Pre-emption cases may support, not undermine, the counterfactual theory of causation

Robert Northcott, Birkbeck College

Abstract

Pre-emption cases have been taken by almost everyone to imply the unviability of the simple counterfactual theory of causation. Yet there is ample motivation from scientific practice to endorse a simple version of the theory if we can. There is a way in which a simple counterfactual theory, at least if understood contrastively, can be supported even while acknowledging that intuition goes firmly against it in pre-emption cases – or rather, only in some of those cases. For I present several new pre-emption cases in which causal intuition does *not* go against the counterfactual theory, a fact that has been verified experimentally. I suggest an account of framing effects that can square the circle. Crucially, this account offers hope of theoretical salvation – but only to the counterfactual theory of causation, not to others. Again, there is (admittedly only preliminary) experimental support for this account.

1. Introduction

In its simplest version, the counterfactual theory of causation states that an event *c* causes an event *e* if and only if *e* is counterfactually dependent on *c*. It was first put forward in its modern form by David Lewis in (1973). Already in that paper, Lewis endorsed only a refined version of the theory in response to a particular kind of pre-emption case. This has proved prescient in two respects: first, that many refinements of the simple theory would be offered; and second, that pre-emption cases, themselves of many different varieties, would prove the biggest problem for these refinements.

Pre-emption cases have been taken by almost everyone to imply the simple counterfactual theory's unviability. I think this is a mistake. The reason is that there are many pre-emption cases in which causal intuition is in accordance with, not contradictory of, the simple theory,

a fact that has been verified experimentally. This creates an inconsistency: in some pre-emption cases intuition favors the simple theory but in others not. Yet the different cases are structurally identical in all theoretically relevant respects. Previous work has used an emphasis on norms and defaults to try to break such impasses (section 5). I will suggest what seems to me an alternative, preferable approach. It has the unexpected consequence that pre-emption cases, taken overall, far from telling against the simple counterfactual theory, in fact tell in favor of it.[1]

Science will enter this story in three ways. The first way is as motivation: the simple counterfactual theory is implicitly endorsed by overwhelming scientific practice. Indeed, the great asymmetry in this respect between it and other theories of causation is often underappreciated. The second way is via psychological experiment, which is the means of establishing the inconsistent causal intuitions across pre-emption cases. The third way is via psychology too: there is also (admittedly only preliminary) experimental support for this paper's resolution of this inconsistency.

Some preliminaries: I will interpret the counterfactual theory as an ontological thesis. Thus, according to it, causation just is counterfactual dependence. As a result, science can be assured that causal knowledge will yield knowledge about interventions. On this view, the infusion of pragmatic factors via framing effects (section 6) does not affect the objective causal facts, and thus science should still put the same weight on causal knowledge as ever. The pragmatic factors affect only our causal judgments.[2] Which brings us to causation as a concept. As a descriptive psychological matter, our concept of causation seems to be, in

[1] In order to reach this result, it will be necessary to interpret the counterfactual theory contrastively, as will become apparent below.

[2] Following others, I will use 'causal judgment' and 'causal intuition' interchangeably.

Brian Skyrms's phrase (1984), an "amiable jumble". Indeed, the judgment reversals in pre-emption cases to be discussed below are themselves evidence for this. But as a normative philosophical matter, should we reform this amiable jumble along counterfactualist lines? I discuss this in section 9, coming out cautiously in favor of indeed so reforming it. Finally, is the counterfactual theory a reductive analysis of causation, or does it instead merely state a non-reductive necessary connection? The answer turns on the controversial issue of whether a reductive semantics can be given for the relevant counterfactuals. Yet the endorsement of the counterfactual theory by scientific practice remains the same either way. And although this paper's analysis of judgment reversals in pre-emption cases can be co-opted by a reductionist account, it too stands independently of whether causation reduces. Accordingly, I remain neutral on the reductionism issue.

Of course, pre-emption cases are not the only challenge to the counterfactual theory. Still, as the most influential of these challenges, there is value in targeting them specifically.

2. A role for science I: Motivation

Defined as counterfactual dependence, causation is intimately linked to practical utility. In Nancy Cartwright's phrase (1979), knowledge of causation becomes the key to "effective strategies". In particular, causation is the guide to effective interventions into, and thus manipulations of, the world. It allows us to infer what would have changed something in the past, thus allowing us to assign causal responsibility. (Indeed, etymologically the function of assigning responsibility is the very root of the word 'cause'.) This in turn explains why we, and science, value causal knowledge so highly, and is at the heart of why we should distinguish causation from mere correlation. The counterfactual theory locks in causation's difference-making feature.

These considerations are familiar, of course. It is not a surprise that they lie at the heart of methods of causal inference in science, which are all built around a difference-making conception. An essential part of these methods is holding 'all else equal', so that changes in the effect variable can be ascribed exclusively to changes in the cause variable. This is the normative ideal that motivates the use of controlled experiments, natural and quasi-experiments, and a range of sophisticated statistical techniques (Angrist and Pischke 2009, Morgan and Winship 2007, Holland 1986). That is, methods of causal inference across almost all biomedical and social sciences – that is, arguably the large majority of science that is actually done – support a counterfactual approach, either implicitly or explicitly. As well as ubiquitous practice, all medical and social methods textbooks are evidence for this too. A simple counterfactual test for causation is also widely used in law.

In philosophy, it is precisely the same ideal that is built into the very definition of causation under the counterfactual theory, according to which we must assess what would happen if we changed the cause but kept everything else the same (except for the downstream effects of that change). Often this difference-making conception takes probabilistic form, i.e. a cause is something that changes the probability of its effect (section 7). The same difference-making conception also lies at the heart of the recent and influential causal modeling literature and of theories of causation closely related to it (Pearl 2009, Spirtes et al 2000, Woodward 2003).[3]

[3] The causal modeling literature works with a probabilistic version of difference-making. In this paper, following the relevant metaphysics literature, I concentrate on the deterministic case. But, as discussed below (section 7), I think an approach analogous to this paper's can be profitably applied to the probabilistic case too. Also, the causal modeling literature, like many methods in social science, operates at the type level, whereas the most discussed pre-emption cases in the metaphysics literature have been at the token level. But the counterfactual account, if accepted, should apply to token and

The sophisticated algorithms that have been developed for causal inference from statistical data all assume such a conception, for instance. Indeed, they could not have been developed in the first place without a prior commitment to it. In turn, their value also rests on this conception because that is what ensures that the causes they discover have the practical implications mentioned above for manipulation and control.[4]

The counterfactual theory has further attractions. Its simplicity and overwhelming relevance to practical concerns readily explain why causation has such a central role in our conceptual repertoire. Moreover, the extensive philosophical work on counterfactual conditionals imposes a helpful discipline on causation's analysis. That in turn enables us to illuminate numerous methodological controversies within science, witness the flourishing in recent philosophy of science of applications of James Woodward's theory of causal explanation.

