Counterfactuals and the fixity of the past

Penelope Mackie

Abstract   I argue that David Lewis's attempt, in his 'Counterfactual Dependence and Time's Arrow', to explain the fixity of the past in terms of counterfactual independence is unsuccessful. I point out that there is an ambiguity in the claim that the past is counterfactually independent of the present (or, more generally, that the earlier is counterfactually independent of the later), corresponding to two distinct theses about the relation between time and counterfactuals, both officially endorsed by Lewis. I argue that Lewis's attempt is flawed for a variety of reasons, including the fact that his own theory about the evaluation of counterfactuals requires too many exceptions to the general rule that the past is counterfactually independent of the present. At the end of the paper, I consider a variant of Lewis's strategy that attempts to explain the fixity of the past in terms of causal, rather than counterfactual, independence. I conclude that, although this variant avoids some of the objections that afflict Lewis's account, it nevertheless seems to be incapable of giving a satisfactory explanation of the notion of the fixity of the past.

Keywords   Counterfactual dependence · Past · Fixity · Open future · Backtracking

1 Introduction: open future, fixed past, and the alleged asymmetry of counterfactual dependence

In 'Counterfactual Dependence and Time's Arrow' (CDTA), David Lewis writes, concerning what he calls '*the asymmetry of openness*':

> We tend to regard the future as a multitude of alternative possibilities, a 'garden of forking paths' in Borges' phrase, whereas we regard the past as a

P. Mackie (✉)
Department of Philosophy, University of Nottingham, University Park, Nottingham NG7 2RD, UK
e-mail: penelope.mackie@nottingham.ac.uk

unique, settled, immutable actuality. These descriptions scarcely wear their meaning on their sleeves, yet do seem to capture some genuine and important difference between past and future. What can it be? (CDTA: 36)

Lewis provides an answer:

I suggest that the mysterious asymmetry between open future and fixed past is nothing else than the asymmetry of counterfactual dependence. The forking paths into the future – the actual one and all the rest – are the many alternative futures that would come about under various counterfactual suppositions about the present. The one actual, fixed past is the one past that would remain actual under this same range of suppositions. (CDTA: 38)

Can the 'asymmetry of openness' be explained, as Lewis here proposes, in terms of a temporal asymmetry of counterfactual dependence?[1] In this paper, I argue that Lewis's attempt to explain the intuition of the fixity of the past in terms of counterfactual independence (or a failure of counterfactual dependence) is unsuccessful.[2] It follows that, regardless of whether the notion of the counterfactual dependence of the future on the past is relevant to the conception of the future as open, Lewis's ingenious attempt to explain the *asymmetry* of openness in terms of an asymmetry of counterfactual dependence does not work.[3]

A preliminary comment is in order about the explanatory task concerning the fixity of the past. Lewis often writes—and, for the sake of brevity, I shall follow him in this—as if the task were to explain the *fact* that the past is fixed, rather than our intuition that it is. However, it is evident (as the first of the quotations above indicates) that for Lewis the principal explanandum is not the fact that the past is fixed, but rather our intuition that it is so—what I shall call '*the fixity intuition*'. Some of my objections depend on the fact that it is the fixity intuition, rather than the existence of an objective correlate, that is the explanandum. The crucial issue is whether the way that we evaluate counterfactuals involves a temporal asymmetry that could explain our intuition that the past is fixed.[4]

I have five objections to Lewis's proposal to explain the fixity of the past in terms of a failure of counterfactual dependence: I summarize them here.

---

[1] Most of Lewis's paper is devoted to the task of constructing a (possible worlds) analysis of the truth conditions for counterfactuals that will yield the desired temporal asymmetry of counterfactual dependence, yet without building a temporal asymmetry into the analysis merely by *fiat*.

[2] For the distinction between counterfactual independence and failure of counterfactual dependence, see §6 below.

[3] Lewis's account of the openness of the future (and, by implication, his account of the fixity of the past) has recently been criticized by Barnes and Cameron (2011; §3). However, my criticisms are quite different from theirs.

[4] Similarly, the corresponding task concerning the future is not to explain why the future is open, but rather our intuition that it is. Again, when Lewis offers his famous account of the criteria for the evaluation of the closeness of possible worlds that he takes to give 'the correct truth conditions' for counterfactuals, the principal desideratum for the correctness of his account is that it yield truth conditions that correspond to our firm intuitions about the truth values of relevant counterfactuals, such as the 'Nixon' counterfactual (CDTA: 46–48).

1. Lewis is wrong in claiming that we normally keep the past (relative to the time of the antecedent) fixed when evaluating *backward* counterfactuals (§3 below).

2. Lewis's claim that we normally keep the more remote past (relative to the time of the antecedent) fixed when evaluating *forward* counterfactuals is at least doubtful (§4).

3. Lewis's claim that, when evaluating forward counterfactuals, we assume a change in at least the immediate past (relative to the time of the antecedent) to allow for a smooth 'transition period' undermines his attempt to explain the fixity of the past in terms of counterfactual independence (§5).

4. A retreat from 'counterfactual independence' to 'failure of (systematic) counterfactual dependence' does not help Lewis (§6).

5. The substitution of 'causal independence' for 'counterfactual independence' would not help Lewis (§7).

Before proceeding to the objections, there is an important distinction to be made between two different elaborations—represented by Thesis F and Thesis B of the next section—of the idea that the past is counterfactually independent of the present and future.

2 Forward counterfactuals, backward counterfactuals, and counterfactual independence

Lewis claims that the past is counterfactually independent of the present—and, more generally, that the earlier is counterfactually independent of the later. But this claim is, as it stands, potentially ambiguous. To explain why, I employ terminology taken from Bennett (1984: 57). I use 'forward counterfactual' to refer to one whose consequent is about a *later* time than any that its antecedent is about; and 'backward counterfactual' to refer to one whose consequent is about an *earlier* time than any that its antecedent is about. (Note that a 'backward counterfactual' in this sense need not say that the earlier would have been different had the later been different: thus the notion must be distinguished from the narrower notion of a backward counterfactual that is based on a 'back-tracking argument'—on which I shall have more to say shortly.[5]) I stipulate that if a counterfactual's consequent is about both an earlier and a later time than any that its antecedent is about, then it is neither a forward nor a backward counterfactual. Finally, I use 'the A-time' to refer to the time that the antecedent of a given counterfactual is about (thus ignoring, for simplicity, any counterfactual whose antecedent is about a multiplicity of times).

