Time travel and counterfactual asymmetry

Alison Fernandes[1]

Abstract

We standardly evaluate counterfactuals and abilities in temporally asymmetric terms—by keeping the past fixed and holding the future open. Only future events depend counterfactually on what happens now. Past events do not. Conversely, past events are relevant to what abilities one has now in a way that future events are not. Lewis, Sider and others continue to evaluate counterfactuals and abilities in temporally asymmetric terms, even in cases of backwards time travel. I'll argue that we need more temporally neutral methods. The past shouldn't always be held fixed, because backwards time travel requires backwards counterfactual dependence. Future events should sometimes be held fixed, because they're in the causal history of the past, and agents have evidence of them independently of their decisions now. We need temporally neutral methods to maintain connections between causation, counterfactuals and evidence, and if counterfactuals are used to explain the temporal asymmetry of causation.

Keywords Time travel · Counterfactuals · Causation · Evidence · Temporal asymmetry · Backwards causation · Open future · David Lewis

1 Introduction

Consider a case of time travel. Tim travels back in time to when his grandfather was a young man, intent on killing his grandfather before he has a chance to grow up, meet grandma, and make the family fortune in munitions. Let's say Tim has an appropriate firearm, is well-trained, and successfully hunts down his grandfather. As Tim waits, finger poised on the trigger, *can* he kill him? What would happen *were* he to try?

Lewis (1976), Sider (2002), Horwich (1987), Smith (1997), Ismael (2003) and Carroll (2010) argue that Tim *can* kill his young grandfather. After all, young grandfather's survival lies in the future of Tim's actions. If we evaluate abilities and counterfactu-

✉ Alison Fernandes
asfernan@tcd.ie

1 Department of Philosophy, Trinity College Dublin, College Green, Dublin 2, Ireland

als in standard temporally asymmetric terms, what happens in the future shouldn't constrain what Tim can do in the present, or what would occur (provided we're not reasoning fatalistically).

In this paper I'll argue that we have strong reasons to adopt more temporally neutral methods of evaluating counterfactuals and abilities in cases of backwards time travel. To explain how, it will be useful to introduce some terminology. There are two ways in which we ordinarily evaluate counterfactuals and abilities temporally asymmetrically: by holding past events 'fixed', and keeping future events 'open'. A counterfactual method holds an event 'fixed' when it directly restricts the event from changing in counterfactual worlds. It keeps an event 'open', when it does not hold it fixed. By definition, I'll take holding the past fixed to imply holding *all* past events fixed, and keeping the future open to imply *not* holding *any* future events fixed. Similarly for abilities: a method that holds the past fixed takes past events to limits an agent's abilities now. A method that keeps the future 'open' implies that no future events can restrict an agent's abilities now.

Except perhaps for a short period leading up to the antecedent, we standardly hold the past fixed when evaluating counterfactuals, and keep the future open. For example, say, in the actual world, a cake arrives, you eat it, and then the crumbs are removed. If you weren't to eat the cake, future events would be different—there would be no crumbs—but past events would be the same—the cake would still arrive. The future is kept open, but the past is held fixed. Note that events may still be kept open, even if they remain unchanged in counterfactual worlds. For example, the weather the following afternoon may remain unchanged in virtue of past events remaining the same, but not because the method more directly restricts tomorrow's weather from changing. A similar temporal asymmetry seems to hold for abilities. A present ability to eat the cake may depend on past events, such as yesterday's dental operation, but not on future events, such as its not being eaten. The past constrains abilities and counterfactuals in a way the future does not.

I'll argue that we need to adopt more temporally neutral methods for evaluating counterfactuals and abilities in cases of backwards time travel. In particular, I'll argue that (i) there are strong reasons to hold *some* future events fixed. These hold whenever future events are in the causal history of the present, or agents have evidence of them independently of their present decisions. I'll also argue that (ii) there are strong reasons to keep the *past* open, and allow past events to change in counterfactual worlds. These hold whenever there is backwards causation.

Note that my argument is not that we can *never* reasonably evaluate counterfactuals in temporally asymmetric terms in cases of backwards time travel. I agree with Lewis (1973, p. 565) and Carroll (2010, p. 89) that how we evaluate counterfactuals and abilities is context dependent. To capture what we ordinarily mean in English by 'can', we may want to continue to evaluate counterfactuals and abilities in temporal asymmetric terms. But evaluations of abilities and counterfactuals matter for other projects as well, such as providing counterfactual accounts of causation and decision theory. For these projects, we need to adopt more temporally neutral methods.

Why care about counterfactuals in cases of backwards time travel? Firstly, getting clear on their evaluation clarifies where the correct response to the grandfather paradox lies. A number of influential papers slide from defending the possibility of time travel

to claiming that it has no interesting consequences for how we evaluate abilities and counterfactuals. But time travel may be possible, even if it has these interesting consequences. Secondly, time travel cases are an important testing ground for methods of evaluating counterfactuals that aim to apply to a full range of causal structures. Even though we might think of cases involving both backwards and forwards causation as unusual, they're in fact the more *general* case. The *special* cases are worlds (presumably like ours) where causation only goes in one direction. Thirdly, considering time travel cases helps us develop more temporally neutral methods of evaluating counterfactuals. Such methods are needed if counterfactuals are to be used to ultimately explain the temporal asymmetry of causation in physical terms. Temporally asymmetric methods cannot be used, because they presume asymmetries not traced back to physical asymmetries.

In the following, I'll assume a broadly 4-dimensionalist view of time, such that talk of the past, present and future is to be treated indexically. When I speak of past and future events, these are events earlier or later than the relevant antecedent. I'll assume that backwards time travel is logically and metaphysically possible and requires backwards causation.

One might already be worried that a metaphysical discussion of time travel such as this pays insufficient attention to physics, and would be better set within the framework of general relativity (Arntzenius 2006; Smeenk and Wüthrich 2011; Daniels 2014). Maybe issues concerning counterfactuals are resolved when time travel goes via closed timelike curves because such spacetimes have no general time ordering, and so travel would involve no local backwards causation (Smeenk and Wüthrich 2011). However, even if we define temporal direction locally, and use a temporally asymmetric local method of evaluating counterfactuals, we still need to evaluate counterfactuals that aren't local. As we'll see (Sect. 5), not even local *causal* structure settles more distant counterfactuals. Nor do facts about the shape of a spacetime region, or other global constraints (such as laws), straightforwardly settle modal facts concerning counterfactuals and abilities. Even if we hold the global constraints fixed (something that itself requires argument) this doesn't settle what else we should hold fixed.

The paper proceeds as follows. In Sect. 2, I show how responders to the grandfather paradox assume counterfactuals and abilities should be evaluated in temporally asymmetric terms (outside of fatalistic contexts)—even though doing so is *not* needed to respond to the paradox. In Sects. 3 and 4, I argue that there are strong evidential and causal reasons for evaluating counterfactuals in more temporally neutral terms. In Sect. 5, I diagnose why we have trouble evaluating counterfactuals in cases of backwards time travel. In Sects. 4 and 5, I also suggest how counterfactuals can be evaluated in more temporally neutral terms.

