

---

---

# THE JOURNAL OF PHILOSOPHY

VOLUME XCVII, NO. 4, APRIL 2000

---

---

## TRUMPING PREEMPTION\*

**E**xtant *counterfactual accounts of causation* (CACs) still cannot handle preemptive causation. I describe a new variety of preemption, defend its possibility, and use it to show the inadequacy of extant CACs.

Imagine that it is a law of magic that the first spell cast on a given day match the enchantment that midnight. Suppose that at noon Merlin casts a spell (the first that day) to turn the prince into a frog, that at 6:00 pm Morgana casts a spell (the only other that day) to turn the prince into a frog, and that at midnight the prince becomes a frog. Clearly, Merlin's spell (the first that day) is a cause of the prince's becoming a frog and Morgana's is not, because the laws say that the first spells are the consequential ones. Nevertheless, there is no counterfactual dependence of the prince's becoming a frog on Merlin's spell, because Morgana's spell is a dependency-breaking backup. Further, there is neither a failure of intermediary events along the Morgana process (we may dramatize this by stipulating that spells work directly, without any intermediaries), nor any would-be difference in time or manner of the effect absent Merlin's spell, and thus nothing remains by which extant CACs might distinguish Merlin's spell from Morgana's in causal status.

### I. TRUMPING PREEMPTION

In order to establish the possibility and relevance of trumping scenarios such as the wizards case, I rebut the following imagined objections: that the causal judgments evoked are unclear, that the laws involved are question-begging, and that the case invoked is unrealistic, either in the sense that it is empirically implausible, given what we know about our world, or in the sense that it is pretheoretically

\* Thanks to John Carroll, Cian Dorr, Barry Loewer, Tim Maudlin, Brian McLaughlin, L. A. Paul, and especially to Ned Hall and David Lewis.

implausible, given what we habitually assume about our world. I conclude that delivering the right verdict in trumping cases is an adequacy condition on an account of causation.

*Objection:* it is not intuitively clear that Merlin's spell is a cause and Morgana's not of the prince's becoming a frog. The judgment of an otherwise successful theory (assuming CACs are otherwise successful) ought to overrule the judgment of unclear intuition. The wizards case, and trumping preemptions in general, should be left (to use David Armstrong's phrase) as spoils to the victor.

*Reply:* to my mind, the judgment that Merlin's spell is a cause and Morgana's not is intuitively clear (and so, to my mind, any account that cannot accommodate this judgment just is not an account of *causation*). But no matter. To the philosopher who does not share my intuitions, or does not attribute as much clarity to these intuitions, I reply that any account of causation which hopes to respect the central connotations of the causal concept had better agree that Merlin's spell is a cause and Morgana's not. To treat trumping preemptions as spoils to the victor is to risk spoiling the point of having a causal concept.

The concept of causation lies in a web of concepts including (most centrally) those of law, explanation, counterfactual implication, agency, and evidence. To a crudest approximation, (1) causes and their effects are subsumed under the laws; (2) causes explain their effects; (3) causes, were they different, would counterfactually imply different effects; (4) causes are means to their effects; (5) causes provide evidence both to and from their effects. Some or all of these relations may need to be refined or even ultimately abandoned. I submit, however, that any pair of events that satisfies *all* these relations deserves to be considered clearly causally related, while any pair of events that satisfies *none* deserves to be considered clearly not causally related.<sup>1</sup>

Consider the relation between causes and laws. Since the law is that the first spell cast on a given day match the enchantment that midnight, and since Merlin's spell is first that day, his spell satisfies the antecedent conditions of the magic law, and so calling his spell a cause of the enchantment that midnight respects the subsumption of causal structure by nomic structure. Morgana's spell, in contrast, does not satisfy the antecedent conditions of the magic law, and so her spell has no relevant lawful consequences. Thus,

<sup>1</sup> Compare D. H. Mellor, *The Facts of Causation* (New York: Routledge, 1995), pp. 58-66, who argues for the viability of indeterministic causation as fitting "the connotations of causation."

calling her spell a cause of the enchantment posits causation not covered by law.

Now, consider the relation between causes and explanations. Intuitively, if you want to know why the prince became a frog that midnight, Merlin's spell is part of a good answer. The deductive account of explanation vindicates this intuition:

- (1) At noon, Merlin cast a spell to turn the prince into a frog.
  - (2) No other spells were cast that day prior to or contemporaneous with Merlin's.
  - (3) It is a law that the first spell cast on a given day match the enchantment that midnight.
- 
- (4) The prince became a frog at midnight.<sup>2</sup>

Thus, treating Merlin's spell as a cause respects the connection between causes and explainers. Morgana's spell, in contrast, intuitively explains nothing about the prince's fate, and the deductive account vindicates this intuition:

- (1) At 6:00 pm, Morgana cast a spell to turn the prince into a frog.
  - (2) ...
  - (3) It is a law that the first spell cast on a given day match the enchantment that midnight.
- 
- (4) The prince became a frog at midnight.

There is no way to fill in (2) so as to get a deductively valid argument without rendering (1) superfluous, since the argument will be deductively valid (in a way that makes essential use of the laws in (3)) only when a first spell is specified, at which point Morgana's spell need not be.<sup>3</sup> Thus, calling Morgana's spell a cause invents causes that explain nothing.