There is no such endorsement for rival theories of causation precisely because they do not define causation in terms of difference-making. Consider, for example, the physical connection view, the most popular version of which defines causation instead as the transfer of some conserved quantity such as energy or momentum (Dowe 2000). As has often been pointed out, this sometimes endorses as causal cases where no difference is made to the effect. It also refuses to endorse as causal cases of negative causation such as prevention, even when a difference is made. Yet prevention is as central as positive causation to

---

type levels alike. Accordingly, its utility at the type level motivates defending it at the token level, and in turn defending it at the token level is necessary to defending it at the type level.

[4] Some work has questioned whether even causes so understood really do license interventions in the way claimed, given various complexities of implementation and extrapolation. The debate continues (Cartwright 2007, Pearl 2010). But that debate concerns only whether to insist on a more local or contextualized understanding of causal knowledge; it does not question causation as difference-making.

manipulation and control, and to assignment of responsibility. Its detection is correspondingly central to science, and science's methods of causal inference reflect that.

Similar negative remarks apply to intricate versions of the counterfactual theory itself. Brad Weslake's recent account (forthcoming), for instance, is one of the most impressive. Yet its definition of causation takes three pages to state and many more to explicate, and each of its four conditions requires many technical notions to be defined, which leaves it far removed from the criteria of causal inference found in the tools of science or from the motivation for those criteria. The definition is so intricate because it is designed ingeniously to accommodate even those causal judgments that contradict the simple counterfactual theory, including especially those from various pre-emption cases. It is this that forces it to become extremely complex – but thus to lose its reliability as a guide to the efficacy of interventions.

When defending the counterfactual theory, we have several options. One is simply to stipulate it, and to appeal to the wider benefits mentioned above in order to justify ignoring awkward intuitions in pre-emption and other cases. On this view, rather than a 'Socratic' approach of trying to reconcile every different aspect of our understanding of a concept, we should instead pursue a 'Euclidean' approach of building more narrowly on a core definition (Glymour 2004). One inspiration is the way physics co-opted the concepts of 'work' and 'energy' to construct theoretical frameworks that eventually – and usefully – left many of the folk connotations of those concepts far behind. Might not the same be desirable for causation?

But I do not advocate simple stipulation here. Even if it were fruitful scientifically, such an approach seems unlikely to find favor among metaphysicians. The latter feel obliged to take

seriously, one way or another, the data of our awkward (for the counterfactual theory) intuitions in pre-emption cases. One way to do that, as noted, is to accommodate those intuitions as best we can by complicating our definition. I will advocate an alternative approach: to stick to a simple definition and show how the awkward intuitions can be explained away. Of course, the task is then to be able to explain away the awkward intuitions in a principled and convincing manner. The reward would be to maintain the close connection between causation and practical efficacy that is so prized by science.

3. Some pre-emption cases

1) *Meteorites*. By coincidence, two different meteorite rocks fall from the sky towards a window. The first rock reaches it fractionally before the second does. When the first rock hits the window, the window shatters. If the first rock had not got there first, the second one would have hit and shattered the window instead.

Did the first rock cause the window to shatter? In this situation, our clear intuition is that it did. Let *c* = the first rock falls, and *e* = the window shatters. Then *c* causes *e* – and yet *e* is not counterfactually dependent on *c*. Therefore, causation and counterfactual dependence come apart. Structurally, there are two potential causes of the effect, one of which pre-empts the other. We endorse the pre-emptor as a cause. Many pre-emption cases have been constructed but they all share this core feature of asymmetric overdetermination – and, in the cases that the literature has highlighted so far, a deviation of causal intuition from the simple counterfactual theory. But now consider some other cases with the same pre-emption structure.

2) *Hotel Coupon*. Fred is staying at a hotel that gives out a coupon to all guests, entitling them to a free drink at the hotel bar. A freshly arriving guest has checked in and so is now entitled to a coupon. Upon leaving the hotel, Fred hasn't used his coupon, so he returns it to reception. The receptionist then hands this particular coupon to the freshly arriving guest, who is on his way to the hotel bar. That guest uses this coupon to get himself a free drink. The receptionist has an ample supply of coupons: if Fred had not returned his coupon, the receptionist would have given the new guest one of the other coupons.

Did Fred returning his unused coupon enable the new guest to get a free drink? Our intuition is that it did not. (I defend this claim below.) Structurally speaking, Fred returning the coupon is a pre-emptor, the receptionist taking a coupon from the stack on their desk is a back-up, and the guest getting a free drink is the effect. Yet this time, unlike in Meteorites, the pre-emptor is not judged a cause.

3) *Baseball Fielder*.[5] A baseball batter hits a pitch. The ball starts off travelling in the direction of a full moon in the sky. A fielder catches the ball. If the ball had not been caught by the fielder, it would have fallen to the ground shortly afterwards.

Did the fielder's catch prevent the ball from reaching the moon? Our strong intuition is that it did not. But once again, this implies that a pre-emptor is not judged a cause: this time the pre-emptor is the fielder catching the ball, the back-up is the Earth's gravity dragging the ball down, and the effect is the ball not reaching the moon.

[5] A version of this example was first put forward in (McDermott 1995) and (Collins 2000).

4) *War*. Two countries, Big and Little, have had a long-running and bitter political dispute. Big has mobilized its army. It delivers an ultimatum to Little, with a threat of war if the ultimatum goes unanswered by a deadline of 9 o'clock. The deadline arrives, and there has been no reply from Little. Accordingly, Big commits itself to invading and begins its preparations. Before those preparations are completed, Little pre-emptively launches its own invasion, and the war begins. If Little had not struck first, Big would have invaded as planned shortly afterwards.

Did Little's pre-emptive strike cause the war? Intuitively – no it did not. (Again, I defend this claim about intuition below.)