2 The grandfather paradox

2.1 An argument against time travel

Tim travels back in time and confronts the man who later grows up to be his grandfather. One might use this case to argue against the logical or metaphysical possibility of time travel as follows.

P1. If time travel is possible, Tim can kill his young grandfather.
P2. Time travel is possible. (assumed for *reductio*)
P3. Tim *can* kill his young grandfather. (from P1 and P2)
P4. If Tim were to kill his young grandfather, Tim both would and would not exist.
P5. Tim *can't* kill his young grandfather. (from P4)
Contradiction. (P3 and P5)

P1 seems true because Tim would travel back in time with all his usual causal abilities to kill whomever he confronts. P4 seems true because if Tim were to kill his grandfather, his grandfather wouldn't meet his grandmother, so Tim wouldn't come to be. P5 seems to follow from P4 because Tim can't do something such that, were he to do it, a contradiction would be true. Since P5 contradicts P3, P2 must be false. So Time travel is impossible.

Lewis (1976) defuses arguments of this form by noting that they equivocate on 'can'. Take 'young-gramps' to refer to the young man before Tim, who, in the actual world, is Tim's grandfather. There is a sense of 'can' in which Tim can kill young-gramps:

I. Tim *can* kill young-gramps (the young man before him).

Tim has the ability, given we hold fixed what we ordinarily do when evaluating abilities: skill, training, opportunity, and so forth.[1] There is also a sense in which Tim *can't* kill young-gramps, given by holding fixed the fact that young-gramps survives to father Tim's parent.

II. Tim *can't* kill young-gramps (Tim's grandfather).

You can evaluate Tim's abilities in either sense I or sense II. But 'what you mustn't do is waver, say in the same breath that he both can and can't, and then claim that this contradiction proves that time travel is impossible' (Lewis 1976, p. 151). Tim's case doesn't threaten the possibility of time travel.

The same response applies to counterfactuals. Assume, following Lewis (1981, p. 116), Sider (2002, p. 121) and Vihvelin (1996, p. 318) that abilities are related to counterfactuals about what would happen (or be caused to happen) were one to do, decide or try.[2] We can then associate the following counterfactuals with sense I and II respectively:

III. Were Tim to try to kill gramps, he would sometimes (or might) succeed.

---

[1] For possible scenarios, see Lewis (1976, p. 152), Smith (1997, p. 372ff.) and Sider (2002, pp. 131–133).

[2] My arguments won't depend on the details of how abilities and counterfactuals are related. One might think that counterfactuals and abilities are only related for a subclass of counterfactuals—those relevant to abilities (Sider p. 137 n. 9). Explicitly adopting this restriction won't change the arguments to come. Nor will it affect the point here: that context-sensitivity is what defuses the grandfather paradox.

IV. Were Tim to try to kill gramps, he would always (or certainly) fail.

We can evaluate counterfactuals in either sense III or IV. But what we mustn't do is evaluate counterfactuals in both senses in a single context and use the resulting contradiction to argue that time travel is impossible.

I agree with Lewis—so far. The context-sensitivity of 'can' defuses these arguments against the possibility of time travel. But defenders of time travel have gone further. They go on to give what I call 'overkill responses'. They defend the possibility of sense I (the sense in which Tim *can* kill young-gramps), by arguing *against* the possibility or reasonableness of sense II (the sense in which Tim *can't* kill young-gramps). These responses are overkill because neither sense needs to be rejected in order to defuse the argument. Simply diagnosing the equivocation is enough.[3]

2.2 Fatalism and foreknowledge

Who makes an overkill response? Well, Lewis, for one. He allows that what we hold fixed when evaluating abilities will vary. But he takes methods that hold parts of the future fixed to be unreasonable, and associates them with *fatalism*. A fact about what Tim does in the future 'is an irrelevant fact about the future masquerading as a relevant fact about the past, and so should be left out of account in saying what, in any ordinary sense, he can do' (1976, p. 151). Even though the causal ordering of events in time travel cases is different from that in the actual world, this only provides the fatalist with new 'methods of disguise' (ibid., p. 151). It doesn't change how we should evaluate abilities (outside of fatalistic contexts). Lewis similarly calls appeals to Tim's knowledge that he will fail 'fatalist trickery' (ibid., p. 151): 'In calling Tim's state at that moment *knowledge*, not just belief, facts about personally earlier but externally later moments were *smuggled* into consideration' (ibid., p. 152, my emphasis). Lewis implies that future facts are irrelevant to abilities, outside of fatalistic contexts.

Others agree. Horwich sees no relevant differences between how we evaluate abilities in time travel and ordinary cases: '[my] inability to go back in time and kill myself as an infant is just a special case of my inability to go back and kill anyone before their death' (1987, p. 119). He claims Tim's knowledge is similarly irrelevant to his abilities (ibid., p. 117). Sider also associates holding parts of the future fixed with fatalism. According to Sider, 'in normal contexts, contexts in which fatalism is false, facts about what occurs after the time in question are irrelevant' (2002, p. 136 n. 8)—in particular 'irrelevant to the interpretation of the counterfactual, at least under a similarity metric relevant to questions of ability and freedom' (ibid., p. 137 n. 9). Carroll similarly associates holding the future fixed in Tim's case with fatalistic reasoning (2010, p. 89).

Lewis (1973, p. 565) and Carroll (2010, p. 89) explicitly allow that what we hold fixed when evaluating counterfactuals and abilities will vary with context. But all the

---

[3] Vihvelin (1996) rejects the first reading for cases in which, she argues, the counterfactual antecedent can't be brought about except by failure of the attempt (without introducing large miracles). Rennick (2015) rejects the first reading for cases in which the time traveller intentionally tries to bring about what they believe will not happen. I'm concerned with a broader class of cases. Moreover, my main concern at this stage is to defend the reasonableness of the '*can't*' reading, not to reject the possibility of the '*can*'.

above authors argue that holding future events fixed in time travel cases is no more justified than doing so in ordinary cases. This is why the charge of fatalism makes sense.

There are other kinds of overkill responses. Smith (1997) considers an argument that time travel is unlikely because it will lead to apparently-miraculous coincidences—coincidences like Tim's gun jamming or a bird flying that prevent Tim succeeding. In response, Smith argues that we shouldn't evaluate counterfactuals by holding future facts fixed—this is to get things 'back to front' (1997, p. 373). If we don't hold future facts fixed (like young-gramps turning out to be Tim's grandfather), killing young-gramps would *not* imply Tim does and does not exist. So Tim's attempts wouldn't necessarily lead to unlikely coincidences. So time travel does not imply unlikely coincidences.