<sup>2</sup> The main extant accounts of explanation are (a) deductive, (b) causal, and (c) pragmatic accounts. While (b) and (c) are not helpful here, because with regard to (b) the cause of the prince's metamorphosis is currently at issue, and with regard to (c) these accounts yield (by design) no objective fact, account (a) proves helpful. I take it that most will concede that, whether or not deductive accounts are ultimately viable, they at least provide reasonable codifications of our intuitions in most cases.

<sup>3</sup> There are tricks around this, such as by filling in (2) with some sort of disjunction, such as: either Morgana's spell was first that day, or it was not but what was first agreed with Morgana's in the enchantment it called for. But these are just the tricks which a deductive account of explanation must rule out as implausible anyway, and so in using the deductive account to justify our intuitions, we may assume (though we may not even be able to say how) such tricks are excluded.

Next, consider the relation between causes and counterfactual variance. Had Merlin's spell been different, such as a spell to turn the prince into a goat, then the prince's fate would have been different: he would have become a goat.<sup>4</sup> Thus, the prince's fate counterfactually varies with Merlin's magic, just as the barometer reading counterfactually varies with the atmospheric pressure, in that in both cases the noncompossible family of propositions representing differences in the former counterfactually depends on the noncompossible family of propositions representing differences in the latter—a relation that David Lewis<sup>5</sup> points out is typical of such intuitively causal processes as measurement, perception, and control. Not so for Morgana's spell: had she cast a spell to turn the prince into a goat, to protect his humanity, or even to make the sky the color of gold, come midnight all that would happen would be the prince's becoming a frog (because we are holding fixed Merlin's prior spell as well as the laws). To call her spell a cause would be to countenance causes whose alleged effects were counterfactually oblivious to them.

Finally, as to the relation between causes and agency, it is widely agreed that, "If an effect is an end, its causes are means to it."<sup>6</sup> This is usually spelled out decision theoretically: if  $E$  is a desired end and  $C$  a prospective means,  $C$  will be effective if and only if the probability of  $E$  given  $C$  is greater than the probability of  $E$  given  $\neg C$ . One way to spell this idea out in more detail is to use agent probabilities (on which actions are assumed free in having no causal antecedents), so that  $C$  will be effective if and only if  $p^A(E/C) >$

<sup>4</sup> This follows on such leading accounts of counterfactual implication as David Lewis's in *Counterfactuals* (Cambridge: Harvard, 1973). According to Lewis, we evaluate the implications of the counterfactual supposition that Merlin had cast a spell to turn the prince into a goat by postulating a minimal "divergence miracle" on which Merlin says 'Presto, prince to goat!' instead of 'Presto, prince to frog!', while holding fixed the course of history up to Merlin's conjuring (so holding fixed the absence of any prior or contemporaneous spells that day) as well as the laws (so presumably keeping the magic laws intact) to the extent compossible with the miracle; all this results in Merlin's casting the first spell that day to turn the prince into a goat, where the first spells cast match midnight enchantments, and so all this results in the prince's becoming a goat that midnight.

<sup>5</sup> "Causation," in his *Philosophical Papers*, Volume II (New York: Oxford, 1986), pp. 159-72, here p. 165. Since I shall show that Lewis's counterfactual account of causation fails to accommodate this variational type of counterfactual dependence in trumping cases, one way to read this result is that Lewis's account of causation as counterfactual dependence between event occurrences fails to capture his own intuitions about the general relation between counterfactual variation and causation.

<sup>6</sup> Mellor, "On Raising the Chances of Effects," in James H. Fetzer, ed., *Probability and Causality* (Boston: Reidel, 1988), pp. 229-40, here p. 230.

$p^A(E/\neg C)$ .<sup>7</sup> On this test, Merlin's spell is an effective means for ( $E$ ) the prince's becoming a frog, since  $p^A(E|\text{Merlin-casts-his-spell})=1$ , while  $p^A(E|\text{Merlin-does-not-cast-his-spell})<1$  (Morgana is assumed free, so she might not cast her spell). Morgana's spell, on the other hand, is ineffective, since  $p^A(E|\text{Morgana-does-not-cast-his-spell})=1$  as Merlin has already acted to determine the prince's fate. Thus, treating Merlin's spell as a cause and Morgana's as a noncause of the prince's becoming a frog respects the agential connotations of causes as means.

Finally, consideration of the relation between causes and evidence will converge on the same conclusion, namely, that Merlin's spell is a cause of the prince's metamorphosis and Morgana's not. Intuitively, causes provide (epistemically valuable) evidence both to and from their effects: if I see someone fire a bullet at Jones's heart, I am licensed to infer that Jones will die; and, conversely, if I see Jones supine with a bullet-shaped entry wound left of his sternum, I am licensed to infer that he was shot. There is typically a relation of scale in both directions: the more details I know about the cause, the more details I have license to predict concerning the effect; and the more details I know about the effect, the more details I have license to retrodict concerning the cause.