With this terminology established, we can see that to say that the past is counterfactually independent of the present—or, more generally, that the earlier is

---

[5] Bennett (2003) tends to use 'backward subjunctive conditionals' (which he prefers to 'backward counterfactuals') to refer to those that say that the earlier would have been different had the later been different. However, he acknowledges that the class of backward subjunctive conditionals does, strictly speaking, include those that say that the earlier would have been the same had the later been different (2003: 275).

counterfactually independent of the later—could imply a commitment to either or both of the following theses:

> Thesis F: When we evaluate *forward* counterfactuals, we keep the past, relative to the A-time, fixed.

> Thesis B: When we evaluate *backward* counterfactuals, we keep the past, relative to the A-time, fixed.

By 'we keep the past, relative to the A-time, fixed' when evaluating a counterfactual, I mean that we assume that, in the counterfactual situation, the past relative to the A-time is exactly the same as the actual past relative to the A-time. In terms of possible worlds: 'the closest possible A-worlds—i.e., worlds in which the antecedent is fulfilled—are worlds that share their past, relative to the A-time, with the actual world'.

It is important to consider Thesis F and Thesis B separately, given the possibility that different standards apply to the evaluation of forward and backward counterfactuals, standards that involve different treatments of the past relative to the A-time. In particular, we need to take seriously the proposal that Thesis F is true but Thesis B false.[6]

Suppose that Jane is a student who is both a perfectionist and bad at meeting deadlines. The deadline for essay submission—2 p.m. on Monday—approaches, but half an hour before the deadline Jane's essay is still only a rough draft. The proponent of the view that Thesis F is true and Thesis B false might hold that when we evaluate, with respect to this scenario, the forward counterfactual

> (C1) If Jane had handed in her essay at 2 p.m. on Monday, it would have got a low mark,

we keep the past relative to the A-time fixed, including the (actual) fact that by Monday afternoon she had only a rough draft of the essay (so that (C1), we may suppose, comes out true), whereas when we evaluate the backward counterfactual.

> (C2) If Jane had handed in her essay at 2 p.m. on Monday, she would have revised it properly first,

we do *not* keep the past relative to the A-time fixed, so (C2) may also come out true. What might explain this discrepancy in evaluation? The obvious thought is that when we consider the backward counterfactual (C2), our focus is on how the antecedent *might* or *would have come about*, whereas this is not our concern when considering the forward counterfactual (C1) (cf. Jackson 1977: 9, 11).

Obviously there is more to be said, since anyone who maintains that the standards for the evaluation of (C1) and (C2) differ in this way must confront the fact that (C2) appears to license the 'mixed' counterfactual[7]

---

[6] This was maintained by Frank Jackson in his (1977). For discussion of whether Thesis F, in the absence of Thesis B, could be regarded as yielding 'counterfactual independence' of the past with respect to the present, see §4 below.

[7] 'Mixed', because its consequent concerns both a time earlier than the A-time and a time later than the A-time; thus (C3) is neither a forward nor a backward counterfactual, as these have been defined.

(C3) If Jane had handed in her essay at 2 p.m. on Monday, she would have revised it properly first, and it would *not* have got a low mark,

and hence the *forward* counterfactual

(C4) If Jane had handed in her essay at 2 p.m. on Monday, it would *not* have got a low mark,

which is in opposition to (C1) and to the thesis (Thesis F) that the past is kept fixed when evaluating *forward* counterfactuals.[8] I shall not discuss this issue here, save to remark that the tension created by endorsing Thesis F while denying Thesis B might be relieved by putting a contextual restriction on the application of Thesis F.

Given the distinction between Thesis F and Thesis B, which does Lewis endorse in holding that the past is counterfactually independent of the present? The answer provided by the text of CDTA is: *both*. More precisely, Lewis's answer is 'both, subject to three qualifications'. These qualifications are: (1) a restriction to contexts that Lewis describes as involving 'the standard resolution' of the vagueness of counterfactuals; (2) a restriction to 'the sorts of familiar cases that arise in everyday life'—for example, we are to ignore bizarre cases involving time machines, black holes, or weird possible worlds consisting of just one atom in the void (CDTA: 35); and (3) a class of exceptions (about which I shall say more later) that is generated by the need to provide a smooth 'transition period' between the actual past relative to the A-time and the fulfilment of the antecedent of the counterfactual.

A forward counterfactual such as (C1) does not, of course, *explicitly* say anything about the past relative to the A-time. However, Lewis's account of the semantics of counterfactuals (under the 'standard resolution' of their vagueness) has consequences for how the past relative to the A-time is envisaged as being in the hypothetical scenario in which the antecedent is fulfilled (cf. Lewis CDTA: 46–48). According to Lewis (assuming that the antecedent of (C1) is in fact false—let's suppose that in fact Jane handed in her essay on Wednesday, rather than Monday) the closest possible A-worlds relevant to the truth conditions of the forward counterfactual (C1) are worlds that share their past with the actual world at least up until very shortly before the A-time (2 p.m. on Monday in this case), and then diverge.[9] Since, in these late-diverging possible worlds, Jane *still* hasn't finished her essay by 2 p.m. on Monday, the essay she hands in is a mess. And since, according to Lewis, the counterfactual (C1) is true if and only if, in the closest possible worlds where Jane hands in the essay at 2 p.m. on Monday, the essay gets a low mark, the forward counterfactual comes out true.

As for (C2), Lewis is quite explicit in stating that, *under what he calls 'the standard resolution of the vagueness of counterfactuals'*, the backward

---

[8] cf. Downing (1959). The moral that Downing draws from the apparent conflict is that a statement such as (C2) that appears to be a 'back-tracking' backward counterfactual conditional (or subjunctive conditional, to use Downing's preferred terminology) is in fact a statement of a different kind altogether, what Downing calls a 'subjunctive implication' (1959: 131–132). Downing does not appear to have considered the possibility of salvaging the consistency of pairs such as (C1) and (C4) via an appeal to a shift of context.