But, while we can evaluate counterfactuals by *not* holding future facts fixed, as Smith suggests, Smith needlessly rejects methods that *do* hold future events fixed (ibid., p. 373). Rejecting such methods is not needed to respond to the grandfather paradox (as we've seen). Nor is it needed to respond to concerns over apparently miraculous coincidences. If we evaluate counterfactuals by holding Tim's future failure fixed, these apparent "coincidences", are no longer coincidences at all. They are just what we would expect from Tim's attempts. They provide no reason to think time travel is unlikely.[4]

Here's a final overkill response, given by Sider (2002) and Ismael (2003). In the actual world, we might 'selectively attend' to the 'permanent bachelors'—men who never get married (Sider 2002, p. 125). But these men are just as able to marry as anyone else. Sider and Ismael argue that the same is true for time travellers. They are also just as able to kill their intended victims as anyone else. We've simply *attended* to a case where they fail, and have been encouraged to do so by the description of the case. Tim's case was described as one where he goes back in time and confronts the person who *is* his grandfather. We've snuck in Tim's failure in our description. But Tim's failure isn't inevitable. So we shouldn't appeal to the impossibility of time travel (or Tim's inability) to account for it. This response is overkill, for the same reason above: once we note the equivocation on 'can' nothing more is needed to respond to the grandfather paradox. Moreover, as I'll argue in the next section, there are reasons to hold parts of the future fixed that don't amount to selective attention.

3 What's So Special about the Future?

Lewis, Horwich, Sider, Carroll, Smith and Ismael argue that Tim retains his ordinary abilities when he travels back in time, and that we should continue to evaluate counterfactuals and abilities (in non-fatalistic contexts) by keeping the future open. Their arguments were:

a. holding the future fixed amounts to *fatalism*,
b. *foreknowledge* is irrelevant,

---

[4] The apparent "coincidences" can also be explained by the system's dynamics (Arntzenius and Maudlin 2013).

c. holding future events fixed gets things *back to front*,

d. holding future events fixed results from *selective attention*.

In this section, I'll dismiss the charge of fatalism, and then present causal and evidential reasons for holding parts of the future fixed. The reasons I present will be relevant under a variety of methods of evaluating counterfactuals and abilities. I'll then use these reasons to reject the remaining overkill arguments.

Before I go on, note that the overkill arguments above were given in terms of both counterfactuals and abilities. It will be useful to keep in mind their relation. Say abilities relate to counterfactuals as above (Sect. 2.1): Tim can ø just in case he would (or might) ø, were he to decide to ø, try to ø, or do something else. Then evaluating *abilities* by keeping the future open requires evaluating *counterfactuals* by keeping the future open. For *reductio*, say we keep some future events fixed when evaluating counterfactuals concerning Tim (including young-gramps' survival), but *not* when evaluating abilities. Then Tim *can* kill young-gramps, since his ability doesn't depend on future events. But, it is false that, were he to try to, he might succeed (since young-gramps' survival is held fixed when evaluating counterfactuals). If counterfactuals and abilities are related as above, Tim *can't* kill young-gramps. Contradiction. To avoid such contradictions, if we hold future facts fixed when evaluating counterfactuals, we must also do so when evaluating abilities (in the same context). *Contraposed*, if we keep the future *open* when evaluating abilities, we must do so when evaluating counterfactuals. So arguments directed *against* holding the future open when evaluating *counterfactuals* are also directed against doing so when evaluating *abilities*.

3.1 Who's afraid of fatalism?

Lewis, Sider and Carroll associate methods of evaluating counterfactuals and abilities that hold future events fixed with fatalism, and, with Horwich, Smith and Ismael, claim that there are no relevant differences between Tim's case and ordinary cases. Let me begin by dismissing the charge of fatalism.

The kind of fatalism Lewis and company presumably have in mind is logical fatalism. In a contemporary guise, the logical fatalist argues from the existence of facts at one time about what occurs to the unavoidable occurrence of all events (Rice 2015). She might argue that Lewis can't vote Republican because it was true hundreds of years ago that he wouldn't (Lewis 1976, p. 151). Logical fatalism, as just explicated, is a view about *what* to hold fixed (all events) and *why*: because there are facts or true claims at other times concerning them.

Insofar as Lewis and company associate *any* attempt to hold parts of the future fixed with fatalism, they're trading in misleading rhetoric.[5] Holding some future events fixed does not imply accepting logical fatalism. Firstly, logical fatalism holds *all* future events fixed. It is partly for this reason that fatalism seems so obviously mistaken: surely we can change *some* future events. Secondly, holding future events fixed in cases of backwards time travel does not imply doing so in ordinary cases in the actual world. One may have reasons for holding future events fixed in time travel cases that

---

[5] Vihvelin (2011) and Rea (2015, p. 270 n. 7) also distinguish between arguments for future events limiting our freedom and fatalist arguments.

do not apply in the actual world. I'll present some examples in Sect. 3. Thirdly, one can't presume, without argument, that any attempt to hold future facts fixed in time travel cases relies on fatalist arguments. So one should not associate *any* attempt to hold future facts fixed with fatalism.

Lewis and company could simply be (implicitly) defining fatalism as any view that holds any future event fixed. But, if they do, the charge of fatalism carries no argumentative force. Being a fatalist would not imply accepting standard fatalist arguments, or having an obviously unreasonable method of evaluating abilities and counterfactuals.

3.2 Causing the past

Having dismissed the charge of fatalism, let me consider a positive reason for holding parts of the future fixed when evaluating counterfactuals in cases of backwards time travel: parts of the future are in the *causal* history of the past. Time travel requires future events (particularly events concerning the time traveller) to *cause* past events. This is a standard requirement on backwards time travel accepted by Lewis and company. Assume also that counterfactual asymmetry in the actual world (a temporal asymmetry in what counterfactuals are true) is related to the temporal asymmetry of causation: either one directly explains the other, or both are explained by some third asymmetry. This assumption follows from a variety of popular accounts of causation and counterfactuals, including reductive counterfactual accounts of causation (Lewis 1979; Loewer 2007), reductive causal accounts of counterfactuals (Jackson 1977; Kvart 1986; Edgington 2004; Wasserman 2018, Ch. 5), non-reductive accounts of counterfactuals and causation (Schaffer 2004), including those from the causal modelling tradition (Pearl 2000; Halpern 2000; Hitchcock 2001; Woodward 2003; Hiddleston 2005), as well as accounts that appeal to a 'third factor' underlying both causal and counterfactual asymmetry (Maudlin 2007, Chs. 1, 5). While one could give up this assumption in order to block the following argument, to do so would go counter to much of the recent work on causation and counterfactuals.