Suppose that all I know about Merlin's spell is that it is the first one cast that day, then all I have license to predict is that there will be some enchantment at midnight. If I learn more detail, such as that it targets the prince, then I can predict the resulting enchantment in more detail, namely, that it will concern the prince. And so on. This relation of scale also holds in reverse. If I know that the prince became a frog, I may infer that a spell of the form 'Presto, prince to frog!' was first that day. And if I then learn that Merlin's spell was first that day, I can come to know its intimate details, though I never witnessed it. Merlin's magic fits the prince's fate as the gun its bullet.

Morgana's spell, once again, fails the test of causes. If I know that hers is not the first that day, I can infer nothing about what will happen

<sup>7</sup> The main versions of decision theory are (a) evidential, and (b) causal (turning on whether "the probability of  $E$  given  $C$ " is interpreted as (a)  $p(EC)/p(C)$ , or (b)  $p(C \text{ will cause } E)$ ). While (b) is not helpful since whether  $C$  causes  $E$  is currently at issue, (a) proves helpful, and since (a) and (b) agree on non-Newcomb cases, (a) can be used neutrally in the case at hand. See William L. Harper and Brian Skyrms, eds., *Causation in Decision, Belief Change, and Statistics* (Boston: Kluwer, 1988). The point of introducing agent probabilities is to free evidentialism from the spurious correlations of Newcomb cases. For further discussion, see Peter Menzies and Huw Price, "Causation as a Secondary Quality," *British Journal for the Philosophy of Science*, XLIV (1993): 187-205.

at midnight from the details of her performance. If I do not know that hers is not the first that day, any inferences I make based on her performance will be, if true, merely accidentally so (and so not licensed).<sup>8</sup> In reverse, if I know what enchantment has befallen the prince that midnight, I have no license for inferring the details of later spells cast that day, and any inferences I make concerning Morgana's spell out of ignorance as to the fact that Merlin's came earlier will be at best accidentally true.

In summary, Merlin's spell is nomically antecedent for, explanatory of, counterfactually variant with, a means to, and evidence both to and from, the prince's becoming a frog at midnight, while Morgana's spell is none of these. So every individual measure of cause-worthiness converges on the result that Merlin's spell is a cause of the prince's becoming a frog, while Morgana's spell is not.

*Objection:* the arguments that Merlin's spell is a cause and Morgana's not of the prince's metamorphosis all turn on the stipulated law that the first spells cast match midnight enchantments. But such stipulation begs the question against those (such as Lewis, the leading proponent of CACs) who hold that laws supervene on occurrent facts. Why not stipulate that last spells cast match midnight enchantments, or that at least they do when first and last spells agree, since in any case the result is the same: the prince becomes a frog at midnight? If these stipulations represent genuine possibilities, then the supervenience of laws on facts is lost.

*Reply:* even if the wizards scenario were incompatible with the supervenience of laws on facts, this would still show (surprisingly) that the success of CACs is tied to the supervenience of laws. Those who would deny supervenience in the manner of Fred Dretske, Michael Tooley, and Armstrong<sup>9</sup> (who hold that laws are contingent necessi-

<sup>8</sup> In a sense, Morgana's spell is evidence for the prince's becoming a frog, since given certain (plausible) prior probability distributions, one may assign a greater subjective probability to the prince's becoming a frog given Morgana's spell. But this is not a sense of 'evidence' relevant here, since in this sense, given certain prior probability distributions, anything may count as evidence for anything (in particular the increased probability that the conditionalizer may assign prince-to-frog given Morgana's spell is exactly parallel to the increased probability assignable victim's death given backup shooter's presence; Morgana's spell carries exactly as much and as little evidence of the effect as any preempted process). The sense of 'evidence' relevant here, which typically holds between cause and effect, which holds between Merlin's spell and the prince's becoming a frog, and fails between Morgana's spell and the prince's becoming a frog, is evidence with *epistemic value*: that which may confer knowledge.

<sup>9</sup> Dretske, "Laws of Nature," *Philosophy of Science*, LXIV (1977): 248-68; Tooley, *Causation: A Realist Approach* (New York: Oxford, 1987); Armstrong, *What Is a Law of Nature?* (New York: Cambridge, 1983).

tation relations between universals) could not avail themselves of this objection, since on their view there may well be a contingent necessitation relation between the universals ‘first spell cast on day  $x$ ’ and ‘midnight enchantment on day  $x$ ’; those who would deny supervenience in the manner of John Carroll and Tim Maudlin<sup>10</sup> (who hold that laws are primitive) could not avail themselves of this objection either, since on their view there is no barrier at all to the existence of a primitive lawful relation between the first spells and midnight enchantments. But no matter. To the philosopher who is willing to tie the success of CACs to the supervenience of laws on facts, I reply that the existence of laws favoring the first spells even when first and later spells agree (as per the wizards) is fully compatible with the supervenience of laws on facts, and is actually entailed by leading supervenientist accounts of laws, given certain facts in the world’s history.