[9] Here and throughout this paper, I assume the truth of determinism, unless otherwise specified.

counterfactual (C2) is *false*. According to Lewis, if we confine our attention to 'the sorts of familiar cases that arise in everyday life' (ignoring non-standard cases involving time machines, black holes, possible worlds consisting of one solitary atom in the void, etc.—none of which is relevant to the current example), and, on the understanding that we are talking only about 'the standard resolution of the vagueness of counterfactuals', then his asymmetry claim is as follows:

> Consider those counterfactuals of the form 'If it were that A, then it would be that C' in which the supposition A is indeed false, and in which A and C are entirely about the states of affairs at two times t_A and t_C respectively. Many such counterfactuals are true in which C also is false, and in which t_C is later than t_A. These are counterfactuals that say how the way things are later depends on the way things were earlier. **But if t_C is earlier than t_A, then such counterfactuals are true if and only if C is true**. These are the counterfactuals that tell us how the way things are earlier does not depend on the way things will be later. (CDTA: 35; bold emphasis mine)

The claim I have emphasized in bold type is an explicit endorsement of Thesis B: i.e., the thesis that backward counterfactuals are assessed in a way that keeps the past relative to the A-time fixed.[10] And the commitment to Thesis B is evident elsewhere in Lewis's paper: for example, when he says:

> The past **would be the same, however we acted now**. The past does not at all depend on what we do now. It **is counterfactually independent** of the present. (CDTA: 38; bold emphasis mine)

Lewis is explicit in claiming that what he describes as 'the standard resolution' of the vagueness of counterfactuals does not allow for 'back-tracking' arguments, whether the counterfactuals are forward counterfactuals or backward counterfactuals, where back-tracking arguments appeal to the idea that since present conditions have their past causes, 'if the present were different then these past causes would have to be different, else they would have caused the present to be as it actually is' (CDTA: 33). Lewis does explicitly note that there are occasions when a back-tracking resolution of the vagueness of counterfactuals is appropriate, and that this is an obstacle to presenting 'a neat contrast between counterfactual dependence in one direction of time and counterfactual independence in the other direction' (CDTA: 33):

> We know that present conditions have their past causes. We can persuade ourselves, and sometimes do, that if the present were different then these past causes would have to be different, else they would have caused the present to be as it actually is. Given such an argument – call it a back-tracking argument

---

[10] This is worth emphasizing, since it has been said that Lewis holds that 'when we conditionalize from times to earlier times,… we adopt standards which are not those we use for forward conditionals' (Bennett 1984: 57). As I hope my quotations from Lewis have made clear, Bennett's statement is misleading if it is intended to suggest that Lewis's attitude to Thesis B is different from his attitude to Thesis F.

– we willingly grant that if the present were different, the past would be different too. (CDTA: 33)

However, he claims that these contexts involve a special, non-standard resolution of the vagueness of counterfactuals:

Under [the] standard resolution [of the vagueness of counterfactuals], back-tracking arguments are mistaken: **if the present were different, the past would be the same**, but the same past causes would fail somehow to cause the same present effects. (CDTA: 34; bold emphasis mine)

3 Objection 1: Lewis, back-tracking, and the 'standard resolution of vagueness'

Why am I placing so much emphasis on Lewis's commitment to Thesis B, as a thesis about the evaluation of backward counterfactuals under the 'standard resolution' of their vagueness? The reason is that Thesis B seems to me plainly false as a description of the way in which we normally evaluate backward counterfactuals. Confronted with the 'rival' backward counterfactuals (C2) and (C2*):

(C2) If Jane had handed in her essay at 2 p.m. on Monday, she would have revised it properly first;

(C2*) If Jane had handed in her essay at 2 p.m. on Monday, it would have been only a rough draft just before she submitted it,

there seems no reason to say that the 'typical' or 'ordinary' verdict favours (C2*) rather than (C2). If anything, I think it is the other way round.[11] Nor is (C2) an isolated case. It is easy to produce examples concerning human behaviour where, in evaluating a backward counterfactual, we naturally suppose that if the agent had acted differently at a time t, then some features of the past relative to t (that are independent of the agent's character) would have been different, otherwise the agent would not have acted thus.[12] Downing's (1959) example involving prideful Jim, resentful Jack, and the quarrel, cited by Lewis (CDTA: 33) may be regarded (contrary to Downing's own verdict) as involving a backward counterfactual (or

---

[11] Although I think this is true, it is a stronger claim than I need. All I need is that the back-tracking reading be at least as natural as the non-backtracking one. Lewis might be thought to accommodate this, given that he says that a counterfactual that requires a back-tracking reading for its truth will 'taken out of context,… not be clearly true or clearly false' (CDTA: 34). However, although this particular statement is consistent with equal naturalness, it does not entail it. And the text of CDTA shows that Lewis maintains that the non-backtracking reading is in some sense the 'default' reading.

[12] Of course, this means that we are keeping something fixed about the past relative to the A-time, namely the agent's character. But it would be a mistake to think that the so-called 'back-tracking' reading keeps one feature of the past fixed (the agent's character), whereas the so-called non-backtracking reading keeps something else fixed (the previous course of events). In Lewis's version of Downing's Jim and Jack case, the non-backtracking reading does *not* change Jim's character prior to the A-time; instead, it supposes that 'somehow Jim would have overcome his pride and asked despite yesterday's quarrel' (CDTA: 34).

subjunctive) conditional of this 'character-accommodating' type.[13] Nor is the phenomenon of the naturalness of the back-tracking reading confined to backward counterfactuals that concern agency. Take the counterfactuals

> (C5) If there had been ice on the pond this morning, the temperature last night would have been lower than it actually was;

> (C6) If the roof had been intact today, it would not have been hit by a falling tree yesterday.

According to Lewis's version of Thesis B, the back-tracking evaluation that is required to make the counterfactuals (C5) and (C6) come out true should require a context that is abnormal or atypical or non-standard. Yet, I submit, this is simply not the case.[14]

Yet if Thesis B is to be rejected, this represents a serious problem for Lewis. We may agree with Lewis that counterfactuals are infected with vagueness, and that different ways of resolving the vagueness are appropriate in different contexts (cf. CDTA: 34). We may thus agree that *one* way of resolving the vagueness of (C2) is to keep the past relative to the A-time fixed, with the result that (C2) comes out false. (On this resolution, if Jane had handed the essay in at 2 p.m. on Monday, it is not the case that she would have taken the precaution of revising it properly first; rather, she would have handed it in despite its unfinished state). However,—and here is my problem—with what right does Lewis call this (*non-backtracking*) resolution 'the standard resolution'?