Given the above assumptions, there are then strong reasons for holding parts of the future fixed when evaluating counterfactuals.[6] To see why, begin by entertaining a (perhaps implausible) view according to which the direction of time is entirely independent of the direction of causation. If so, there can be worlds with vastly more backwards causation than forwards. One such world is a world like time-travelling Tim's, but with the direction of time reversed: Tim travels forwards in time to meet his grandfather, and everyone else is a backwards time traveller. Lewis and company would be committed to keeping the future *open* when evaluating counterfactuals at

---

[6] Kiourti (2008, pp. 349–350) also argues that future events are relevant to an agent's abilities when they're in the personal past. The arguments below are more general, and do not rely on demarcating the personal past. While Rea (2015) eschews talk of counterfactuals and abilities, he also uses causal features to argue that time travellers are not free. However, Rea relies on the premise that we should keep the *whole* causal past of an agent fixed when evaluating what the agent is free to do at *t*, where the causal past includes all events that stand in the ancestral of the causal relation to events at *t* concerning the agent (ibid., p. 271). As Rea acknowledges (ibid., pp. 269–270), his justification for this premise relies on assuming freedom is *incompatible* with determinism—an assumption that I, as well as Lewis and company, would reject. I'll argue that *some* future events (in the causal past) should be held fixed, not that all events in the causal past (in Rea's sense) should be held fixed—see Sect. 5.

such worlds. Future events (which have the causal features of past events at our world) *won't* be held fixed, no matter how distant these events are from Tim's activities. Lewis and company are committed to this result because none of their arguments against holding the future fixed in Tim's case were sensitive to the amount of backwards causation involved. If the future is kept open, ordinary events can counterfactually depending on their effects, implying that 'backtracking' counterfactuals are true. For example, take the time-reverse of the cake example above: the cake arrives at t_2 (A), causally contributing to your eating it at t_1 (B), causing the crumbs to be left at t_0 (C). If the future is kept open when evaluating counterfactuals, whether the cake arrives (A) can counterfactually depend on whether you eat it (B). That is, taking $\square\rightarrow$ as the counterfactual conditional connective, it can be that B$\square\rightarrow$A and ¬B$\square\rightarrow$¬A.[7] This result is contrary to the forwards case, where backtracking is ruled out by holding past events fixed.

But this first argument presupposed an unlikely view of time and causation. A more plausible view is that the direction of time matches the dominant direction of causation at a world. This view, however, still allows there to be worlds with only slightly more forwards causation than backwards causation. Perhaps slightly more regions of the world contain forwards causation. In such worlds, there is no reason to always decisively privilege one temporal direction when evaluating counterfactuals. Yet Lewis and company would be committed to always keeping the future open when evaluating counterfactuals at such worlds, leading to the same problem as above: backtracking counterfactuals can be true at such worlds, when they concern events in the 'backwards' regions. For example, if the cake-eating above takes place in a backwards region, whether the cake arrives at t_2 (A) can counterfactually depend on whether you eat it at t_1 (B): B$\square\rightarrow$A and ¬B$\square\rightarrow$¬A. Keeping the future open when evaluating counterfactuals disconnects the direction of counterfactuals from the direction of causation.

Another view is that the direction of time in a local region corresponds to the direction of causation in that region. According to this view, worlds described above as containing both forwards and backwards causation may actually only contain forwards causation. In this case, keeping the future open will not allow for backtracking counterfactuals to be true in ordinary cases. But while such a view is plausible, it doesn't help Lewis and company. Tim's failure is forwards in time of his attempt in one region (the region containing grandfather's life), and backwards in time in another (the region containing Tim's time travel). So this response doesn't rule out holding Tim's failure fixed when evaluating counterfactuals. Without a general time ordering, moreover, analogies that Lewis and company appeal to between the actual world and time travel worlds break down. In the actual world, an event being in the future implies that it is not in the past. Not so if we deny global temporal order.[8]

---

[7] Whether ¬B$\square\rightarrow$¬A is true depends on details of how counterfactuals are evaluated. The claim I make here is comparative: to the extent that holding past events fixed is required to rule out backtracking counterfactuals in the actual world, as in Lewis (1979) and on causal methods of evaluating counterfactuals, holding future events fixed is required to rule out backtracking in the above case.

[8] Local approaches are also difficult to square with global methods of evaluating counterfactuals like Lewis' (1979).

Here's a second way in which causal features of Tim's case support holding parts of the future fixed. Assume that causal asymmetry is ultimately explained in terms of other physical asymmetries, via counterfactuals. Lewis (1979), for example, attempts to derive counterfactual and causal asymmetry from an asymmetry of 'traces'. Traces are local events that individually nomically entail the occurrence of another event.[9] According to Lewis, events leave multiple traces in their futures, but not their pasts. In other words, many local events that come after (but not before) an occurrence can individually nomically entail that occurrence. Because of this asymmetry, Lewis argues, many 'miracles' would be required for a counterfactual state to nomically entail the counterfactual antecedent, if the miracles were to occur *after* the antecedent. But, given there aren't traces of the future, only a small miracle just *prior* to the antecedent is required for the counterfactual state to nomically entail the antecedent. This asymmetry is then used to explain why most of the past is the same in counterfactual worlds. If an account is to explain causal asymmetry at our world, a plausible requirement is that it must be able to derive causal directionality in worlds with more complex causal structures. In worlds with backwards time travel, this will require showing that there are areas where the local *past* depends on a present antecedent in a way the local *future* does not.[10] Under many approaches, this will require holding parts of the local future fixed. In Lewis' case, for example, backwards time travel will require there to be areas in which events leave multiple traces in the (local) *past*, but not the (local) *future*, such that only a small miracle in the *future* is required to produce the antecedent—implying a *future* that is held mostly fixed.[11]

Similar remarks apply to more recent attempts to explain counterfactual asymmetry (Albert 2000; Loewer 2007). Whatever physical asymmetries ultimately explain counterfactual asymmetry at our world provide reasons for keeping parts of the future fixed in cases of backwards time travel. Albert and Loewer derive counterfactual asymmetry from statistical-mechanical asymmetries. If their accounts are to deliver violations of causal asymmetry in backwards time travel cases, there must be something like local regions where the usual statistical-mechanical asymmetries are reversed, and where the temporal past depends on the present in a way the temporal future does not. This will require holding parts of the future fixed.

I've presented two causal reasons for holding parts of the future fixed when evaluating counterfactuals in cases of backwards time travel: the relation between counterfactuals and causation, and the need to explain causal asymmetry in terms of

---

[9] Lewis defines traces non-causally. However, he was wrong about the precise form and origin of the trace asymmetry (Elga 2001; Field 2003), and debate has continued over whether asymmetries of traces or 'records' can be explained in non-causal terms. See Reichenbach (1956), Horwich (1987, Ch. 5), Albert (2000, Ch. 6) and Loewer (2007) for attempts, and Earmam (1974) and Frisch (2010) for criticism. Earman (1974, p. 41), for example, argues that traces must be analysed in causal terms.

[10] Even if cases of backwards time travel involve causal loops, accounts still need to give *direction* to the causal loop. The presence of causal loops does, however, create other trouble—see Sect. 5.

[11] Lewis' approach doesn't actually imply that future events are held fixed. But this is due to a general problem his account has with cases of backwards time travel. Lewis' method is global in character, seeking perfect match between largest possible spatiotemporal regions, and so has trouble capturing the local violations of causal order in cases of backwards time travel. For examples of this trouble, see Tooley (2002), Collins, Hall and Paul (2004, p. 11), and Wasserman (2015). For a related diagnosis and a causal solution, see Wasserman (2018, pp. 171–182).

other physical asymmetries. These reasons don't rely on fatalist arguments, and so are relevant to how we evaluate counterfactuals and abilities in non-fatalistic contexts.