5) *Soccer*. John is a professional soccer player. He is playing a one-on-one challenge match against a 16-year-old. Not surprisingly, John is winning easily and requires only one more goal to finish the match. Unexpectedly, a weather system suddenly picks up and a rumble of thunder is heard. The match is immediately ended because of the risk of lightning, as agreed in the rules. The rules of the match also state that in the event of a premature finish like this, if one player is winning easily he is declared the victor. As a result, John is declared the victor. If the weather had not gotten worse, John would have won soon afterwards anyway.

Did the weather getting worse cause John's victory? Again, intuitively not.

Comparing case 1 with cases 2 to 5, there is clear inconsistency. In Meteorites, our causal judgment contradicts the counterfactual theory; but in Hotel Coupon, Baseball Fielder, War, and Soccer, it is in accordance with it. Obviously, the bare counterfactual theory itself has no

way to account for this inconsistency. But the problem is wider than that because no other theory of causation can account for it either. The inconsistency is a problem for everyone.

To see this, Table 1 compares the five cases. With respect to all of the elements that different theories take to be relevant to causation, the cases are identical: there is the same pattern of counterfactual dependence; equally reliable nomological regularities hold; and there are analogous physical connections too. Yet when it comes to deciding whether *c* is a cause of the final effect *e*, still our judgment reverses. The only difference between the cases lies in what should be irrelevant window-dressing, namely the background framing story: why should it matter whether the very same structure is presented in terms of meteorites, baseballs, soccer matches or hotel coupons?

Table 1: Pre-emption cases compared

	Meteorites	**Hotel Coupon**	**Baseball Fielder**	**War**	**Soccer**
Cause *c*	First meteorite rock falls	Fred returns his spare coupon	Fielder prevents the ball passing	Little's pre-emptive strike	Weather gets worse
Salient contrast C_0	First meteorite misses Earth	Fred keeps it in his pocket	Fielder lets the ball pass	Little remains peaceful	Weather stays good
Effect *e*	Window shatters	New guest gets a free	Ball does not reach the moon	War begins	John wins the game

		drink			
Salient contrast E_0	Window is still intact	New guest has to pay for his drink	Ball does reach the moon	Peace continues	The 16-year-old wins the game
Counterfactual dependence?[6]	No – C_0 does not imply E_0	No – C_0 does not imply E_0	No – C_0 does not imply E_0	No – C_0 does not imply E_0	No – C_0 does not imply E_0
Appropriate transmission of conserved quantity?	Yes – the rock transmits energy to the window	Yes – there is a continuous physical process connecting Fred handing	Yes – there is a continuous physical process by which the fielder keeps	Yes – there is a continuous physical process connecting	Yes – there is a continuous physical process connecting

[6] The dependence here is that between the contrasts C_0 and E_0. This can be thought of as a clarification of the one usually quoted, i.e. that between the simple negations not-*c* and not-*e*. It is arguably superior as it incorporates our judgment's contrastive sensitivity (Maslen 2004, Schaffer 2005, Northcott 2008). Introducing contrasts does raise the issue though of which of the many possible specifications of contrast should be taken as salient (McDonnell 2017). The literature generally agrees that this is determined by context but that there is no single objective procedure for exactly how, and so that these determinations must be assessed case by case. For our purposes, what matters is whether the experimental prompts imply contrasts C_0 and E_0 sufficiently clearly that in each case whether there is dependence (or not) between contrasts is unambiguous. I think that they do, and so we may proceed using the contrasts cited in the table.

		the coupons back to the new guest getting his free drink	the ball on the ground[7]	the pre-emptive strike to the declaration of war	the thunder to the game being ended
Appropriate nomological regularity?	Yes – the laws imply that whenever (ceteris paribus) the rock hits the window, the window shatters	Yes – the laws imply that whenever (ceteris paribus) Fred hands the coupons back, the new guest gets his free drink	Yes – the laws imply that whenever (ceteris paribus) the fielder prevents the ball passing, it does not reach the moon	Yes – the laws imply that whenever (ceteris paribus) there is a pre-emptive strike, war breaks out	Yes – the laws imply that whenever (ceteris paribus) there is thunder, the game is ended
Intuitive judgment: does *c* cause *e*?	**Yes**	**No**	**No**	**No**	**No**

[7] Baseball Fielder is a case of prevention rather than positive causation. Thus, although a physical-connection view would indeed endorse that the fielder caused the ball to stay on the ground, strictly speaking it would join intuition and the counterfactual theory in denying that the fielder caused the ball not to reach the moon. A similar point might apply to War too if the relevant causal chain includes double-preventions. But in any case, the physical-connection view still cannot explain the judgment reversal between Meteorites and Hotel Coupon and Soccer.

4. A role for science II: Checking intuitions

Are our causal judgments in these various pre-emption cases indeed as I have claimed? This can be tested. Accordingly, I ran experiments that presented, among others, precisely the five vignettes quoted above (Northcott 2011). The results confirm the judgment reversals. In brief: respondents answered on a Likert scale of 1-7 whether they thought *c* was a cause (or preventer), where 1 = 'Strongly Agree' through to 7 = 'Strongly Disagree'. Samples were around 40, and the deviation in mean responses from the neutral score of 4 was statistically significant in each case. Those mean responses were: Meteorites 2.24; Hotel Coupon 5.59; Baseball Fielder 5.80; War 5.02; Soccer 4.62.

These data support the claim that judgment reversals across different pre-emption cases are a real phenomenon. A complete account of causation needs to be able to account for them. So far, none can.

Note that the judgments against the pre-emptor being a cause (or preventer) in Hotel Coupon and Baseball Fielder were just as strong – i.e. far from the neutral score of 4 – as was the judgment in favor of the pre-emptor being a cause in the paradigm case of Meteorites. The judgments in War and Soccer were also far from 4. Yet Hotel Coupon, Baseball Fielder, War and Soccer are not outré cases of magic in strange worlds; they are perfectly quotidian. They cover a range of types: positive causation, negative causation, and 'enable'. Accordingly, it can no longer simply be said that pre-emption cases tell against the counterfactual theory; it is more accurate to say that only some of them do. The impression otherwise is an artefact of the one-sided range of cases that have achieved wide circulation.