If 'standard' means, as Lewis appears to intend it to mean, 'typical' or 'ordinary' or 'normal', then, for the reasons that I have indicated, Lewis's claim seems to be false.[15] Nor is Lewis well placed to assert that the 'standard' (in the sense of ordinary or typical) resolution of the vagueness of a backward counterfactual is one that outlaws back-tracking, given his view that the very assertion of a backward counterfactual such as (C2)—a counterfactual that requires support from a back-tracking argument for its defence—may create a context that is hospitable to its

---

[13] The relevant backward conditional is [(Jb)] 'If Jim were to ask Jack for help today, there would have been no quarrel yesterday', which, I claim, is naturally taken to be true, given the background assumption that Jim's pride would be an almost insuperable obstacle to his asking for help after a quarrel. Note that I make this claim only for the backward conditional. I do *not* automatically extend the claim to a forward conditional that might be supposed to be derived from (Jb), such as [(Jf)]'If Jim were to ask Jack for help today, Jack *would* help him'. Given my separation of Thesis B and Thesis F, I am willing to concede that the most natural evaluation of (Jf) is one that keeps fixed the past quarrel and Jack's consequent resentment, and thus supports the verdict that (Jf) is false.

[14] Perhaps Lewis might claim that the plausibility of (C5) and (C6) depends on the fact that their consequents involve changes to the past before the A-time that are required to avoid an 'abrupt discontinuity' between the past before the A-time and the fulfilment of the antecedent (cf. CDTA: 40). However, this response would simply push the wrinkle in the carpet to another place. See my 'Objection 3' in §5 below.

[15] Lewis suggests no other interpretation. And as well as describing back-tracking contexts as 'special', he uses the word 'ordinarily' when talking about the allegedly 'standard resolution' of backward counterfactuals (CTDA: 34).

truth (CDTA: 34).[16] I conclude that Lewis has taken a feature (the 'keeping fixed' of the past relative to the A-time) that may with some plausibility be regarded as a feature of the standard resolution of the vagueness of forward counterfactuals, and has extended it, quite implausibly (and with dubious consistency), to the standard resolution of the vagueness of *backward* counterfactuals.

Even if I am right about this, though, does it really matter? Perhaps Lewis should not have described the resolution of the vagueness of backward counterfactuals that rules out back-tracking arguments as 'the standard resolution'. But so what? As long as there is *some* clearly identifiable resolution of the vagueness of backward counterfactuals (call it 'the Lewis resolution') that rules out back-tracking arguments and keeps the past fixed, isn't that enough for his purposes, as long as this 'Lewis resolution' is one that we do employ at least some of the time in evaluating backward counterfactuals (which is not in dispute)? My answer is that is not enough, given Lewis's ambition to explain the intuition of the fixity of the past. There is a dilemma here. Either Lewis's case for the fixity intuition rests partly on a version of Thesis B (concerning our treatment of the past when evaluating backward counterfactuals), or it does not. If it does not, then the appeal to the 'Lewis resolution' version of Thesis B is, of course, completely irrelevant to the explanatory project. But if it does, then the appeal to the 'Lewis resolution' version of Thesis B must be insufficient. If all that Lewis can maintain is that, when we evaluate backward counterfactuals, we sometimes keep the past fixed, although we sometimes do not, this is not a version of Thesis B that could help to explain the intuition of the fixity of the past.

At this point, it might be objected that the resolution that Lewis describes as 'the standard resolution'—the one that keeps the past fixed and outlaws back-tracking— is not just any old resolution that we happen to use some of the time in our counterfactual thinking, but is the resolution of the vagueness of counterfactuals that gives the result that the direction of counterfactual dependence is the (standard) direction of causal dependence. Even though a backward counterfactual that is asserted on the basis of a back-tracking argument [such as my (C2)] may be (as I think) perfectly respectable as an illustration of the *counterfactual* dependence of the earlier on the later, no one wants to say that it represents a case of the *causal* dependence of the earlier on the later. On the contrary, it is of the very nature of a back-tracking argument that it infers the counterfactual dependence of the earlier on the later from the causal dependence of the later on the earlier.

---

[16] Lewis says that a 'counterfactual saying that the past would be different if the present were somehow different… [that comes out] true under the special resolution of its vagueness, but false under the standard resolution' may be called 'a *back-tracking counterfactual*' (CDTA: 34). However, this should not, I think, be taken as a definition, for reasons independent of (what I regard as) Lewis's tendentious use of the expression 'standard resolution'. Independently of my quarrel with Lewis on that issue, Lewis's characterization seems too narrow to serve as a definition of 'back-tracking counterfactual', since it does not apply to cases like my forward counterfactual (C4), which requires support from a back-tracking argument even though it does not explicitly *say* that the earlier would have been different if the later had been different. (Bennett, who introduced the term 'back-tracking', used it for the phenomenon of 'counterfactualizing back in time and then forward again' (Bennett 2003: 208), as exemplified by my 'mixed' counterfactual (C3). However, subsequent usage has not followed Bennett in this respect).

I shall return (in §7) to the question of the relation between the allegedly 'standard' (Lewisian) resolution of the vagueness of counterfactuals and the temporal asymmetry of causal dependence. For the present, I set it aside, and proceed to my second objection to Lewis, which concerns Thesis F.

4 Objection 2: Forward counterfactuals and keeping the past fixed

Thesis F appears considerably more plausible, as a claim about our ordinary (standard, normal, typical) use of counterfactuals than does Thesis B. Or, at any rate, it is plausible when modified to allow for the possible exception of a 'transition period' leading from the actual past to the fulfilment of the antecedent. To accommodate this last point, let us consider, as an alternative to Thesis F, the following:

> Thesis F*: When we evaluate *forward* counterfactuals, we keep the past, relative to the A-time, fixed, with the possible exception of a transition period leading from the actual past to the fulfilment of the antecedent.