3.3 Records of the future

I'll now give an evidence-based argument for holding parts of the future fixed in cases of backwards time travel.[12] This argument builds on observations above, but does not directly invoke causal considerations. In cases of backwards time travel, agents can have evidential access to the future *of a kind* they usually only have to the past. Ordinarily we have *records* of the past and not the future, where records are, roughly, local states that are highly reliable (but not necessarily veridical) indicators of other states, given limited relevant background information. For example, photographs are reliable indicators of how people *used to* look, and our memories are (at least reasonably) reliable indicators of what *has* befallen us. Even though, conceptually, we may have records of the future, we don't seem to in ordinary cases in the actual world. We typically require much more information to reason as reliably about the future as we do about the past. For example, finding equally reliable indicators of how we're *going to* look and what is *going to* happen to us is hard, and requires much more background information about surrounding events.

Exactly how records should be characterised is controversial. If one gives a causal account of records, the reasons given in this section for holding some future events fixed will largely overlap with those of the last.[13] But one might give a non-causal account of records (Albert 2000, Ch. 6; Loewer 2007). For example, records may be states that raise the objective probabilities of the states they record, compared to the record's absence, given limited relevant background information (Loewer 2007). Objective probabilities themselves may then derive from laws and a probability measure (applied to macrostates). Under these approaches, the reliability of records is underwritten by nomic and probabilistic rather than causal relations. These approaches aim to explain why we have records of the past (and not the future) in non-causal terms.[14] Regardless or whether one adopts a causal or non-causal approach, whatever causal or nomic mechanisms allow us to have records of the past in the actual world allow backwards time travellers like Tim to have records of the future that are as reliable as our usual records of the past. Tim can travel back in time with a vast store of reliable indicators about his grandfather's life and death: memories, photographs, testimonials,

---

[12] Ismael (2003, p. 314) and Arntzenius (2006, p. 613) also note that violations of evidential asymmetry in time travel cases may spell trouble for asymmetric counterfactual and causal reasoning. But neither takes this to affect the time traveller's abilities. Rennick (2015) argues that a time traveller's freedom may be limited by her beliefs. The approach I take below provides a more general route to a similar conclusion.

[13] Even so, one might hold future events fixed in virtue of the fact that they *cause* present events, or in virtue of the fact that they are *evidenced* by present events. As I discuss in the next section, these options suggests different approaches to evaluating counterfactuals.

[14] Accounts differ on whether the asymmetry of records is strict. For example, while Albert begins by drawing attention to particular *methods* that apparently work towards the past (and not the future) (2000, p. 113), the asymmetry he ultimately derives (ibid., p. 122) doesn't actually entail a strict asymmetry of local records. Loewer (2007, p. 303) defines records somewhat differently, and explicitly takes the asymmetry to be matter of degree—we only have *more* reliable records of the past (compared to the future). For criticisms of this program, see, again, Earman (1974) and Frisch (2010).

newspaper clippings. These records concern what will befall young-gramps later in life—events that now lie in the future.

Records are relevant to evaluating counterfactuals and abilities, particularly in cases involving deliberation. When we're deliberating about what to do, and evaluate counterfactuals concerning our decisions, we typically hold fixed what we believe to be the case independently of our decisions. Say you're thinking about where to go for dinner, and what would happen were you to decide to go to one restaurant or another. In evaluating these counterfactuals, you hold fixed what you believe about how busy each restaurant will be. In general, if we are good epistemic agents, and not too unlucky, our beliefs will conform reasonably well to our reliable *external* evidence—what the sources of evidence we have encountered reliably indicate. For example, if you have read several reputable restaurant reviews saying that a given restaurant is always extremely busy (and you have encountered no significant countervailing evidence), you will believe that the restaurant will be busy.

Typically we have more (and more reliable) external evidence about the past than the future. So much more of the past gets held fixed (if we are good epistemic agents). But time travellers like Tim can have external evidence of the future in the form of records that is as reliable as any of our usual evidence of the past. If Tim is considering what would happen were he to shoot young-gramps, has reliable evidence of young-gramps's survival (independently of his decision), lacks countervailing evidence, and is a good epistemic agent, this gives him reason to hold young-gramps' future survival fixed.[15] Similarly, say Tim is thinking about whether he *can* kill young-gramps. If he has evidence that, overall, settles his failure, independently of what he decides or does now, and is a good epistemic agent, then he shouldn't take himself to be able to kill young-gramps.[16]

If one is convinced by this line of thinking, then a tempting view is that it is Tim's having reliable evidence that settles young-gramps' survival that *directly* implies that young-gramps' survival should be held fixed when evaluating counterfactuals and abilities concerning Tim. Independently of whether Tim comes to form the appropriate *beliefs* about whether young-gramps' survives, he has the evidence, and so is not able to kill young-gramps.[17] There are more subjectivist and objectivist alternatives one could

---

[15] If Tim is unsure of the metaphysics of time, perhaps we'd not fault him for failing to believe his grandfather will survive—but he'd still be failing to follow his external evidence.

[16] Won't Tim have countervailing evidence in the form of evidence about what he can do? Certain of his ability to kill the man before him, shouldn't he be certain it isn't his grandfather after all? (My thanks to a reviewer for the case.) In particular scenarios, one's *alethic* modal evidence about what one can do (and what that would imply) may outweigh one's *epistemic* modal evidence about what does happen in the actual world. (See Ismael (2017, p. 117) for more on the distinction.) What I'm committed to is a priority of epistemic modality over alethic in new settings: in these settings, modal evidence takes priority over alethic evidence in cases of conflict. My reasons for this are themselves epistemic: in new settings, we learn about what would happen if we tried only by observing what does happen when we try. So Tim shouldn't remain certain of his ability to kill young-gramps and take this to throw in doubt his evidence of young-gramps' identity.

[17] There are difficult issues to explore concerning what it is for Tim to 'have' evidence. For example, does Tim have evidence when he has an envelope that (unbeknownst to him) contains relevant information, but has, a) left it behind, unopened, or b) brought it with him, unopened, or c) brought it with him and read it? (My thanks to a reviewer for the case.) My responses are: a) no, b) probably not, c) yes. The answers (to my mind) turn on what we take the relevant epistemic standards of a world to be (noting that these may

explore. For example, one could appeal to the evidence present in Tim's *surrounds*, or take counterfactuals to be relative Tim's beliefs. While my own preference is for a mixed approach, any of these approaches will potentially license holding future events fixed in Tim's case.