5. Salvation for the counterfactual theory?

Two possible responses to this evidence are primitivism and pluralism. I return to them below (section 8). Another response is to postulate some as yet unknown theory sensitive to a factor that reveals the different cases to be dissimilar. But no such theory is currently on offer (although see below). So, the response I will pursue instead is to recognize that, for the purposes of metaphysical theorizing, we must disregard at least some of the divergent intuitions. What might govern this in a principled way? An obvious move is to attend to the cognitive science of our causal judgments, and in particular to examine how they might be distorted by framing effects.[8] Of course, this requires a psychological theory to explain the framing effects, plus evidence for that theory. I address that shortly (section 6). But first, what might follow if we did manage to establish such a psychological theory?

The answer is salvation for the simple counterfactual theory of causation. This is because it would be granted a lifeline for wriggling free from pre-emption counterexamples. In particular, this will be so if the intuitions we are licensed to disregard turn out to be precisely those problematic (to the counterfactual theory) ones that endorse pre-emptors as causes.

Why salvation only for the counterfactual theory – why not for rival theories too? The answer is that all of the judgment reversals occur in pre-emption cases, which are problem cases only for the counterfactual theory. Therefore, only the latter is offered the happy prospect of its

[8] Others have also noted the desirability of an account of framing effects, although coming from a different perspective than this paper's (Collins et al 2004, 37; Hitchcock 2006, 428). L.A. Paul comments: "Philosophy … involves the construction of models and takes ordinary judgments to be constraints on such models, and hence needs to attend to the cognitive science of ordinary judgments" (2010, 475). Ernest Sosa (2007, 99-107) highlights the need for an error theory to explain divergent intuitions about individual cases. True, strictly speaking in the pre-emption experiments in the text the intuitive response in each particular case is uncontroversial; rather, the divergence is between cases. But this paper's analysis is in the spirit of Sosa's (and Paul's) suggestions.

problem cases being defused. An analogous maneuver is certainly open to other theories too. What would be required are judgment reversals in those other theories' own problem cases. Alas, neither I nor (to my knowledge) anyone else has been able to establish these, although I certainly have no proof that it cannot be done. But until it is, the investigation of framing effects offers hope only to the counterfactual theory.

The crucial point is that it is therefore an advantage for a theory if its own problem cases engender judgment reversals. First, its problem cases, or more precisely the judgment reversals between them, now become equally problems for everyone else too. And second, it offers a route by which the inconvenient judgments in those problem cases can potentially be explained away as judgment errors.[9]

To be clear – it is no strike against this strategy to note the incorrigibility of our intuition that the first meteorite rock did indeed cause the window's shattering. L.A. Paul and Ned Hall remark: "There is thus little room—none, really—for arguing that our intuitive judgments about [pre-emption cases] should not be taken seriously, or can be explained away as resulting from misleading or overly unfamiliar features." (2013, 132) But, in reply: first, as we have seen, intuitions in other pre-emption cases are equally firmly in favor of the counterfactual theory so, absent primitivism or pluralism, a firm intuition will have to be disregarded one way or the other. And which features are the "misleading" ones anyway: those in Meteorites or those in Hotel Coupon, Baseball Fielder, War and Soccer? That is precisely what is at issue. Second, anti-counterfactualist intuitions in Meteorites and similar cases should indeed be taken seriously. That is why simply stipulating the counterfactual

[9] To be sure, other theories too could use framing effects to explain away the judgment reversals in pre-emption cases. But, unlike for the counterfactual theorist, that would still leave them having to solve the additional problem cases specific to them.

dependence theory is unsatisfactory. Instead we must explain away the troublesome intuitions in a principled manner, with supporting evidence. But there is no reason why the possibility of incorrigible causal illusions should somehow be ruled inadmissible, any more than are incorrigible optical ones.

There is one strand of the causation literature that does already address the issue of judgment reversals square on. I have in mind work that explores the relevance of norms and defaults. A leading instance of this approach will illustrate. Joseph Halpern and Christopher Hitchcock (2015) note that while any situation typically features many counterfactual dependencies, our judgment picks out as causal only a few of those dependencies, as when we blame a fire on a short circuit but not on the oxygen in the air. Building on both empirical evidence and previous philosophical work, they provide a sophisticated analysis of what determines our causal judgment, presented in terms of different worlds ordered by their normality along several dimensions. This analysis can then be used to differentiate between situations that are isomorphic with respect to counterfactual dependence, in a way that tracks our causal judgment. Might this analysis, or one in a similar vein, be the key to explaining the judgment reversals in our pre-emption cases?

In reply, Halpern and Hitchcock do not explicitly apply their formalism to cases of late pre-emption but their basic account would seem to count both pre-emptor and back-up as causes in Meteorites. A further variable would need to be added to the causal model to break this symmetry. But perhaps this could be done in a principled rather than ad hoc way. If so, there would then be some hope of tracking the judgment reversals because in the non-Meteorites examples the back-up causes – but not the pre-emptors – are arguably more normal than their associated contrasts, in which case, as desired, only the pre-emptors would be endorsed as

causes.[10] But even if the details here could be worked out, still by this paper's lights two worries would remain. First, although I spoke above of their normality ordering as being a theory of causal judgment, Halpern and Hitchcock interpret it instead as a theory of 'actual causation'. That is, they posit a second kind of causation over and above simple counterfactual dependence. The formalism of normality orderings thus, in their interpretation, defines a metaphysical theory not a psychological one. I discuss this choice further in section 9 below. I argue there in favor of the psychological option, in which case Halpern and Hitchcock's framework would be re-interpreted as an interesting contribution to relevant psychological theorizing (section 6). Second, Halpern and Hitchcock's theory is a complex version of the counterfactual one. As argued earlier, I think this very complexity counts against it.

Similar remarks apply to other work rooted in the contextualist/interventionist approach derived from a structural equations analysis of causation. Woodward himself discusses how judgments in different cases, including different pre-emption cases, are sensitive to various psychological and other factors (2003, 85-90). He too defines a notion of 'actual causation', distinct from simple counterfactual dependence, that seeks to track our causal judgments. Many have pursued and refined this approach in various ways, defining and making use of sophisticated formal machinery (Hitchcock 2001, Halpern and Pearl 2005, Woodward 2006, Pearl 2009 ch 10, Fenton-Glynn 2017). As noted, many from the same tradition have also explored the role of defaults and norms (Hall 2007, Hitchcock 2007, Menzies 2007, Hitchcock and Knobe 2009, Blanchard and Schaffer 2017). For our purposes, this literature has similar strengths and weaknesses as Halpern and Hitchcock (2015). So far as I am aware, none of it advocates the discounting of certain causal judgments in the service of a simple

[10] I thank an anonymous referee for highlighting this possibility.

counterfactual theory, in the manner of this paper.[11] Rather, it is always steadfastly accommodationist of those judgments to some degree or another.