Even if Thesis B is false, it might nevertheless be suggested that the truth of Thesis F or Thesis F* would be sufficient to secure, for the past, a counterfactual independence with respect to the present and future that could explain our intuition of the fixity of the past.[17] It appears to be undeniable that forward counterfactuals play a much more prominent role in our counterfactual thinking than do backward counterfactuals. Hence, if Thesis F (or Thesis F*) is true, then, even if Thesis B is false, it would follow that 'keeping the past fixed' is a feature of the evaluation of most of the counterfactuals that we are actually inclined to assert or consider, simply because of the fact that most such counterfactuals are forward counterfactuals rather than backward counterfactuals.[18]

Now, it might be objected that, strictly speaking, to say that the past is *counterfactually independent* of the present requires the explicit endorsement of backward counterfactuals that say that the past would have been the same if the present had been different, and thus that the mere endorsement of Thesis F/F*, without the endorsement of Thesis B, is not sufficient. However, even if this is, strictly speaking, true (of what it would take to establish that the past is counterfactually independent of the present), the more interesting issue is whether there is a temporal asymmetry in our counterfactual thinking that could explain the fixity intuition, even if the asymmetry does not involve the explicit assertion of any backward counterfactuals that say that the past would have been the same had the present been different. So I shall not rely on this objection (to the employment of Thesis F/F* in isolation from Thesis B).

---

[17] Here I am driven to use the phrase 'counterfactual independence with respect to' as an alternative to 'counterfactual independence of' because of the awkwardness of speaking of 'the past's counterfactual independence of the present and future'.

[18] Thanks to Philip Percival for suggesting this point.

What I shall maintain, however—and this is my second objection to Lewis—is that, although Thesis F—or at least its modification Thesis F*—has some plausibility, it is not *obviously* correct. How are we to tell whether Thesis F (or Thesis F*) is true of our normal practice, especially if (as the arguments of the last section imply) we cannot appeal, in testing Thesis F (or F*) against our practice, to the *backward* counterfactuals that we accept? In his initial remarks about the (alleged) asymmetry of counterfactual dependence, Lewis claims:

> In reasoning from a counterfactual supposition, we use auxiliary premises drawn from (what we take to be) our factual knowledge. But not just anything we know may be used… If the supposition were true, the future would be different and some things we know about the actual future might not hold in this different counterfactual future. But we do feel free, ordinarily, to use whatever we know about the past…. [I]n reasoning from a counterfactual supposition about any time, we ordinarily assume that **facts about earlier times are counterfactually independent of the supposition** and so may freely be used as auxiliary premises. (CDTA: 33; bold emphasis mine)

If we take Lewis's claims in this passage as restricted to our use of forward counterfactuals, then they have some plausibility. But the data to which Lewis appeals here are not conclusive in favour of Thesis F/F*. They are consistent with something much weaker—for example, that we keep the past relative to the A-time fixed in respect of its salient or significant features. As Jonathan Bennett convincingly remarks (convincingly, that is, if the topic is taken to be subjunctive conditionals of the 'forward' variety):

> The plain person using a subjunctive conditional has a *vague* thought of a world that does not significantly differ from the actual one until a divergence leading to the truth of his antecedent. (Bennett 2003: 218)

Bennett goes on to say that although *one* way to sharpen this vague thought is with (Lewis's) idea of a world that is exactly the same as the actual world up until a divergence leading to the truth of the antecedent, it is not the only way. Another way to sharpen the vague thought is with the idea of an 'exploding difference', that is:

> the idea of a world that is like [the actual world] in every respect we would ever think about and then suddenly, legally, and improbably embarks on a short course of events through which it becomes noticeably unlike [the actual world]. (Bennett 2003: 218)

In other words, whereas Lewis sharpens the vague 'does not significantly differ' into what we can call 'Exact Match', leading to Thesis F or F*, a rival way of sharpening the vague 'does not significantly differ' is into 'Exploding Difference'. And Exploding Difference, unlike Exact Match, does *not* support Thesis F or F*. What it supports is at most:

> Weakened Thesis F: When we evaluate forward counterfactuals, we keep the past, relative to the A-time, fixed *in salient respects*,

and

Weakened Thesis F*: When we evaluate forward counterfactuals, we keep the past, relative to the A-time, fixed *in salient respects*, with the possible exception of a transition period leading from the actual past to the fulfilment of the antecedent.

And neither Weakened Thesis F nor Weakened Thesis F* is sufficient to explain or support the intuition of the fixity of the past. Why so? Because the fixity intuition does not discriminate between salient features and insignificant features of the past. If you were asked: 'If a feature of the past is one "that you would never think about", does this mean that this feature of the past might not be fixed?', then surely your answer would be 'No'. (Nor, of course, is there any paradox in asking people whether features that they would never think about (cf. Bennett's phrase in the passage quoted above) are fixed, since they can coherently consider the general question without thinking about any of the relevant unthought-of features.) Yet if Bennett is right in suggesting that our ordinary counterfactual thinking (with regard to forward counterfactuals) does not discriminate between keeping the past before the A-time fixed in the sense required by Exact Match, and keeping the past before the A-time fixed in salient respects (Exploding Difference), then this casts doubt on the claim that our practice supports Thesis F/F*.

Speaking for myself, my intuition that the past is fixed—an instance of the intuition that Lewis seeks to explain by reference to features of our counterfactual thinking—is significantly more robust than my intuition that the past is to be kept fixed in the 'Exact Match' sense when evaluating forward counterfactuals. I think I could be persuaded that the notion of an exploding difference is sufficient to match my intuitions about the extent to which the past is kept fixed in the evaluation of forward counterfactuals. If I were so persuaded, then I would reject Thesis F and Thesis F*. But this would not in the least tempt me to give up my intuition that the past is fixed; nor do I think it should.

5 Objection 3: The 'transition period' and keeping the past fixed

My first two objections have been that neither Thesis F/F* nor Thesis B is sufficiently plausible to warrant the claim that we keep the past fixed when evaluating counterfactuals, whether these are forward counterfactuals (Thesis F/F*) or backward counterfactuals (Thesis B).