We can now see where the second overkill argument went wrong. Recall, Lewis (1976, pp. 151–152) and Horwich (1987, p. 117) claimed that Tim's knowledge of his future was irrelevant to his abilities now. Lewis argued that we can parse Tim's knowledge into one part about the present (his belief) that doesn't imply his failure, and another part about the future (the belief's justified truth) that is irrelevant to his abilities. But merely because the belief's justified truth concerns the future *doesn't* imply that it is irrelevant to Tim's abilities. To think so is analogous to equating holding the future fixed with fatalism. Nor is it the *truth* of the belief that explains why we should hold part of the future fixed. It is Tim's having reliable external evidence that settles his failure, evidence that also (externally) *justifies* his belief, that provides reason for holding parts of the future fixed when evaluating counterfactuals and abilities concerning Tim.

3.4 Against an open future

Let me summarise where the overkill arguments against holding the future fixed went wrong. Firstly, it's a mistake to associate holding the future fixed with fatalism. There are strong evidential and causal reasons for holding parts of the future fixed that don't rely on fatalist arguments. Secondly, *contra* Lewis and Horwich, Tim's knowledge of the future *is* relevant to his freedom now, insofar as his knowledge is based on reliable external evidence of the future. Thirdly, *contra* Smith, to hold future events fixed in cases of backwards time travel cases is not to get things 'back to front'. Holding future events fixed is permissible when there are violations to causal and evidential asymmetries. Fourthly, *contra* Sider and Ismael, there are important disanalogies between holding parts of the future fixed in ordinary cases, and doing so in cases of backwards time travel. We don't merely 'selectively attend' to cases where the time traveller fails. In Tim's case, his presence and his attempts are caused by his future failure, and there is external evidence now that settles his failure (independently of his attempts). In the case of permanent bachelors, their presence and their attempts are not caused by their future failures, and there isn't external evidence now that settles their future failures.

Let me again emphasise that holding parts of the future fixed doesn't imply that time travel is impossible. Lewis' solution to the grandfather paradox stands. Nor do apparently miraculous coincidences make time travel unlikely. If we evaluate counterfactuals by holding parts of the future fixed, these apparent "coincidences" are in fact expected events. It is only by combining both future-directed and past-directed expectations that we run into trouble: such as holding Tim's future failure fixed, and expecting his abilities to remain normal.

In the actual world, evidential and causal asymmetries align with temporal asymmetry. So no harm typically comes from evaluating counterfactuals and abilities in

---

Footnote 17 continued

be different from standards of praise or blame). I suspect someone isn't necessarily failing as an epistemic agent when they don't open letters, but they are if they fail to take into account information contained in letters they've read. But one could argue for different standards, or take a non-normative approach.

temporally asymmetric terms. But in backwards time travel cases, these asymmetries don't align. So we can't continue to evaluate counterfactuals temporally asymmetrically without violating plausible connections between counterfactuals, causation and evidence. Ultimately we may still want to evaluate some counterfactuals by keeping the future open, even in cases of backwards time travel. But we shouldn't ignore that there are strong reasons against doing so that aren't fatalistic.

I've given both evidential and causal reasons for holding parts of the future fixed. Each kind of reason suggests a different approach to evaluating counterfactuals: hold the causal past of (some part of) the present fixed, or hold events evidenced by (some part of) the present fixed. Are these approaches equivalent? Not straightforwardly. Each approach takes a different reason as basic, and may produce different results.[18] In order not to force a choice between these approaches, I've included both causal and evidential reasons.

For example, say we take causal reasons as more basic. Under a simple causal view of evidence, B counts as evidence of A only if A appropriately causes B. Keeping the causal past of the present (outside the antecedent) fixed entails keeping events evidenced by the present (outside the antecedent) fixed. But the converse does not hold: not all causal relations need be evidential. Alternatively, say we take evidential reasons as more basic. Some accounts attempt to derive causal relations from evidential relations, and so give non-causal accounts of evidence. Albert (2000, Ch. 6) and Loewer (2007), for example, appeal to statistical-mechanical probabilities. Their methods keep events recorded in the present outside the antecedent (for the most part) fixed. This entails keeping much of the causal past fixed, but there are exceptions—events in the causal past no longer recorded in the present.

My own view (2017) is closer to Albert's and Loewer's—causal relations derive from evidential relations—but takes the evidential relations to be those relevant to deliberating agents. This view fits well with the following method for evaluating counterfactuals: hold fixed what the relevant deliberating agent has reliable evidence of, independently of her decision (using a sufficiently externalist notion of evidence). In Tim's case, he has evidence of his grandfather's survival, independently of what decision he makes now. So in evaluating counterfactuals concerning Tim, young-gramps' survival is held fixed. A similar result holds, when evaluating Tim's abilities. Tim can actively explore the various ways in which young-gramps will survive his murderous attempts. But they are bound to fail. Tim is not able to kill young-gramps.

Much more would need to be said to lay out and defend such a view. What I offer here is merely a sketch of how a more temporally neutral method might go. For further details on how such a view might be developed and defended, see my (2016; 2017; Forth.). Under this proposal, interruptions to causal asymmetry in cases of backwards time travel are indicative of interruptions to evidential asymmetry. But the evidential reasons for holding parts of the future fixed remain the more basic.[19]

---

[18] Differences may also arise due to how fine-grained the relata are, or how probabilistic relations are treated.

[19] There may also be events evidenced in the present that aren't in the causal past, and events in the causal past that aren't evidenced in the present. So holding evidenced events fixed is, again, not equivalent to holding the casual past fixed.

4 Holding the past fixed

There are two ways in which we might evaluate counterfactuals and abilities temporally asymmetrically: by holding the (whole) past fixed *or* by keeping the future open—that is, not holding any future events fixed. So far I've argued that there are strong reasons to hold some future events fixed in cases of backwards time travel. Can we nevertheless evaluate counterfactuals asymmetrically by holding the past fixed?

We should not. Backwards time travel implies backwards causation. Assume that the temporal direction of causation is related to the temporal direction of counterfactuals, such that one explains the other, or they are both explained by some third asymmetry. Say we explain the asymmetry of causation in terms of counterfactuals, and take causal relations to require a chain of counterfactual dependencies between events (Lewis 1979). Then backwards causation *requires* backwards counterfactual dependence. That is, if A (at t_2) causes an earlier event D (at t_1), then at least one of the following must be true: i) D counterfactually depends on A (A$\square\rightarrow$D and ¬A$\square\rightarrow$¬D), ii) there exists an event *after* D on which D counterfactually depends ($\exists$B (at $t>t_1$) such that B$\square\rightarrow$D and ¬B$\square\rightarrow$¬D), or iii) there exists an event *before* A that counterfactually depends on A ($\exists$C (at $t<t_2$) such that A$\square\rightarrow$C and ¬A$\square\rightarrow$¬C). If any of these are true, there is counterfactual dependence of a past event on a future event. So the past is *not* held fixed. Say we explain the asymmetry of counterfactuals in terms of causation. Then backwards causation typically *implies* backwards counterfactual dependence. For example, Tim's entering his time machine in the future causes his appearance in the past. So Tim's appearing in the past counterfactually depends on his entering his time machine in the future. So the past is not held fixed. Under this alternative there could, strictly, be backwards causation without backwards counterfactual dependence. But this would produce rather odd results: Tim causes his appearance in the past, even though his past appearance in no way counterfactually depends on what he does now. Furthermore, under a difference-making account of explanation (Strevens 2008), if causal asymmetry is to explain counterfactual asymmetry, a difference in causal direction must standardly imply a difference in counterfactual direction (absent a principled reason for the implication to fail). So, in worlds with backwards causation, there will be some event D that counterfactually depends on a later event A (A$\square\rightarrow$D and ¬A$\square\rightarrow$¬D). So the past is not held fixed. Similar arguments hold if counterfactual and causal asymmetry are explained by some third asymmetry.