Suppose laws supervene on facts, and suppose for the sake of definiteness that (1) laws supervene as per the Mill-Ramsey-Lewis<sup>11</sup> (MRL) account:  $L$  is a law of nature if and only if  $L$  is a theorem of the axiomatization of the facts that best balances simplicity and strength. Now (2) let the world history contain “decisive competitions” between spells, where first and later spells disagree, and where the first spell cast always wins (that is, you try first to turn the queen into a frog, I try later to turn her into a goat, and she becomes a frog). Finally (3), let the laws summarizing the nonmagical remainder of the world history be independent of the magical laws (say, the world is otherwise Newtonian). This suffices to fix the laws as per the wizards case: when first and later spells agree, it will be simpler (and just as strong) to regard these cases as extensions of the same overall pattern of first-spell dominance found in (2), rather than as new examples of a second pattern of later-spell dominance or first-and-later-spell equality; this gain in simplicity will not compromise the simplicity/strength of the other laws since they are assumed independent by (3), and so this local and not-compromised-elsewhere gain in simplicity without loss in strength will by (1) entail laws favoring the first spells even when first and later spells agree.

Somewhat more precisely, consider how the laws of a world in which the first spells dominate when first and later spells disagree might cover cases in which first and later spells agree:

<sup>10</sup> Carroll, *Laws of Nature* (New York: Cambridge, 1994); and Maudlin, “A Modest Proposal” (unpublished manuscript).

<sup>11</sup> Lewis, *Counterfactuals*; also John Earman, “Laws of Nature: The Empiricist Challenge,” in R. J. Bogdan, ed., *D. M. Armstrong* (Boston: Reidel, 1984), pp. 191-223.

- (a) *Asymmetric* in allowing only the first spells to satisfy their antecedents (as per the wizards): for all spells  $x$  and days  $y$ , if  $x$  is first on  $y$ , then the  $x$ -matching enchantment obtains that midnight.
- (b) *Symmetric* in allowing both first and later spells to satisfy their antecedents: for all spells  $x$  and days  $y$ , if  $x$  is first on  $y$  or there is a spell  $z$  which is first on  $y$  and which agrees with  $x$ , then the  $x$ -matching enchantment obtains that midnight.
- (c) *Asymmetric* in allowing only later spells to satisfy their antecedents: for all spells  $x$  and days  $y$ , if ( $x$  is first on  $y$  and there is no spell  $z$  which is later on  $y$  and which agrees with  $x$ ) or (there is a spell  $z$  which is first on  $y$  and which agrees with  $x$  and there is no  $z'$  which is both later than  $x$  and which agrees with  $x$ ), then the  $x$ -matching enchantment obtains that midnight.
- (d) *Silent* in these cases: for all spells  $x$  and days  $y$ , if  $x$  is first on  $y$  and there is no spell  $z$  which is later on  $y$  and which agrees with  $x$ , then the  $x$ -matching enchantment obtains that midnight.

Now, it is apparent that (a) is simpler than (b), (c), or (d): (a) subsumes all cases under one pattern/clause, that of first-spell dominance, while (b)-(d) require multiple clauses to accommodate multiple patterns;<sup>12</sup> it is especially revealing that (d), which is silent on first-later agreements, is actually more complicated than (a)—to get the laws not to say anything about first-spell dominance, you need to add a silencing clause. Further, it is obvious that (a) is as strong as (b) or (c) and stronger than (d): (a)-(c) have identical consequences as to which events are entailed, and how informatively described each event is; whereas (d) is as strong but for cases of first-later agreement under which no events are entailed. So, on the MRL account, given that laws are the theorems of the maximally simple and strong deductive system (I have been supposing that the laws of magic are independent of the other laws), it follows that the laws of magic say that the first spells entail matching enchantments.

Generalizing away from the MRL account, the above result shows that any account of laws which rules simplest laws at least possible (which I consider an adequacy condition on an account of lawhood, given the role of simplicity in scientific practice) must deliver at least some worlds with laws as per the wizards. So even someone who

<sup>12</sup> We could, of course, be more precise, or employ different vocabulary, but no added precision or shift of (natural) vocabulary should alter the greater simplicity of (a) as compared to (b)-(d). Of course, sufficiently "gruesome" predicates would alter this verdict, but the MRL account of lawhood, which I am presently assuming, is viable only to the extent that our similarity judgments, based in part on our vocabulary, have objective bite; so for purposes of showing that the wizards laws beg no questions against supervenientists like Lewis, we may assume what the MRL advocate must.

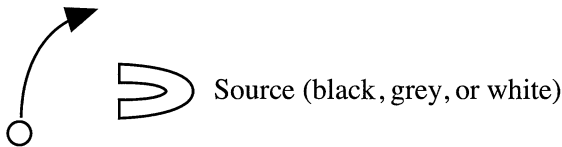


thinks laws are contingent necessitation relations between universals in the manner of Dretske, Tooley, and Armstrong, or who thinks laws are primitive in the manner of Carroll and Maudlin, should accept that there are at least some worlds with laws as per the wizards (they will presumably add that there are other worlds, factually identical to the wizards with laws as in (b), (c), and (d), and maybe with no laws at all). Thus, the laws invoked should be unproblematic on reflection.<sup>13</sup>

*Objection:* causation is an empirical concept, and an account that mishandles fairy-tale cases like the wizards will be no less empirically adequate. Even if the wizards case is a conceptual possibility, it involves magic and action-at-a-distance, and so is too far-fetched to worry about.

*Reply:* I think accounts of causation need to be not just empirically but conceptually adequate. Without prior conceptual analysis, how could we tell whether an empirical investigation had hit its mark? But no matter. To the philosopher who scorns such “merely conceptual possibilities” as those involving magic and action-at-a-distance and demands empirically plausible scenarios, I reply that there exist empirically plausible trumping scenarios.