But a further problem for Lewis is evident, connected with the transition period that has already been mentioned. Famously, Lewis's account of counterfactuals delivers the result that, in perfectly ordinary cases (no time machines, black holes, weird possible worlds consisting of just one atom in the void, etc.), there are portions of the past that are *not* kept fixed when evaluating counterfactuals even under what Lewis calls 'the standard resolution' of their vagueness. These are the portions of the past that, according to Lewis's account, concern the transition period (or 'ramp', as Bennett (2003) calls it) from an initial divergence from the actual course of events to the fulfilment of the antecedent.

According to Lewis, the closest possible worlds relevant to the evaluation of a counterfactual, under the 'standard resolution' of vagueness, and assuming determinism, are ones in which the fulfilment of the antecedent typically comes about via a transition period that starts with a divergence from the actual course of events at some time *prior* to the time of the antecedent. [Under determinism, the divergence has to involve a breach of the actual laws of nature (hence, a 'miracle' in Lewis's quasi-technical sense of the term).[19]]

One might ask why Lewis needs the 'transition period': why can't the 'divergence miracle' itself be the fulfilment of the antecedent? We need the transition period, according to Lewis, in order to

> [avoid] abrupt discontinuities. Right up to *t*, the match was stationary and a foot away from the striking surface. If it had been struck at *t*, would it have travelled a foot in no time at all? No; we should **sacrifice the independence of the immediate past** to provide an orderly transition from actual past to counterfactual present and future. (CDTA: 39–40; bold emphasis mine)[20]

There are two serious problems arising from this concession, given Lewis's ambition to explain the fixity of the past in terms of counterfactual independence.

First, although in this passage Lewis speaks of the transition period as involving only the immediate past, it seems impossible that, on his own account, the transition period can be so confined. For example (as Jonathan Bennett has pointed out), the details of Lewis's own theory (in particular, his insistence on the paramount importance of minimizing big miracles) appear to lead inevitably to the result that if we evaluate a counterfactual with the antecedent:

(D) If dinosaurs had been roaming the earth today,

the closest possible A-worlds will be ones whose divergence from the actual course of events occurred millions of years ago (cf. Bennett 2003: 220).

The dinosaurs example is a dramatic case. But it is easy to construct examples where the requirement to ensure a smooth transition from the actual past to the fulfilment of the antecedent requires a 'ramp' from the actual course of events that starts months or years before the A-time. Consider, for example, counterfactuals with antecedents such as 'If the Duchess of Cambridge had given birth to a child in January 2012', 'If John F. Kennedy had been alive today', or 'If in 1930 the indigenous population of Japan had been three times what it actually was'.

If (like Lewis) we are committed to the idea that there must be a transition period in order to allow the fulfilment of the counterfactual antecedent to be 'smoothly

---

[19] Two possible worlds that have exactly the same past up until a certain time and then diverge cannot also have exactly the same deterministic laws. So, if the actual world is deterministic, any possible world that shares its past up until time *t* with the actual world and then diverges cannot have exactly the same laws as the actual world.

[20] It is true that these remarks are made by Lewis in connection with an analysis of counterfactuals that he rejects ('*Analysis 1*'). However, he endorses this feature of it—that is, its appeal to the transition period. (He rejects *Analysis 1* only because it builds the temporal asymmetry into the analysis by *fiat*, and favours his own analysis (*Analysis 2*) as yielding the asymmetry without doing so by *fiat*.)

grafted' on to the actual past, then it seems inevitable that in some cases this will require the transition period to begin days, months, years, centuries, or even millions of years before the A-time. At any rate, this consequence seems to be unavoidable on Lewis's own 'minimization of big miracles' account. If so, then it is impossible for Lewis to maintain, with an appropriate degree of generality, the thesis that the past relative to the A-time is kept fixed in the evaluation of counterfactuals (even forward counterfactuals). And if this is the case, then it seems impossible for him to maintain that our intuition of the fixity of the past is to be explained by the extent to which the past is kept fixed in the evaluation of counterfactuals.

The second problem is that even if the transition period *were* confined to the immediate past relative to the A-time, this would still undermine Lewis's explanatory project. Remember that Lewis aims to use considerations about the evaluation of counterfactuals in order to explain the fixity intuition—an intuition that entirely ignores the distinction between immediate and more remote past. Nobody thinks that although what happened last century, last year, last month, last week, or yesterday, is 'fixed', what happened in the past five minutes (or even the past five seconds) may still be 'open' simply because it occurred such a short time ago. The result is that there is an apparently fatal mismatch between, on the one hand, the extent of counterfactual independence that Lewis's theory can consistently attribute to the past, and, on the other hand, the fixity intuition that his account purports to explain. Even setting aside the arguments of §§3–4 above, the prospects look bleak for Lewis's attempt to explain the fixity of the past in terms of counterfactual independence.[21]

6 Objection 4: Lewis's retreat from 'counterfactual independence' to 'failure of (systematic) counterfactual dependence'

Lewis is curiously untroubled by the problem posed by the transition period. In discussing the class of exceptions that the transition period requires, he argues that,

---

[21] Lewis is committed, not only to there being, under determinism, prior changes to the past, but also to there being prior 'miracles'. Thus, given his own treatment of backward counterfactuals, he is committed (with regard to the notorious Nixon example (cf. CDTA: 43–48)) to the assertibility, *in standard contexts*, not only of:

(N1) If Nixon had pressed the button at *t*, then some events prior to *t* would have been different in certain respects from the way that they actually were,

and perhaps:

(N2) If Nixon had pressed the button at *t*, then a few extra neurons would have fired in his brain shortly before *t*,

but also of:

(N3) If Nixon had pressed the button at *t*, then the laws of nature would have been different in some respect resulting in a difference in the course of events before *t*.

I agree that both (N1) and (N2) are plausible. But that is because I think that they are plausible examples of backward counterfactuals asserted on the basis of back-tracking arguments. Lewis can't say that. (N3), on the other hand, is something to which Lewis is firmly committed (on the assumption of determinism). But (N3) is an intuitively bizarre counterfactual, and not one that would be commonly accepted as true except as a consequence of a philosophical theory.

even though they do represent a 'sacrifice' of the counterfactual independence of the past, they do not bring with them a type of counterfactual *dependence* of past on future that would, under the 'standard resolution', and in conjunction with his counterfactual theory of event causation, imply a *causal* dependence of past events on future ones, thus introducing unwanted (and unacceptable) cases of backward causation 'even in cases that are not at all extraordinary' (CDTA: 40).