What differences between the various cases explain our divergent intuitions? I discuss this shortly (section 6). Here, note that one possibility turns out to be that the relevant differences are purely some subtleties of the counterfactual structure, for instance what we might call counterfactual robustness, i.e. the perceived inevitability of the effect. If so, one option would then be to incorporate these subtleties into our metaphysical theory. The resultant theory would still be a counterfactual one but now more complex than simple dependence. In effect, it would be an instance of the 'actual causation' approach from above that potentially could explain the judgment reversals in our examples.[12] A second option, by contrast, would be to stick to simple counterfactual dependence and to incorporate factors such as counterfactual robustness not into metaphysical theory but instead into a psychological theory of framing effects. In this paper I will pursue this second option because, roughly speaking, I think the simple dependence account is independently plausible and coheres better with scientific practice (sections 2 and 9). Moreover if, as seems likely, yet further complexities become necessary to track causal intuition fully, the first option will cohere with scientific practice ever worse as metaphysical theory is forced to become ever more Byzantine. Better – and more plausible – to transfer the complications, as it were, from the metaphysics to the psychology.

[11] Occasionally, this possibility is briefly mentioned or implied, but not pursued (e.g. Blanchard and Schaffer 2017, 197-8). Blanchard and Schaffer do use psychological theory to disregard some causal judgments in cases of causal selection.

[12] I thank an anonymous referee for highlighting this possibility.

Finally, on a separate note, causation has long had a close connection to the apportionment of moral and legal responsibility. Might this tell against the simple counterfactual theory? Suppose, for instance, that two vandals throw bricks and the first one's brick shatters a window just before the second one's would have. On this paper's view, the first brick does not cause the window-shattering – yet it seems strange to deny here that the first vandal is morally responsible.

In reply: first, the situation is reversed in Hotel Coupon, Baseball Fielder and War, in all of which the pre-empting agent is arguably *not* morally responsible – thus now in accordance with the simple counterfactual theory. Second, moral responsibility anyway deviates from token causal responsibility in several systematic ways: it takes account of type as well as token considerations, especially in probabilistic cases, as when risky actions are blameworthy even if they don't lead to disaster in a particular case; agents' intentions are morally significant, again independently of actual physical outcomes; and in the case of legal responsibility (and perhaps moral too), the institutional and historical context is relevant too. All of this muddies the water because it shows that moral responsibility and our judgments of it are sensitive to a range of factors to which causal responsibility is not, and so we should expect some divergence between the two. Overall, it is not clear that causation's connection to moral responsibility bears either way on the simple counterfactual theory.[13]

6. A role for science III: Framing effects

Can we make good on the promissory note of a theory of framing effects? The central idea is that our causal judgment is sensitive to many things left unspecified by metaphysical theory.

[13] Arguably, a simple counterfactual theory does usefully illuminate the analysis of *harm* (Northcott 2015).

In particular, in common with our perceptual faculties generally, it is greatly influenced by the manner in which a particular situation is presented. The vivid salience of some rather than other aspects may trump any merely verbal stipulation.[14]

The basic suggestion here is that judgments that pre-emptors are causes should be seen as analogous to optical illusions, and an error theory may explain why we have these illusions – and thereby also point the way to cases in which we should expect veridicality rather than illusion, i.e. in which our judgment does not endorse pre-emptors as causes. The truth of this suggestion is quite consistent with there being more cases of illusions than of veridicality (assuming raw numbers can even be meaningfully assessed), just as the fact that some optical illusions are ubiquitous does not show that they are not nevertheless illusions. One thing that does matter is that there is veridicality where the error theory predicts; that is, a test is whether we can use the error theory reliably to engineer cases in which pre-emptors are judged not to be causes.

One such error theory, for example, is that, given that overdetermination is rare, a good heuristic for our judgment is for it to associate causation with an event that *in non-overdetermination cases* would indeed cause an effect – even though (by counterfactualist lights) in overdetermination cases that heuristic leads us astray.[15] This heuristic may be strong because it is so useful. However, like all heuristics, it can be overridden in certain

[14] One drawback of the neuron diagrams popular in the causation literature since Lewis is that they abstract away from such details of presentation and so hide the importance of framing effects.

[15] This might be conjoined with other heuristics, based around factors such as those to be mentioned in the text shortly. For instance, in Meteorites the first rock may be picked out in preference to the second because only it had *physical contact* with the window, conjoined with the fact that such physical contact would indeed cause the window to shatter in normal non-overdetermination circumstances.

circumstances. In particular, in this case it can be overridden – and thus in overdetermination cases the correct causal judgment returned – when context makes especially salient the inevitability of the effect, which serves to redirect attention away from the specific pre-emptor event. Given the unlimited supply of different contexts, there should be an unlimited supply of such cases. (The details of what determines salience in these cases is precisely what is addressed by a theory of framing effects, on which more shortly.)

The cases in Table 1 are examples of this pattern at work: in all except Meteorites, context makes salient the effect's inevitability and thus that the pre-emptor does not make a difference to the effect and thus is not a cause of it. Further examples can be generated on demand.

In this section, I report some preliminary evidence about such framing effects. It joins other psychological work already being discussed in the philosophical literature. Most of the experiments in (Northcott 2011) concern pre-emption cases that, like those discussed earlier, are structurally identical with respect to counterfactual dependence, choice of contrasts, physical connection (except for Baseball Fielder), and nomological regularity. The experiments tested the impact on our causal judgment of four factors, in each case varying one factor while keeping the other three fixed:

(1) *Positive versus negative causation*. Did the scenario involve positive causation or prevention? (Baseball Fielder was the prevention case concerned.)

(2) *Physical contact*. Did the relation between the cause and effect events involve physical contact? (Meteorites was the physical contact case concerned.)

(3) *Teleological context*. Was the cause-effect relation embedded in a wider context that made the effect seem designed or intended?

(4) *Enable versus cause*. Did it matter if the cause 'enabled' the effect or 'caused' it? This line of inquiry was motivated by the ubiquity of this distinction in neuroscience. (Hotel Coupon was the 'enable' half of this comparison.)