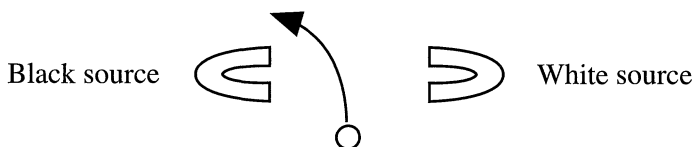
Imagine a world whose laws of physics are very much like our own, but for recognizing types of fields, ‘black’, ‘grey’, and ‘white’, which do not superpose. Whenever a particle is subject to just one of these fields (assume for simplicity that no further forces are present), the particle will accelerate along a curved trajectory:



Whenever a particle is subject to multiple fields, the particle will ac-

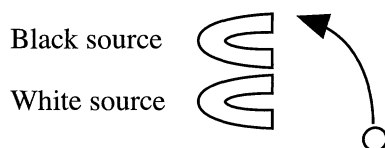
<sup>13</sup> Perhaps some will balk at the laws invoked, not because of worries about supervenience, but because of disbelief in the possibility of laws projecting extrinsic properties, such as being first cast on a given day. I see no grounds for such disbelief (moreover none of the leading accounts of lawhood provides grounds for such disbelief). Further, I take worlds such as the wizards world, in which there is perfect covariance between events with the extrinsic property of being-the-first-spell-cast-on-day-*x* and events of midnight enchantments, and in which there may well be no such covariance between events merely-intrinsically-described and midnight enchantments, to show exactly why such disbelief is unwarranted.

celerate as if subject only to the darkest field, for example:<sup>14</sup>



Thus, the laws of physics at this world would include a color-field law, most simply formulated as: the intensity of the darkest fields on a given object = the color force on that object. Although this world does not have the physics we now believe our world to have, it is relatively easy to imagine actual physicists discovering further types of fields, performing the experiments represented above, and supplementing our laws of physics in accord. Moreover, this world is free of magic, action-at-a-distance, and other types of “far-fetched” causal connections. It is in this sense that it is empirically plausible.

Now, if a particle is subject to both a black and a white field that “pull in the same direction” with the same magnitude, it is still only the black field that causes the resultant acceleration:



Here, we have an empirically plausible trumping scenario: replace Merlin with the black source, Morgana with the white source, the prince’s fate with the particle’s acceleration, spells with fields, and laws favoring earlier spells with laws favoring darker fields—the arguments from cause-worthiness all carry over.

The fields case shows that all that is needed for trumping are three events, C1, C2, and *E*, where *E* would be produced by C1 alone or by C2 alone (redundant causation), but where the laws render C1

<sup>14</sup> This does not mean that white fields in general are always trumped: there may be times when no black fields exist, or the black fields might have finite spatial extent, or the black fields might trump only insofar as they exceed a certain threshold. Nor does it mean that this particular white field is completely epiphenomenal: it might trump on other types of particles, or trigger a white flag in the distance, or have any number of other effects as long as they are not manifest on this particle. More interestingly, there might be a functional law on which the ratio of darker field to lighter field influences cycles through the epochs, with the relevant experiments done in a 1:0 epoch. This raises the empirically open possibility that our world is such a world (with respect to, say, types of electromagnetic force): our epoch just happens to be 1:1, so we do not notice.

the exclusive factor in the entailment of *E* (trumping preemption). This is why the causal exotica of the wizards case is inessential. Nothing empirically implausible remains.

*Objection:* causation is a folk concept, and an account of causation which mishandles fairy-tale cases like the wizards and faux-physical cases like the fields will be no less useful in our everyday dealings with people, rocks, and the like. Even if the wizards case is a conceptual possibility and the fields case an empirical possibility, both involve behaviors so unlike those typical of local middle-sized dry objects that both cases are too removed from our habitual assumptions about the world for us to expect our concept of causation to be adequate to them.

*Reply:* I think the objector is overly pessimistic about the viability of our concepts in unfamiliar scenarios. The intuitive clarity of a judgment in a given case strikes me as strong evidence for the viability of the judged concept in that case (I hope to have already convinced the reader of the intuitive clarity of causal judgments in trumping cases). But no matter. To the philosopher skeptical of judgments in unfamiliar scenarios who requires scenarios that accord with the world we habitually take ourselves to inhabit, I reply that there exist pretheoretically plausible trumping scenarios.

Imagine that in a world that could well be our own, the major and the sergeant stand before the corporal, both shout 'Charge!' at the same time, and the corporal decides to charge.<sup>15</sup> Orders from higher-ranking soldiers trump those of lower rank. I hope you agree that the major's order, and not the sergeant's, causes the corporal's decision to charge (for the same reasons as in the wizards and fields cases: ranking orders are like the first spells and darkest fields). The soldiers case shows that everyday scenarios can stand in the pattern constitutive of trumping preemption. Nothing pretheoretically unfamiliar is needed.