Now, Lewis may or may not be right in claiming that he can avoid the conclusion that the transition period introduces cases of counterfactual dependence that his theory of event causation would be required to treat as cases of causal dependence, and hence as cases of backwards causation.[22] Leaving that aside, there is the further question why Lewis thinks that merely avoiding the kind of systematic counterfactual dependence that would (by his lights) amount to causal dependence defuses the problem posed by the transition period. In this connection, he writes:

> [W]e should **sacrifice the independence of the immediate past** to provide an orderly transition from actual past to counterfactual present and future. **That is not to say, however, that the immediate past depends on the present in any very definite way**. There may be a variety of ways the transition might go, hence there may be no true counterfactuals that say in any detail how the immediate past would be if the present were different. I hope not, since if there were a definite and detailed dependence, it would be hard for me to say why some of this dependence should not be interpreted – wrongly, of course – as backward causation… in cases that are not at all extraordinary. (CDTA: 40; bold emphasis mine)

And again, in the opening paragraph of CDTA:

> Suppose today were different. Suppose I were typing different words…. Would yesterday… be different? If so, how? Invited to answer, you will perhaps come up with something. But **I do not think there is anything you can say about how yesterday would be that will seem clearly and uncontroversially true**. (CDTA: 32; bold emphasis mine)

However, there are two reasons why this retreat does not help Lewis. The first is that it makes the letter of what Lewis says inconsistent. The retreat undermines his right to make assertions such as:

> The past would be the same, however we acted now. The past does not at all depend on what we do now. It is counterfactually independent of the present. (CDTA: 38)

> [I]f the present were different the past would be the same, but the same past causes would fail somehow to cause the same present effects. (CDTA: 34)

The second reason is that if the past would have been different *in some way or other* had the present been different, then there is a clear sense in which the past *does* depend counterfactually on the present. But if the past does depend counterfactually

---

[22] For some doubts, see Bennett (2003: 290).

on the present, how can the thesis that the dependence is not systematic (even if correct) save the proposed explanation of the fixity of the past?[23]

One might attempt to respond by pointing to the fact that, if the openness of the future really is a matter of the *systematic* counterfactual dependence of future events on the present, then, technically, there could still be an asymmetry between past and future as long as past events are not *systematically* counterfactually dependent on the present, even if they are not, strictly speaking, counterfactually *independent* of the present. And there is logical space for this in Lewis's account.[24]

However, although this revision would introduce *an* asymmetry of counterfactual dependence, I do not believe that it is one that can play the role of explaining the fixity intuition. If someone were to tell me that my intuition that the past is fixed either amounts to, or can be explained by, the idea that although the past would indeed have been different had the present been different, there is no *definite* way in which it would have been different, then I would find the suggestion totally mysterious. How can the idea that, if today had been different, then yesterday would have been different, although not in any definite or specifiable way, coherently be invoked to explain the idea that the past is fixed? (Even the idea that the past *might* have been different had the present been different—let alone the idea that it would (though unspecifiable) have been different—appears to conflict with the idea that failure of counterfactual dependence has anything to do with the fixity of the past.[25])

7 Objection 5: The substitution of 'causal independence' for 'counterfactual independence' would not help Lewis

Another possibility is that, in suggesting that the failure of counterfactual independence that is required by the transition period, as long as it does not also involve systematic counterfactual dependence, would not undermine his explanation

---

[23] Remember that what Lewis is attempting to explain, in terms of an asymmetry of counterfactual dependence, is an intuition of the fixity of the past. If he concedes, as he appears to do, that the past in the transition period *would have been different* had the present been different, although there is no *definite* way in which it would have been different, the concession undermines this explanatory project.

[24] See the characterizations of 'counterfactual dependence' and 'counterfactual independence' in Lewis (1973: 164–165; 168). Lewis there characterizes counterfactual dependence primarily as a relational characteristic of a *family* of propositions with respect to another family of propositions, whereas 'counterfactual independence' is characterized as a feature that a single proposition has with respect to a family of propositions. This leaves it open that a proposition about the past might not be counterfactually *independent* of a family of propositions about the present, and yet might also fail to belong to a family of propositions about the past that *depends* counterfactually on a family of propositions about the present. This would appear to be a case of failure of counterfactual independence without (systematic) counterfactual dependence.

[25] Bennett (2003: 290–291) makes the further objection that even if Lewis can maintain that the admission of the transition period introduces no systematic counterfactual dependence of earlier *events* on later *events* of a type that would cause trouble for his counterfactual account of event causation, this does not give Lewis a justification for rejecting the thesis that there is a systematic counterfactual dependence of earlier facts or states of affairs on later ones. But to allow the latter, Bennett plausibly maintains, makes trouble for Lewis's attempt to explain the fixity of the past in terms of a failure of counterfactual dependence.

of the fixity of the past, Lewis is, implicitly, appealing to the following: the idea that, even if the past may not be *counterfactually* independent of the present, it is nevertheless *causally* independent of the present—in the sense that, had the present been different, this would not have *caused* the past to be different.

But could Lewis legitimately appeal to these considerations about causation in his explanation of the asymmetry of openness? I believe that he could not. Lewis makes it explicit that he wishes to explain both the temporal asymmetry in the direction of causation *and* the 'asymmetry of openness' in terms of an asymmetry of counterfactual dependence (CDTA: 35–36). If it should turn out that the asymmetry of openness is to be explained in terms of an asymmetry of causal dependence that is more fundamental than the asymmetry of counterfactual dependence, then Lewis's explanatory project would be undermined.

8 Concluding remarks

My criticisms of Lewis's attempt to explain the fixity intuition in terms of counterfactual independence naturally prompt two questions. One concerns the extent to which my criticisms can be generalized. Might there not be, for all that I have said here, some modification of Lewis's account that would escape my criticisms? The second question is: if the fixity intuition is not to be explained in terms of counterfactual independence, how *is* it to be explained? To conclude this paper, I briefly address these questions in turn.