As far as I know, no one who accepts the possibility of backwards time travel argues that we should hold the past fixed when evaluating counterfactuals.[20] For this reason, the focus of my arguments has been on whether we can keep the future open.

5 Trapped by causal loops

Even though there are causal and evidential reasons for holding parts of the future fixed when evaluating counterfactuals and abilities in cases of backwards time travel,

---

[20] Horacek (2005, p. 424), however, argues that we should hold the past fixed when determining chances.

Lewis and company prefer to keep the future open. Let me offer a diagnosis, and explain why backwards time travel counterfactuals are often difficult to evaluate.[21]

Tim's case doesn't just involve backwards causation. It also involves a causal loop: a chain of events where each event is causally related to the next, and the last event is causally related to the first. Young-gramp's survival is causally related to Tim's time travelling, which is causally related to Tim's meeting him, which is causally related to young-gramps' survival. The existence of causal loops implies that some simple causal-based methods of evaluating counterfactuals won't work. Say we evaluate counterfactuals by holding events in the causal past of an area surrounding the antecedent fixed (excluding perhaps a brief transition period), and keeping events in the causal future open. In cases involving causal loops, the very same events are in the causal past *and* in the causal future. Young-gramps' survival is both in the causal past and the causal future of Tim's meeting him. Because the same events can't be held both fixed and open, the method can't be consistently applied.[22]

Tim's case looks especially bad because it leads us to expect a *self-defeating* causal loop: a causally connected chain of events that are compatible when laid out on a line, but which become incompatible when the first event is identified with the last. Tim seems able to kill young-gramps. But if he does, this will cause him not to be there in the first place. However, the expectation of self-defeat isn't the crucial feature that makes causal loop counterfactuals hard to evaluate. Trouble arises in cases where this expectation does not arise. Say old-Samuel Taylor Coleridge travels back in time and teaches *Kubla Khan* to young-Coleridge. Young-Coleridge's knowing *Kubla Khan* is a cause of his later knowing it. No unlikely coincidences are required for this loop to be consistent. But the same trouble arises when we try to evaluate counterfactuals. Would old-Coleridge know *Kubla Khan* were he *not* to teach it to his younger self? The consequent is in the casual past and causal future of an area surrounding the antecedent, so it is unclear whether we should hold it fixed or keep it open.

Lewis and company have preferred to keep events in the causal future open. This preference may be due to asymmetric tendencies in their thinking. It may also be due to wanting to keep Tim's abilities 'normal': as they would be in a world without backwards time travel. I suspect both these tendencies have played a role in generating overkill responses. However, a preference for keeping the future open might be more principled, and not based on temporally asymmetric assumptions. Lewis and company might think we should always assign events to the 'open' rather than the 'fixed' category when both principles can't be followed. An event's being causally open trumps any attempt to hold it fixed. Perhaps, in Tim's case, events in the casual future of his encounter with young-gramps are necessarily kept open, implying that Tim *can* kill young-gramps. This preference might also explain why holding fixed the future has been associated with fatalism: fatalists crucially neglect our causal control of the future when they claim we can't influence it. Lewis and company may also have been keen to avoid the unwelcome result that *no one* is free in a time travel world. Rea argues that we should keep the *whole* causal past of an agent fixed when evaluating

---

[21] Arntzenius and Maudlin (2013) discuss independent problems concerning indeterminacy.

[22] Rea (2015, pp. 273–274) also notes this concern with respect to what we are free to do. Rea prefers to keep the causal past fixed in a very wide sense. But his justification for doing so is an incompatibilist one.

what the agent is free to do at *t*, where the causal past includes *all* events that stand in the ancestral of the causal relation to events at *t* involving the agent 'as subject' (2015, p. 271). If we evaluate freedom in these terms, Rea argues, no one is free in a time travel world. If keeping the causal past fixed necessarily leads to these unwelcome results, this is a powerful reason for Lewis and company to prefer keeping the causal future open.

We may ultimately want to keep the causal future open when evaluating counterfactuals in cases of backwards time travel. But we shouldn't settle for this response too easily. Firstly, this approach does not fit well with the project of explaining the temporal asymmetry of causation using counterfactuals (Lewis 1979; Albert 2000; Loewer 2007). This project requires non-causal methods of evaluating counterfactuals. Secondly, a mere *preference* for holding the causal future open is not enough. We need arguments for why we should, particularly when doing so conflicts with holding the causal past fixed. These arguments should not rely on temporal asymmetric assumptions, or the assumption that there are no relevant differences between time travel cases and ordinary cases—as overkill responses do. Nor should the relevant contrast be holding the causal past fixed in Rea's sense, which requires holding fixed *all* events in the casual ancestor relation to *all* events now concerning the agent. One can hold fixed more limited parts of the causal past, or the causal past of a more limited part of the present. Overall, if what's driving the overkill responses is a principled preference for keeping the causal future open, the terms of the debate need to shift. We need arguments for why causal abilities downstream *always* trump causal conditions upstream. There remains a tension, moreover, between this approach and using counterfactuals to explain the temporal asymmetry of causation.

Let me end by returning to the temporally-neutral approach I suggested above: hold fixed what the relevant deliberating agent has evidence of, independently of her decision. Applied to Tim's case, because Tim has reliable external evidence of his grandfather's survival independently of his decision, young-gramps' survival is held fixed when evaluating counterfactuals and abilities concerning Tim. So Tim *can't* kill young-gramps. Applied to the Coleridge case, if old-Coleridge has reliable evidence of his teaching the poem to his younger self independently of his decision, his doing so should be held fixed when we consider what would happen were old-Coleridge to attempt not to.[23]

Note that under this approach, the counterfactuals used to determine causal structure may differ from those relevant to the actual agent. For example, we might derive causal structure by considering counterfactuals relevant to agents who can 'properly deliberate', and who lack appropriate evidence of what will occur (Fernandes 2016, 2017; Forth.). Tim can't properly deliberate on young-gramps' survival, because he has evidence that settles young-gramps' survival. But we can still consider what counterfactuals would obtain for a properly deliberating agent in a relevantly similar situation—one that mirrors the local but not global features of Tim's case. So we may

---

[23] What would happen were old-Coleridge *not to* teach the poem to his younger-self depends on details of the case. If Coleridge's evidence of knowing the poem is had independently of his evidence of how he came to learn it, his knowing the poem is held fixed—he will come to learn the poem by some other means. Otherwise, his knowing it is not held fixed.

still use counterfactuals to recover the causal structure of the case, such that Tim's actions are causally relevant to young-gramps' survival.