The sophisticated philosopher might presume that the soldiers (unlike the wizards and the fields) must be a case of standard preemption, by presuming that, as the major's and sergeant's orders engage the corporal's mental mechanisms en route to his decision module, there must be some (at least unconscious) intermediate filter through which the major's order is passed but the sergeant's order blocked. But those who would reject the wizards and fields cases on the grounds

<sup>15</sup> This particular case is a variant of one due to Bas van Fraassen, though the discovery of this type of case is Ned Hall's. A variant of one of Hall's cases illustrates just how common these cases are: the Bradys are deciding where to vacation, and Cindy and Bobby both plead "Hawaii!" Since Ma and Pa always do what Cindy wants, the Bradys decide on Hawaii. Trumping is everywhere.

that causation is a folk concept are not entitled to such sophisticated presumptions. Perhaps the corporal decides to charge via some such unconscious intermediate filter (as in standard preemption), or perhaps he decides to charge because his training has simply made his decision module exclusively sensitive to ranking orders (as in trumping preemption). The folk do not know how the mind is wired here, and from the point of view of attributing causation, they do not care.

In conclusion, trumping preemptions are intuitively clear, unproblematic in their assumptions, and fully realistic in being both empirically and pretheoretically plausible. Pending further objections, I take the above to show that delivering the right verdict in trumping cases is an adequacy condition on an account of causation.

## II. COUNTERFACTUAL ACCOUNTS OF CAUSATION

CACs, developed especially by Lewis, in their simplest form claim that: *C* causes *E* if and only if (i) *C* and *E* are actual, distinct events, and (ii) if *C* had not occurred, *E* would not have occurred.

Preemptions show that simple counterfactual dependence is not necessary for causation, because, in evaluating the counterfactual implications of *C*'s not occurring, we hold the presence of the backup fixed, and so still get *E*. In response, defenders of CACs have refined the counterfactual-dependence relation. The leading extant refinements (Lewis's ancestral dependence, fragility, and quasi-dependence; Peter Menzies's continuous processes; Michael McDermott's minimal-counterfactual sufficiency; Murali Ramachandran's minimal-dependence sets<sup>16</sup>) all rely on failed intermediaries along the backup process or on would-be differences in the effect absent the main process to distinguish preempting causes from preempted backups. For this reason, none has the resources to distinguish trumping causes from trumped backups.

Ancestral dependence, quasi-dependence, continuous processes, and minimal-dependence sets all rely on failed intermediaries along the backup process to distinguish preempting causes from preempted backups. According to ancestral dependence (suggested by Lewis to transitivize simple counterfactual dependence and handle preemption cases in one swoop), *C* may cause *E* even though *E* does not counterfactually depend on *C*, as long as there is an event *D* (or some finite set of *D*s) such that *E* counterfactually depends on *D* and

<sup>16</sup> These accounts are offered, respectively, in Lewis, "Causation" (*op. cit.*) and "Postscripts A-F to 'Causation'," in *Philosophical Papers*, Volume II, pp. 172-213; Menzies, "Probabilistic Causation and Causal Processes: A Critique of Lewis," *Philosophy of Science*, LVI (1989): 642-63; McDermott, "Redundant Causation," *British Journal for the Philosophy of Science*, XL (1995): 523-44; Ramachandran, "A Counterfactual Analysis of Causation," *Mind*, CVI (1997): 263-77.

*D* on *C*. Thus, where there are two gunmen *C* and *C'* but only *C* shoots, the death *E* does not counterfactually depend on either *C* or *C'*, but does presumably depend on the bullet-flying *D*, which in turn depends on *C*. Ancestral dependence works only if there are failed intermediaries on the backup process prior to the effect (early), since these failures are what create counterfactual dependence of the effect on some post-failure intermediary *D* along the main process, by spoiling the dependency-breaking redundancy. For this reason, ancestral dependence miscounts the trumping cause as a noncause. There is no intermediary along the trumping process such that the effect depends on it, either because the causation is direct (no intermediaries), as in the wizards, or because the trumped process is fully intact, as with the white field that propagates all the way to the particle, and with the sergeant's order that remains with the corporal throughout the decision process, breaking the dependency of the effect on all intermediaries along the entire trumping process.

According to quasi-dependence (tentatively adopted by Lewis at the end of postscript E to "Causation"), *C* may cause *E* even though *E* does not counterfactually depend on *C*, as long as *C* and *E* are part of a process that is intrinsically such as typically to induce counterfactual dependence (this relation of counterfactual or quasi-dependence is then transitivized by taking the ancestral). Thus, where there are two gunmen *C* and *C'* but only *C* shoots, the *C* to *E* process will still be intrinsically such as typically to induce counterfactual dependence, because the presence of a backup gunman is presumably atypical. What about the *C'* to *E* process? The only reason I can see why this process (understood, for example, as the sequence of events <*C'* armed and malevolent, *E* dead>) is intrinsically not such as typically to induce counterfactual dependence is that there are failed requisite intermediaries along this process, such as a failure of *C'* to pull the trigger.<sup>17</sup> For this reason, quasi-dependence miscounts the trumped backup as a cause. If Merlin's spell is to count as a cause because the great majority of intrinsic duplicates of the sequence <Merlin casts prince to frog, prince turns to frog> exhibit counterfactual dependency, then Morgana's spell must count as a cause as well, since the great majority of intrinsic duplicates of <Morgana casts prince to frog, prince turns to frog> will exhibit just as much dependency, since in both cases we abstract away from the atypical presence of a rival spell caster; likewise for the field-process and sol-

<sup>17</sup> Jonardon Ganeri, Paul Noordhof, and Ramachandran come to the same understanding of quasi-dependence in "Counterfactuals and Preemptive Causation," *Analysis*, LVI (1996): 219-25, here p. 220.

dier-process duplicates. Trumping cases show that the intuition that motivates quasi-dependence—namely, that causation is determined by the intrinsic character of events (plus the laws)—is hasty, since the laws may well be sensitive to extrinsic characteristics, such as which spell is earlier.

