the changing past. *Philosophia*, 44(1), 137–142.
- Kant, I. (1783/2004). *Prolegomena to any future metaphysics*. In: G. Hatfield (Ed.). Cambridge University Press.
- Khalidi, M. A. (2015). Three kinds of social kinds. *Philosophy and Phenomenological Research*, 90(1), 96–112.
- Kim, J. (1976). Events as property exemplifications. In M. Brand & D. Walton (Eds.), *Action theory* (pp. 310–326). D. Reidel.

- Korman, D. Z. (2020). The metaphysics of establishments. *Australasian Journal of Philosophy*, 98(3), 434–448.
- Loss, R. (2015). How to change the past in one-dimensional time. *Pacific Philosophical Quarterly*, 96(1), 1–11.
- Mellor, D. H. (1998). *Real time* (2nd ed.). Cambridge University Press.
- Oddie, G. (1990). Backwards causation and the permanence of the past. *Synthese*, 85, 71–93.
- Peijnenburg, J. (2006). Shaping your own life. *Metaphilosophy*, 37(2), 240–251.
- Roache, R. (2009). Bilking the bilking argument. *Analysis*, 69, 605–611.
- Roache, R. (2015). What is it like to affect the past? *Topoi*, 34, 195–199.
- Schaffer, J. (2016). Grounding in the image of causation. *Philosophical Studies*, 173(1), 49–100.
- Schaffer, J. (2019). Anchoring as grounding: On Epstein's the ant trap. *Philosophy and Phenomenological Research*, 99, 749–767.
- Searle, J. (1995). *The construction of social reality*. The Free Press.
- Silver, K. (2021). Emergence within social systems. *Synthese*, 199(3–4), 7865–7887.
- Smith, N. J. J. (1997). Bananas enough for time travel? *British Journal for the Philosophy of Science*, 48(3), 363–389.
- Smith, N. J. J. (2015). Why time travelers (still) cannot change the past. *Revista Portuguesa de Filosofia*, 71(70th Anniversary Issue on Metaphysics), 677–694.
- Taylor, E. (2022). Backing without realism. *Erkenntnis*, 87(3), 1295–1315.
- Thomasson, A. L. (2003a). Foundation for a social ontology. *Protosociology*, 18, 269–290.
- Thomasson, A. L. (2003b). Realism and human kinds. *Philosophy and Phenomenological Research*, 67(3), 580–609.
- Tooley, M. (1997). *Time, tense, and causation*. Clarendon Press.
- Torrenço, G. (2018). Nunc pro tunc. The problem of retroactive enactments. *Philosophia*, 46(1), 241–250.
- Van Putten, C. (2006). Changing the past: Retrocausality and narrative construction. *Metaphilosophy*, 37(2), 254–258.
- Wahlberg, T. H. (2021). The creation of institutional reality, special theory of relativity, and mere Cambridge changes. *Synthese*, 198, 5835–5860.
- Wilson, A. (2018). Metaphysical causation. *Noûs*, 52(4), 723–751.

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