The biggest impact was made by switching between positive and negative causation (almost 2.5 points on the Likert scale). There were smaller but (in many of the comparisons) still significant impacts for physical contact and for enable versus cause, while the smallest impact was for teleological context. A pre-emptor was more likely to be judged a cause if it was a case of, respectively: positive rather than negative causation; physical contact rather than no contact; cause rather than enable; a teleological rather than non-teleological context. The nascent theory of framing effects would be that the former halves of these pairs tend to create the illusion of causation, perhaps as suggested above by triggering archetypes that trump lack of counterfactual dependence.[16]

I have reported the experimental results here only very briefly. This is because I do not wish to imply that they are sufficient fully to establish a theory of framing effects, which would be a task far beyond any single study. Perhaps the most obvious shortfall is that the experiments so far merely probe what factors influence our causal judgment. That still leaves the task of constructing and testing a detailed positive theory that seeks to explain all of those judgments. The above is investigating merely what the ingredients of such a theory might be.

[16] The experimental results themselves are also consistent with it being the latter, rather than former, halves of the pairs that are misleading us. This second interpretation would imply that the illusions are our causal judgments in Baseball Fielder, Hotel Coupon, War, and Soccer rather than in Meteorites. But, as the paper as a whole argues, that interpretation would not cohere with any satisfactory wider metaphysical account – so, since the experimental results leave us free to reject it, we should do so.

Other interesting work about the cognitive science of causal judgments has also been discussed in the philosophical literature but none as yet quite bears on our pre-emption issue. Christopher Hitchcock and Joshua Knobe (2009) report on a number of experiments that show causal judgment to be influenced by various norms. Again, the method is to compare judgments in cases that are structurally identical and that differ only in the narrative framing. Three categories of norm have an impact on what factors are more likely to be picked out as causal: statistical, moral, and functional. But these experiments all concern the problem of causal selection, i.e. of when causal judgment picks out only one of several different necessary factors, regarding the rest as mere background conditions. Pre-emption cases, on the other hand, are, so to speak, problems the other way round, with multiple sufficient causes rather than multiple necessary ones, so that the effect is now counterfactually dependent on none, rather than several, of the factors in question. Thus, these norm studies do not yet speak directly to pre-emption. Similar remarks apply to several other studies, such as Halpern and Hitchcock (2015).[17]

John Collins (2000) and Michael McDermott (2002) each discusses what framing effects might be affecting causal judgment in the original Baseball Fielder case. Although they conduct no formal empirical investigations, to my knowledge these are the only such discussions of framing effects in regard to a pre-emption case. Christopher Clarke (unpublished manuscript) constructs a set of linked pre-emption cases and argues that they exhibit a judgment reversal, although again without a formal empirical investigation. Clarke shares this paper's wider dialectical goal of defending the counterfactual theory. There have been linguistic analyses of the assertibility conditions of causal claims (Swanson 2010), albeit

[17] The teleological context factor in the (Northcott 2011) experiments does incorporate a functional norm to some degree, although it did not prove to be especially influential.

to my knowledge none has so far been applied to pre-emption cases. Finally, contrastive analyses (footnote 6 above) also incorporate contextual considerations, albeit in a different way to this paper.

A whole field in psychology studies human causal cognition and learning, but to my knowledge none of it speaks directly to this paper's concerns. Most pertinent is work by Tania Lombrozo (2010) that studies some pre-emption cases. Although it does not consider judgment reversals, in our terms it can be thought of as a preliminary investigation of framing effects.

7. Probabilistic causation

Following most of the pre-emption literature, I have been concerned solely with deterministic cases, but a broadly analogous strategy is viable for probabilistic causation too.[18] Scientific practice again overwhelmingly endorses a difference-making definition, which now becomes (roughly speaking) that a cause is something that changes the probability of its effect. There are probabilistic versions of pre-emption cases, to which a similar analysis applies as for deterministic versions.

But there are in addition new problem cases specific to probabilistic difference-making. How these may be tackled provides independent support for this paper's general strategy, since it turns out that here too causation as difference-making may be offered psychologistic salvation. More precisely, the new problem cases are when probability-changers are not

[18] Twardy and Korb (2011) and Fenton-Glynn (2017) ingeniously extend to the probabilistic case the 'actual causation' approach derived from structural equations modeling. As discussed in section 5, that approach deviates from this paper's by seeking to accommodate rather than explain away our various causal judgments in the relevant cases.

judged causes and when causes are not judged probability-changers. There seems to be no possibility of the particular framing effects discussed above explaining why our causal judgments in these new cases are illusory and should be disregarded. Yet these new cases too are amenable to experimental elucidation – but this time by a different kind of framing effect. In particular, the relevant psychological theory now is that our causal judgment interprets the world under the assumption that the world is deterministic. Accordingly, in the eyes of our judgment, causes are only probabilistic for epistemic reasons, i.e. because of our ignorance of the complete causal situation. How exactly this offers salvation to the probabilistic version of the simple counterfactual account is an intricate story, detailed in (Northcott 2010).[19] There is space here only to summarize it briefly, in order just to indicate how it and this paper can be seen as two parts of a more general program.

The standard difference-making definition of probabilistic causation – and the one I ultimately want to defend – is that a cause is something that changes the probability of its effect, where the relevant probabilities are evaluated at the time of the cause-event. Call this the Standard Formula. But suppose that these probabilities are evaluated instead at the time of the *effect* event. It follows that the probability of the effect will then always be either 1 or 0, depending simply on whether the effect occurs or not. It turns out that this shift in the time at which the probabilities are evaluated can be interpreted as the natural result of a determinist attitude: roughly, 'really it was determined all along whether the effect has been caused or not; the uncertainty about this at the time of the cause-event just reflects our ignorance; the full truth is revealed *to us* only in hindsight, i.e. at the time of the effect; so that is the time at which causation may be judged.' This yields an adjusted formula for probabilistic causation:

[19] One loose end from (Northcott 2010) was that it offered no treatment of pre-emption cases. In this way, the present paper is complementary to it.

a cause is something that changes the probability of its effect, but now with this probability evaluated at the time of the effect rather than cause event. Call this the Adjusted Formula. The Adjusted Formula turns out to deliver the intuitively correct causal verdict in all of the key problem cases (and in several other cases of interest too). Why is this significant? Because the Adjusted Formula is just what you would expect to govern our causal judgment if, first, our causal judgment assumes determinism and, second, causation is really described by the Standard Formula, i.e. by probability-changing with the probabilities evaluated at the time of the cause. Thus, we now have a principled way to defend the Standard Formula definition of probabilistic causation from these particular counterexamples, namely that our judgments in these counterexamples are explicable illusions arising from a prejudice in favor of determinism. It is in this sense that (Northcott 2010)'s treatment of the probabilistic case is analogous to this paper's treatment of deterministic pre-emption cases: once again, a simple counterfactual account may be saved from its counterexamples by appeal to an independently supported psychological error theory.