The continuous-process refinement (suggested by Menzies as a solution to “making it the hard way” preemption cases) requires that *C* and *E* be connected by a temporally continuous chain of counterfactual dependencies. Here, failed intermediaries function as discontinuities in the dependency chains from the backup. Without these discontinuities, the backup will spoil the final dependency of effect on the immediately prior event along the causal process. Thus, because the trumped backup runs continuously in the fields case, the final dependency of the particle’s acceleration on the black source is spoiled by the continued propagation of the white field, and so the continuity requirement miscounts the trumping cause as a noncause. Worse, requiring temporal continuity rules out by brute stipulation the intuitive possibility of action-at-a-distance and so could not count Merlin’s spell as a cause even if Morgana had not cast a spell at all.<sup>18</sup>

According to minimal-dependence sets (devised by Ramachandran), preempter *C* counts as a cause of *E* because the *C*-process involves only actual events, while backup *C'* is not a cause because the *C'*-process includes failed events. First, Ramachandran defines an *M*-set for *E* as a nonempty set of events *S* such that:

- (i) *E* is not a member of *S*, (ii) if no members of *S* had occurred, *E* would not have occurred, and (iii) no proper subset of *S* meets (ii).

He then shows that *M*-sets will contain exactly one event, actual or not, from each process, preempted or not, leading to *E*, and concludes that *C* causes *E* if and only if:

- (iv) *C* and *E* are actual events, (v) *C* is in an *M*-set for *E*, and (vi) there

<sup>18</sup> Menzies has recently moved to a theoretical definition of causation on which causation is that intrinsic relation between actual, distinct events which typically induces counterfactual dependence between them—“Probabilistic Causation and the Pre-emption Problem,” *Mind*, cv (1996): 85-116. This definition is still troubled by trumping, since it takes as platitudinous that causation is an intrinsic-to-its-pairs relation between events. Take different-day duplicates of the pair <Merlin’s spell, prince’s metamorphosis>. Whether a given duplicate is causal will depend, *inter alia*, on whether there was a Morgana spell at 9:00 am. Thus, trumping refutes Menzies’s central assertion that, “The distinctive mark of our intuitive concept of causation...is that it takes causal relations to be determined by the natural properties of the relata and the natural relations holding between them, taken in isolation from everything else happening in the world” (*ibid.*, p. 100).

are no two  $M$ -sets for  $E$  which differ only in having one or more nonactual events in place of  $C$ .

The failed-intermediaries idea is implemented in (vi), since given that  $M$ -sets contain exactly one event from each process,  $C$  will be replaceable by a nonactual event (and thus not a cause) only if there is a nonactual event on its process. Ramachandran is fully explicit in his reliance on the failed-intermediaries idea:

It seems true in all genuine cases of causal pre-emption,...that the pre-empted processes do not run their full course,.... For any pre-empted cause  $x$ , of an event,  $y$ , there will be at least one possible event...which fails to occur in the actual circumstances *but which would have to occur in order for  $x$  to be a genuine cause of  $y$ ....* All genuine causes, on the other hand *do* seem to run their full course; indeed, they presumably count as genuine precisely because they do so (*op. cit.*, p. 273)

Thus, minimal-dependence sets miscounts the trumped backup as a cause, because it runs its full course. In the wizards case, {Merlin's spell, Morgana's spell} will be an  $M$ -set for the prince's metamorphosis, but there will be no nonactual event substitutable for Morgana's spell, since there is no nonactual event along the process from Morgana's spell to the effect; likewise for the fields and soldiers cases.

Thus, ancestral dependence, quasi-dependence, continuous processes, and minimal-dependence sets cannot distinguish trumper from trumped. It is important to see that this failure is due, not to the mere details of each refinement, but to the underlying assumption they share (call this the *cutting assumption*) that backup processes are such because they have failed intermediaries. Trumping preemptions show that the root idea of these refinements is inadequate.

Both fragility and minimal-counterfactual sufficiency add the strictest standards of event individuation to the failed-intermediaries idea. According to fragility (proposed though tentatively rejected by Lewis on grounds of spurious causation), the above ancestral-dependency analysis is to be implemented on the understanding that events have their times and manners of occurrence essentially.<sup>19</sup> Fragility-based approaches are intended to handle *late* preemption cases in which the failed intermedi-

<sup>19</sup> We may distinguish temporal fragility from fragility of manner, and fragility of effects from fragility of events generally. All Lewis needs (for his reply to *late* preemption) is the temporal fragility of effects. In fact, all that is needed is the stipulation that the effect not be delayed, as L. A. Paul has shown in, "Keeping Track of the Time," *Analysis*, LVIII (1999): 191-98. Nothing here will turn on the details of how the fragility proposal is implemented.

aries on the backup process occur at or after the time of the effect (since it then follows, assuming no simultaneous/backward causation, that had the backup been left to produce the effect, the effect would have been delayed) and are no help at all with trumping. In the wizards case, for instance, the prince's metamorphosis, absent Merlin's spell, would still take place at exactly midnight, and in precisely the same manner; the analogous point applies to the fields and soldiers cases. Thus, the fragility refinement miscounts the trumping cause as a noncause.