6 Conclusion

The grandfather paradox seemed to threaten the possibility of backwards time travel. In their eagerness to respond to the challenge, however, Lewis, Horwich, Sider, Carroll, Smith and Ismael go too far. Rather than simply showing that the argument against time travel equivocates on two different senses of 'can', they reject methods of evaluating counterfactual and abilities that hold parts of the future fixed. But this asymmetric preference is not justified in cases of backwards time travel. Backwards time travel requires backwards causation, and allows agents to have records of the past. While there may be principled reasons to keep the *causal* future open, this approach would need to be argued for, and fits poorly with the project of explaining the temporal asymmetry of causation using counterfactuals.

An alternative approach is to hold fixed events evidenced by some part of the present—such as those events a deliberating agent has evidence of independently of her decision. Methods such as this are temporally neutral, and allow parts of the future to be held fixed. So they can deal well with the interruptions to causal temporal asymmetry found in cases of backwards time travel. No doubt there are other alternatives to explore. But what we can't do is merely *presume* a fixed past or an open future. We need more nuanced considerations to guide how we evaluate counterfactuals in cases of time travel—particularly if counterfactuals are to explain the temporal asymmetry of causation.

References

Albert, D. Z. (2000). *Time and chance*. Cambridge, MA: Harvard University Press.

Arntzenius, F. (2006). Time travel: Double your fun. *Philosophy Compass, 1*(6), 599–616.

Arntzenius, F., & Maudlin, T. (2013). Time travel and modern physics. In Edward N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition). http://plato.stanford.edu/archives/win2013/entries/time-travel-phys/. Accessed October 24, 2016.

Carroll, J. W. (2010). Context, conditionals, fatalism, time travel, and freedom. In J. K. Campbell, M. O'Rourke, & H. S. Silverstein (Eds.), *Time and identity* (pp. 79–93). Cambridge, MA: MIT Press.

Collins, J., Hall, N., & Paul, L. A. (2004). Counterfactuals and causation: History, problems, and prospects. In J. Collins, N. Hall, & L. A. Paul (Eds.), *Causation and counterfactuals* (pp. 1–57). Cambridge, MA: MIT Press.

Daniels, P. R. (2014). Lewisian time travel in a relativistic setting. *Metaphysica, 15*(2), 329–345.

Earman, J. (1974). An attempt to add a little direction to "the problem of the direction of time". *Philosophy of Science, 41*(1), 15–47.

Edgington, D. (2004). Counterfactuals and the benefit of hindsight. In P. Dowe & P. Noordhof (Eds.), *Cause and chance: Causation in an indeterministic world* (pp. 12–27). Abingdon: Routledge.

Elga, A. (2001). Statistical mechanics and the asymmetry of counterfactual dependence. *Philosophy of Science, 68*(3), S313–S324.

Fernandes, A. Forth. Freedom, self-prediction, and the possibility of time travel. *Philosophical Studies*.

Fernandes, A. (2016). Varieties of epistemic freedom. *Australasian Journal of Philosophy, 94*(4), 736–751.

Fernandes, A. (2017). A deliberative approach to causation. *Philosophy and Phenomenological Research, 95*(3), 686–708.

Field, H. (2003). Causation in a physical world. In M. J. Loux & D. W. Zimmerman (Eds.), *The Oxford handbook of metaphysics* (pp. 435–460). Oxford: Oxford University Press.

Frisch, M. (2010). Does a low-entropy constraint prevent us from influencing the past? In G. Ernst & A. Hüttemann (Eds.), *Time, chance and reduction: philosophical aspects of statistical mechanics* (pp. 13–33). Cambridge: Cambridge University Press.

Halpern, J. (2000). Axiomatizing causal reasoning. *Journal of Artificial Intelligence Research, 12,* 317–337.

Hiddleston, E. (2005). A causal theory of counterfactuals. *Noûs, 39*(4), 632–657.

Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *Journal of Philosophy, 98*(6), 273–299.

Horacek, D. (2005). Time travel in indeterministic worlds. *The Monist, 88*(3), 423–436.

Horwich, P. (1987). *Asymmetries in time*. Cambridge, MA: MIT Press.

Ismael, J. (2003). Closed causal loops and the bilking argument. *Synthese, 136*(3), 305–320.

Ismael, J. (2017). An Empiricist's guide to objective modality. In M. Slater & Z. Yudell (Eds.), *Metaphysics and the philosophy of science: New essays* (pp. 109–125). New York: Oxford University Press.

Jackson, F. (1977). A causal theory of counterfactuals. *Australasian Journal of Philosophy, 55*(1), 3–21.

Kiourti, I. (2008). Killing baby Suzy. *Philosophical Studies, 139*(3), 343–352.

Kvart, I. (1986). *A theory of counterfactuals*. Indianapolis: Hackett.

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556–567.

Lewis, D. (1976). The paradoxes of time travel. *American Philosophical Quarterly, 13*(2), 145–152.

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

Lewis, D. (1981). Are we free to break the laws? *Theoria, 47*(3), 113–121.

Loewer, B. (2007). Counterfactuals and the second law. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality* (pp. 293–326). New York: Oxford University Press.

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

Rea, M. (2015). Time travelers are not free. *Journal of Philosophy, 112*(5), 266–279.

Reichenbach, Hans. (1956). *The direction of time*. In M. Reichenbach (Ed.). Berkeley: University of California Press.

Rennick, S. (2015). Things mere mortals can do, but philosophers can't. *Analysis, 75*(1), 22–26.

Rice, H. (2015). Fatalism. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition). http://plato.stanford.edu/archives/sum2015/entries/fatalism. Accessed 24 November, 2016.

Schaffer, J. (2004). Counterfactuals, causal independence and conceptual circularity. *Analysis, 64*(4), 299–308.

Sider, T. (2002). Time travel, coincidences and counterfactuals. *Philosophical Studies, 110*(2), 115–138.

Smeenk, C., & Wüthrich, C. (2011). Time travel and time machines. In C. Callender (Ed.), *The Oxford handbook of philosophy of time* (pp. 577–630). New York: Oxford University Press.

Smith, N. J. J. (1997). Bananas enough for time travel? *British Journal for the Philosophy of Science, 48*(3), 363–389.

Strevens, M. (2008). *Depth*. Cambridge, MA: Harvard University Press.

Tooley, M. (2002). Backward causation and the Stalnaker–Lewis approach to counterfactuals. *Analysis, 62*(3), 191–197.

Vihvelin, K. (1996). What time travelers cannot do. *Philosophical Studies, 81*(2/3), 315–330.

Vihvelin, K. 2011. Two objections to the possibility of time travel. http://vihvelin.typepad.com/vihvelincom/2011/05/two-objecti/. Accessed April 28, 2016.

Wasserman, R. (2015). Lewis on backward causation. *Thought, 4,* 141–150.

Wasserman, R. (2018). *Paradoxes of time travel*. Oxford: Oxford University Press.

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