8. Primitivism and pluralism

Primitivism typically takes causal powers to be ontologically basic.[20] Accordingly, although it may be used to endorse a close relation between causation and counterfactuals, it is not reductive. How might it accommodate judgment reversals? At first glance, it can do so easily – just by stipulation. Thus, for instance, we can simply assert that a causal power is exercised by the pre-emptor in Meteorites but not in Hotel Coupon. But this maneuver is ad hoc in the absence of any independent account of what constrains it. An explanation of the reversal is needed. After all, the operation of the hotel coupon's and meteorite rock's causal powers, in

[20] The view arguably dates at least to Aristotle. Notable recent advocates include (Harré and Madden 1975), (Cartwright 1989), (Heil 2003), and (Bird 2007).

terms of the effect they produce given the particular interactions with the other causal powers present, seem exactly analogous, yet still our causal judgment reverses. So, primitivism too is in some tension with that reversal.

On the other hand, one of the most powerful arguments for primitivism echoes that in this paper for the counterfactual theory – namely, its endorsement by scientific practice. Cartwright (1989), for instance, argues at length that science's methods of causal inference assume an ontology of, roughly speaking, stable causal powers. In reply: Cartwright does still also endorse the central role of difference-making. Her main target is analysis of causation in terms of Humean regularity. It is true that some versions of the counterfactual theory, most famously Lewis's, are Humean in this way. But others, such as Woodward's, are not committed to that. As with the general reductionism issue (section 1), this paper may remain neutral.[21] The point is that whether or not scientific practice supports an ontology of primitive causal powers, either way it still also endorses causation as difference-making.

Turn now briefly to pluralism. In various versions, this denies that there is any univocal definition of causation (Skyrms 1984, Hitchcock 2003, Hall 2004, Cartwright 2004, Reiss 2009, Godfrey-Smith 2010). Similarly to primitivism, pluralism may accommodate judgment reversals simply by stipulation. Thus, it could declare Meteorites to be a case of one kind of causation and Hotel Coupon a case of another. The challenge, again, would be to justify this otherwise ad hoc maneuver. After all, the examples feature structurally identical situations that differ only with respect to narrative window-dressing. Why should we think that they feature two metaphysically different kinds of causation?

[21] Of course, other theoretical options are available here besides powers and Humeanism, such as primitivism about subjunctive facts (Lange 2009) or a pragmatism-inspired quietism about the entire issue (Price 2011). This paper is neutral about these too.

9. A conceptual choice

Throughout, I have framed the discussion in favor of the counterfactual theory. The argument has been that, ontologically speaking, causation is counterfactual dependence but we aren't very good at applying it in some cases due to framing effects. As a result, in those cases our intuition gets the answers wrong. This view requires us to endorse our judgments in some pre-emption cases, such as Hotel Coupon, Baseball Fielder, War, and Soccer, while declaring our judgments in others, such as Meteorites, to be explicable mistakes or illusions.

What of causation as a concept? Should we refine this along similar counterfactualist lines? We face a choice. Either:

1) Causation is counterfactual dependence and therefore a reliable guide to the efficacy of interventions but unfortunately framing effects sometimes trigger ascriptions of it inappropriately. Thus, in the Meteorites case the rock is not a cause of the window shattering even though we intuitively judge that it is. Or:

2) Causation is some mixture of simple counterfactual dependence and other factors. Accordingly, it is now only a fallible guide to interventions. In Meteorites, the rock is indeed a cause of the window shattering.

Either way, the following will remain true: first, our intuitive causal judgment is only a fallible guide to the efficacy of interventions. Second, scientific practice overwhelmingly endorses a counterfactual, difference-making concept of causation. Third, there is an asymmetry between counterfactual dependence and other views because judgment reversals occur only in the problem cases of the counterfactual theory. That said, 1 or 2?

In favor of Option 2 is its fidelity to some of our intuitive judgments. In particular, perhaps Option 1's rejection of intuition in Meteorites and cases like it represents too radical a revision. We might define a notion of 'actual causation' that tracks our actual norm-infested causal judgments, perhaps usefully, and which can be distinguished from counterfactual dependence, the latter being the only thing relevant to assessing interventions (Halpern and Hitchcock 2015).[22]

It is not clear to me how best to weigh up the pros and cons of conceptual revision. That said, my own preference is for Option 1. It makes our concept of causation much simpler, and dovetails with a plausible psychological back-story. It keeps causation itself objective, restricting the subjective factors behind framing effects just to our judgments of it. Thus, the mishmash of complications is restricted to the psychological level, where it is plausible, and which seems preferable to introducing the mishmash to the metaphysical level. Option 1 also aligns with scientific practice, and it gives a clear rationale for why we should value causal knowledge. Finally, it supports many recent and fruitful applications of causation theory by philosophers of science, by giving them license to disregard the muddying of the conceptual water implied by pre-emption cases.

[22] Note that, as we saw in section 5, this approach so far gives no explanation of why our judgment – i.e. what it calls 'actual causation' – reverses between different pre-emption cases, sometimes coinciding with counterfactual dependence and other times not.

Bibliography

Angrist, Joshua, and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics* (Princeton: Princeton University Press).

Bird, Alexander (2007). *Nature's Metaphysics* (Oxford: Oxford University Press).

Blanchard, Thomas, and Jonathan Schaffer (2017). 'Cause without default', in Helen Beebee, Christopher Hitchcock and Huw Price (eds.), *Making a Difference* (Oxford: Oxford University Press), 175-214.

Cartwright, Nancy (1979). 'Causal laws and effective strategies', *Nous* 13, 419-437.

Cartwright, Nancy (1989). *Nature's Capacities and their Measurement* (New York: Oxford University Press).

Cartwright, Nancy (2004). 'Causation: One word, many things', *Philosophy of Science* 71, 805-819.

Cartwright, Nancy (2007). *Hunting Causes and Using Them* (Cambridge: Cambridge University Press).

Clarke, Christopher (unpublished manuscript), 'Causation is counterfactual dependence: Your intuitions are corrupt'.