According to minimal-counterfactual sufficiency (due to McDermott), the root causal notion is of counterfactual sufficiency between strictly individuated events, where a set of actual events  $C$  is counterfactually sufficient for  $E$  if and only if:

There is no set of actual events  $D$ , where  $E$  is not a member of  $D$ , such that (if the members of  $C$  had all occurred and the members of  $D$  had all not occurred, then  $E$  would not have occurred.)<sup>20</sup>

Now,  $C$  is minimally counterfactually sufficient for  $E$  if and only if:

- (i)  $C$  is counterfactually sufficient for  $E$ , (ii) no proper subset of  $C$  is counterfactually sufficient for  $E$ , and (iii) there is no actual event  $D$  (distinct from  $E$  and the members of  $C$ ) such that for some proper subset  $C'$  of  $C$ ,  $C' \cup \{D\}$  is counterfactually sufficient for  $E$ , and for some proper subset  $C''$  of  $C$ ,  $C'' \cup \{\neg D\}$  is counterfactually sufficient for  $E$ .

McDermott equates direct causation with minimal-counterfactual sufficiency, and claims that, "a part of a minimal sufficient condition is always a cause" (*op. cit.*, p. 535). He denies the converse, and so supplements his account by defining processes via chains of minimally counterfactually sufficient conditions, relativizing causes to processes, and distinguishing between realizing the real as opposed to nominal essence of an event. I need not spell out these (considerable) complications here, since the trumped backup disproves the claim that a part of a minimally counterfactually sufficient condition is always a cause, and since these complications merely countenance more causes. Consider the set {Morgana's spell}. It is a sufficient condition, according to the strictly-individuated counterfactual-based definition, since its occurrence by itself would still lead to the prince's becoming a frog exactly as and when he actually does. It is

<sup>20</sup> This is a questionable use of counterfactuals. It is difficult enough to say what would happen on the counterfactual supposition that one event not occur, and almost completely indeterminate what would happen had most of the rest of the world other than  $C$  and  $E$  not occurred: space-time might simply collapse. And the strict individuation of events which McDermott adds only makes things worse, because it tightens the standards for saying that  $E$  still occurs.



clearly minimal, since (i) it is sufficient, (ii) its only proper subset,  $\{ \}$ , is not sufficient for the prince's becoming a frog, and (iii) there is no actual event  $D$  such that both  $\{ \} \cup \{D\}$  as well as  $\{ \} \cup \{\neg D\}$  are sufficient for the prince's becoming a frog (in particular Merlin's spell fails the latter). But Morgana's spell, despite being part of a minimally counterfactually sufficient condition, is not a cause of the prince's becoming a frog.<sup>21</sup>

Thus, fragility and minimal counterfactual sufficiency cannot distinguish trumper from trumped. This failure is due to the underlying assumption they share (call this the *precision assumption*), namely, that backup processes are such that had they run to completion, the effect would have been somewhat different in time or manner. Trumping preemptions show that the root idea of these refinements is inadequate.

In conclusion, extant CACs still cannot handle preemptive causation. Trumping preemptions show that the failed intermediary and would-be differences strategies that extant CACs use are inadequate as general solutions to the preemption problem.<sup>22</sup> Of course, there may yet be some new strategy for CACs which will prove adequate, but at this point the prospects look dark.

JONATHAN SCHAFER

University of Houston

<sup>21</sup> Adding the rest of McDermott's machinery: the set  $P$  of all events involved in the process is {Morgana's spell, the prince becoming a frog}, the set  $P_i$  of intermediate events is  $\{ \}$ , and the set  $P_c$  of primary causal factors is {Morgana's spell}. Now, it is clear that Morgana's spell is counterfactually sufficient relative to  $P$ , and minimally so, and so vis-à-vis realizing the nominal essence of  $E$ , which is just the prince's becoming a frog. The added machinery will not help with the fields or soldiers, either. In the fields case (the analogous points will hold for the soldiers), the process is something like {the insertion of the white source, the propagation of the white field, the white field hitting the particle, the particle accelerating along a curved trajectory}. The set  $P_c$  is {the insertion of the white source}. Now, all the intermediaries in the process are held fixed, and so the insertion of the white source now counts as counterfactually sufficient relative to  $P$ , and minimally so, and so vis-à-vis the nominalization "the particle's accelerating along a curved trajectory," and thus the presence of the white source is miscounted as a cause.

<sup>22</sup> One solution to trumping is to analyze causation via nomic subsumption rather than counterfactual dependence, since Merlin's spell is subsumed under the magic law while Morgana's is not (likewise for the fields, and likewise presumably for the soldiers given that the corporal's decision module works as described). But this is a whole other story.