Early in this paper I identified two candidates for the claim that the past is counterfactually independent of the present: Thesis F and Thesis B. I argued (§3) that Thesis B is false, and I expressed scepticism about Thesis F and its modification (favoured by Lewis), Thesis F* (§4). In addition, I argued that the adoption of Thesis F* is inimical to the project of explaining the fixity intuition in terms of the counterfactual independence of the past with respect to the present, on the grounds that Thesis F* requires exceptions to the 'counterfactual independence' rule that undermine the proposed explanation (§5). If my arguments against Thesis B and against the employment of Thesis F* are cogent, then the only option, for the defender of the claim that the fixity intuition is to be explained in terms of counterfactual independence, appears to be to abandon both Thesis B and Thesis F*, and argue (as Lewis does not) for the following two claims: (i) that the *unmodified* version of Thesis F represents our practice in the evaluation of forward counterfactuals, and (ii) that Thesis F (without the support of Thesis B) is sufficient to explain the fixity intuition. Although I shall not argue for it here, I think that the prospects for success in either of these undertakings are not at all promising.

The second question is: if the fixity intuition is *not* to be explained in terms of the counterfactual independence[26] of the past with respect to the present and future, what does explain the fixity intuition? In this paper, I have said little about what the fixity intuition involves, except for the following:

---

[26] Or failure of (systematic) counterfactual dependence (cf. §6 above). For simplicity, I omit this qualification in the remainder of this section.

The fixity intuition does not discriminate between the immediate past and the more remote past (§5).

The fixity intuition does not discriminate between salient features of the past and insignificant features of the past (§4).

The fixity intuition is not satisfied by the claim that although the past does depend counterfactually on the present, the dependence is not systematic (§6).

The fixity intuition involves an asymmetrical attitude to past and future (although the asymmetry need not be totally exceptionless).

To these points, we might add, following Lewis:

The fixity intuition associates fixity with uniqueness, and openness with multiplicity (cf. the quotations from Lewis in §1 of this paper).

The fixity intuition associates fixity with being settled, and openness with being unsettled (ibid.).[27]

These points appear to leave open the possibility that the fixity intuition is simply the intuition that the past (unlike the future) is *causally* independent of the present and future. I have argued that Lewis himself cannot legitimately appeal to this 'causal independence thesis', because his project is to explain both causal independence and the fixity intuition by appeal to a more basic notion of counterfactual independence (§7). However, this leaves it open that the fixity intuition that Lewis tries and (if I am right) fails to capture is really nothing more than an intuition of causal independence.

I admit that it is natural to associate the idea that the past is fixed with the idea that nothing that we can now do, and nothing that can now happen, could have any *effect* on the past: that the past is, in that sense, immutable. Nor does the possibility of backwards causation necessarily threaten this association between causal independence and fixity, since it seems plausible to suppose that, were there to be a case of backwards causation, it would represent an exception to the general rule of the fixity of the past.

Nevertheless, I am sceptical about whether the notion of causal independence can be the key to the notion of the fixity of the past. The reason is that, intuitively, causal independence appears to be neither necessary nor sufficient for fixity.

Reflection on fatalistic thinking suggests that causal independence is not *necessary* for fixity—or, to put it another way: that fixity is compatible with causal dependence. By killing his father, Oedipus, we may suppose, brought it about that he married his mother: his incestuous marriage was causally dependent on his previous act of parricide. Yet (to the extent that we can take the fatalistic story seriously) it seems that we can accept that causal claim, yet still question whether

---

[27] Fischer (1994, 2011), followed by Holliday (2012), uses the expression 'The Principle of the Fixity of the Past' to describe a thesis about a limitation on our abilities, to the effect that we cannot do anything such that, were we to do it, the past would have been (or have had to be) different. However, Fischer's principle seems more appropriately regarded as a (controversial) claim about a *consequence* of the fixity of the past, rather than an expression of what it means to say that the past is fixed.

Oedipus had before him, at any time in his existence, a future that was 'open' rather than 'fixed'.

In addition, it seems that causal independence is not sufficient for fixity. A future that is a completely random continuation of the present, if we can coherently envisage such a thing, is surely a future that is causally *independent* of the present. Yet such a 'random' future, far from being a 'fixed' future, would seem to be a paradigm of one type of openness, even if it represents a type of openness that brings with it no prospect of control over the course of events.

If this is right, and neither counterfactual independence nor causal independence is the key to the explanation of the fixity intuition, what is? Unfortunately, I do not have an answer to that question. However, my aim in this paper has been, not to solve this problem, but the more modest project of attempting to show that the idea that the past is counterfactually independent of the present and future cannot be invoked to give a satisfactory account of the intuition of the fixity of the past. As a consequence, Lewis's claim that 'the mysterious asymmetry between open future and fixed past is nothing else than the asymmetry of counterfactual dependence' (CDTA: 38) cannot be sustained.

References

Barnes, E., & Cameron, R. (2011). Back to the open future. *Philosophical Perspectives, 25*(1), 1–26.

Bennett, J. (1984). Counterfactuals and temporal direction. *The Philosophical Review, 93*(1), 57–91.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*. Oxford: Clarendon Press.

Downing, P. B. (1959). Subjunctive conditionals, time order, and causation. *Proceedings of the Aristotelian Society, 59*, 125–140.

Fischer, J. M. (1994). *The metaphysics of free will*. Oxford: Blackwell.

Fischer, J. M. (2011). Foreknowledge, freedom, and the fixity of the past. *Philosophia, 39*(3), 461–474.

Holliday, W. (2012). Freedom and the fixity of the past. *The Philosophical Review, 121*(2), 179–207.

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

Lewis, D. (1973). Causation. *The Journal of Philosophy, 70*(17), 556–567. Reprinted, with additional *Postscripts*, in Lewis 1986. (Page references in the text are to the reprinted version).

Lewis, D. (1979). [CDTA] Counterfactual Dependence and Time's Arrow. *Noûs, 13*(4), 455–476. Reprinted, with additional *Postscripts*, in Lewis 1986. (Page references in the text are to the reprinted version).

Lewis, D. (1986). *Philosophical Papers* (Vol. 2). Oxford: Oxford University Press.