Collins, John (2000). 'Preemptive prevention', *Journal of Philosophy* 97, 223-234.

Collins, John, Ned Hall, and L.A. Paul (eds) (2004). *Causation and Counterfactuals* (Cambridge MA: MIT Press).

Dowe, Phil (2000). *Physical Causation* (Cambridge University Press).

Fenton-Glynn, Luke (2017). 'A proposed probabilistic extension of the Halpern and Pearl definition of 'actual cause'', *British Journal for the Philosophy of Science* 68, 1061–1124.

Glymour, Clark (2004). Review of Woodward (2003), *British Journal for the Philosophy of Science* 55, 779-790.

Godfrey-Smith, Peter (2010). 'Causal pluralism', in Helen Beebee, Christopher Hitchcock, and Peter Menzies (eds.), *Oxford Handbook of Causation* (Oxford: Oxford University Press).

Hall, Ned (2000). 'Causation and the price of transitivity', *Journal of Philosophy* 97, 198-222.

Hall, Ned (2004). 'Two concepts of causation', in Collins et al, 225-276.

Hall, Ned (2007). 'Structural equations and causation', *Philosophical Studies* 132, 109-36.

Halpern, Joseph, and Judea Pearl (2005). 'Causes and explanations: A structural-model approach. Part I: Causes', *British Journal for the Philosophy of Science* 56, 843-87.

Halpern, Joseph, and Christopher Hitchcock (2015). 'Graded causation and defaults', *British Journal for the Philosophy of Science* 66, 413-57.

Harré, Rom, and Edward Madden (1975). *Causal Powers* (Oxford: Blackwell).

Heil, John (2003). *From an Ontological Point of View* (New York: Oxford University Press).

Hitchcock, Christopher (2001). 'The intransitivity of causation revealed in equations and graphs', *Journal of Philosophy* 98, 194–202.

Hitchcock, Christopher (2003). 'Of Humean bondage', *British Journal for the Philosophy of Science* 54, 1-25.

Hitchcock, Christopher (2006). 'Conceptual analysis naturalized: a metaphilosophical case study', *Journal of Philosophy* 103, 427-451.

Hitchcock, Christopher (2007). 'Prevention, pre-emption, and the principle of sufficient reason', *Philosophical Review* 116, 495-532.

Hitchcock, Christopher, and Joshua Knobe (2009). 'Cause and norm', *Journal of Philosophy* 106, 587-612.

Holland, Paul (1986). 'Statistics and causal inference', *Journal of the American Statistical Association* 81, 945–60.

Lange, Marc (2009). *Laws and Lawmakers: Science, Metaphysics and the Laws of Nature* (Oxford: Oxford University Press).

Lewis, David (1973). 'Causation', *Journal of Philosophy* 70 (1973), 556-567.

Lombrozo, Tania (2010). 'Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions', *Cognitive Psychology* 61, 303–332.

Maslen, Cei (2004). 'Causes, contrasts and the nontransitivity of causation', in Collins et al, 341-357.

McDermott, Michael (1995). 'Redundant causation', *British Journal for the Philosophy of Science* 46, 523-544.

McDermott, Michael (2002). 'Causation: influence versus sufficiency', *Journal of Philosophy* 97, 84-101.

McDonnell, Neil (2017). 'The non-occurrence of events', *Philosophy and Phenomenological Research* (online, doi:10.1111/phpr.12476).

Menzies, Peter (2007). 'Causation in context', in Huw Price and Richard Corry (eds.) *Causation, Physics, and the Constitution of Reality* (Oxford: Oxford University Press), 191-223.

Morgan, Stephen, and Christopher Winship (2007). *Counterfactuals and Causal Inference* (New York: Cambridge University Press).

Northcott, Robert (2008). 'Causation and contrast classes', *Philosophical Studies* 139, 111-123.

Northcott, Robert (2010). 'Natural-born determinists: a new defense of causation as probability-raising', *Philosophical Studies* 150, 1-20.

Northcott, Robert (2011). 'Pre-emption and causation experiments', unpublished manuscript (full results available on request).

Northcott, Robert (2015). 'Harm and causation', *Utilitas* 27, 147-164.

Paul, L.A. (2010). 'New roles for experimental work in metaphysics', *Review of Philosophy and Psychology* 1, 461-476.

Paul, L.A., and Ned Hall (2013). *Causation: A User's Guide* (Oxford University Press).

Pearl, Judea (2009). *Causality* (2nd edn, Cambridge University Press).

Pearl, Judea (2010). 'Nancy Cartwright on Hunting Causes', *Economics and Philosophy* 26, 69-77.

Price, Huw (2011). *Naturalism Without Mirrors* (New York: Oxford University Press).

Reiss, Julian (2009). 'Causation in the social sciences: evidence, inference, and purpose', *Philosophy of the Social Sciences* 39, 20-40.

Schaffer, Jonathan (2005). 'Contrastive causation', *Philosophical Review* 114, 297-328.

Sosa, Ernest (2007). 'Experimental philosophy and philosophical intuition', *Philosophical Studies* 132, 99-107.

Spirtes, Peter, Clark Glymour, and Richard Scheines (2000). *Causation, Prediction, and Search* (2nd edn, Cambridge, Massachusetts: MIT Press).

Skyrms, Brian (1984). 'EPR: Lessons for metaphysics', in Peter French, Theodore Uehling, and Howard Wettstein, *Midwest Studies in Philosophy IX* (Minneapolis: University of Minnesota Press 1984), 245-255.

Swanson, Eric (2010). 'Lessons from the context sensitivity of causal talk', *Journal of Philosophy* 107, 221-242.

Twardy, Charles, and Kevin Korb (2011). 'Actual causation by probabilistic active paths', *Philosophy of Science* 78, 900-13.

Weslake, Brad (forthcoming). 'A partial theory of actual causation', *British Journal for the Philosophy of Science*.

Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation* (Oxford: Oxford University Press).

Woodward, James (2006). 'Sensitive and insensitive causation', *Philosophical Review* 115, 1-50.

Acknowledgements
I thank Jonathan Livengood, Joshua Knobe and others for extensive help in running the relevant experiments. I thank two anonymous referees for helpful feedback. For earlier feedback I also thank anonymous referees from other journals, plus audiences at: University of California San Diego, Saint Louis University, University of Kansas, Society for Exact Philosophy, University of Missouri-Saint Louis, Birkbeck College, and University of Cambridge.