For What Must We Pay?
Causation and Counterfactual Baselines

MICHAEL MOORE*

TABLE OF CONTENTS

1182

I. INTRODUCTION

A. Negative Moral Obligations

It is often thought that the hard core of our moral obligations—the social minimum, so to speak—is that we do no harm. Regardless of what may be virtuous, supererogatory, or obligatory in making the world a better place, at a minimum we should not make it a worse one. Further, we make the world a worse place whenever we worsen it compared to the way the world would have been without our actions. On this view, our actions must make a difference for us to violate our basic obligations. What gets recorded in our moral ledgers are the bad states of affairs that would not have existed but for our actions. Our actions were, in a word, necessary to the bad states of affairs for which we are responsible.

This view of our responsibility is naturally captured by a certain kind of counterfactual test, one that compares how the world is after our actions with how the world would have been if, contrary to fact, we had not done the actions in question. Most lawyers, and not a few philosophers, believe that this counterfactual test is also a test of causation. This, because on the counterfactual theory of causation, to cause something is to make the kind of difference tested by the contrary-to-fact conditional stated above. According to the counterfactual theory of causation, we cause harm when our actions are a necessary condition for that harm having occurred; if it is true that but for our acts, some harm would not have existed, then, and only then, have we caused it to come about. Because we are, on this view, responsible for what we cause, we are responsible for harms that do exist but which, absent our actions, would not have existed.

B. Legal Reflections of Our Moral Obligations

The legal applications of this view of the hard core of our moral obligations are legion. The following subsections will consider ten of them.

1. The Basic Case of Misfeasance in Tort

Despite the adoption of the "substantial factor" test of causation in fact

1183

by the *Restatement of Torts*,[1] tort liability in general is standardly thought to require successful application of the "but for" test of causation in fact.[2] If a plaintiff seeks to recover damages for cancer caused by the defendant's cigarettes, cancer caused by a cut from the defendant's negligently maintained transom,[3] or cancer caused by any other action or state of affairs for which a defendant is responsible, the plaintiff has to show that absent the smoking, cutting, or other acts, there would have been no cancer. This is usually termed the "but for," necessary condition, or "sine qua non" (literally, without which, not) test of causation and thus, for liability, in tort.

This but-for test is no peripheral part of tort law. On the contrary, it is thought to be the central core of the causation requirement, which is itself a core requirement of assigning tort liability. The but-for test occupies such a position of prominence because it is thought to mark the line between mere temporal succession and true causation.[4] For instance, just because this cancer followed upon this cut does not mean the cut caused the cancer; just because this kind of cancer usually or even always follows upon this kind of cut does not mean that cuts cause cancer.[5] Indeed, "post hoc, ergo propter hoc" ("after this, therefore, because of this") is the fallacy thought to be exposed by the but-for test for causation. The difference between mere temporal succession of events and true causal relations between events is said to lie in the counterfactual question: But for the first event, would the second have occurred?

2. *The Basic Case of Misfeasance in Contract*

The difference between tort and contract liability lies in the different bases of the primary obligation breached by a defendant. In tort, that obligation is nonpromise-based, whereas in contract, the obligation

---

1. The substantial factor test was first proposed by Jeremiah Smith, *Legal Cause in Actions of Tort* (pts. 1–3), 25 HARV. L. REV. 103, 103–28, 223, 223–52, 303, 303–21 (1911–1912). It was adopted by both the first and the second *Restatements*. RESTATEMENT OF TORTS §§ 431–35 (1934); RESTATEMENT (SECOND) OF TORTS §§ 431–33 (1965).

2. The *Restatements* hold that, except for the overdetermination cases, a but-for cause of a harm is a substantial factor and, thus, a cause in fact of that harm. RESTATEMENT OF TORTS §§ 431–33 (1934).

3. *See* Kramer Serv. Inc. v. Wilkins, 186 So. 625 (Miss. 1939) (discussing injury caused by a negligently maintained transom).

4. RESTATEMENT (SECOND) OF TORTS § 431 cmt. a (1965). This section assumes that the factual or scientific part of the legal tests of causation treats as a cause "every one of the great number of events without which any happening would not have occurred." *Id*.

5. The correlation-cause distinction is described briefly in Moore, *Causation and Responsibility*, *supra* note *, at 7–8.

arises out of the defendant's promise. When the promise in question is negative in character ("negative" meaning the promisor promises *not* to do some act *A*), breach of the promise in question is an action by the defendant, namely, the one he promised not to do, *A*. When the promise in question is positive in character ("positive" meaning the promisor promised to *do* some act *A*), breach of the promise is an omission by the defendant, namely, an omission to perform act *A*.

Increasingly, contract law has eliminated any distinction between misfeasance and nonfeasance by treating all cases as instances of misfeasance. When promisees seek reliance damages, they treat the promisor's promise as the touchstone of liability, not the act or omission in breach. The causal chain is then from promise by defendant, through change of position by plaintiff in reliance on the promise, and then harm caused plaintiff by his change of position. Such reliance damage cases in contracts thus look very much like standard misfeasance cases in tort. Indeed, the causal analysis is the same. The defendant's promise must cause the plaintiff's harm, and this is standardly translated into the counterfactual question: But for the defendant's promise, would the plaintiff have suffered the harm?

Despite the increasing number of reliance damage cases,[6] the standard measure of damages in contract law is the plaintiff's expectancy.[7] Here we ask not whether the defendant's promise caused the harm but rather, whether the defendant's act breaching his (negative) promise caused the harm. If the defendant promised not to disclose the plaintiff's secret, and then did so, with loss of customers to the plaintiff, the relevant question is whether the loss of customers was caused by the defendant's disclosure of the secret. This causal question is thought to be the counterfactual question: If the defendant had not breached his promise, that is, if the defendant had not disclosed the secret, would the plaintiff have lost his customers anyway? In the familiar language of the expectancy measure of contract damages, our law seeks to give the plaintiff "the benefit of his bargain," or in other words, it seeks to place him in the position he would have been in if, contrary to fact, the defendant had not breached his promise.

---

6. This trend is particularly true in cases based on the theory of promissory estoppel, which are already tort-like in their theory. Grant Gilmore famously proclaimed that contract would be absorbed by torts because of such developments. GRANT GILMORE, THE DEATH OF CONTRACT 87 (1974).

7. *See, e.g.*, E. ALLAN FARNSWORTH, CONTRACTS 784 (3d ed. 1999).

3. The Occasional Liability for Nonfeasance in Tort

As is well-known, Anglo-American tort law only occasionally imposes liability for breach of nonpromissory, positive obligations.[8] While we do have a general duty not to cause certain harms by our actions, tort law does not recognize a similarly general duty to prevent the same harms. Still, there are some pockets of true omission liability in tort law. These are based on relationships of various kinds, voluntary undertakings, or causal responsibility for the predicament of the plaintiff.[9]

In such cases of occasional liability for a defendant's nonfeasance, liability is standardly said to turn on the question of whether the defendant's omission caused the plaintiff's harm. If I starve another to death by not feeding him when I have a duty to do so, I am said to have caused his death by my omission. Again, such causation is standardly unpacked in terms of the counterfactual: If I had not omitted to feed the victim, that is, if I had fed him, would he have died?

4. The Basic Case of Nonfeasance in Contract

Unlike in torts, there is no reluctance to hold parties liable for nonfeasance in contracts. After all, the duty not to omit is readily supplied by the defendant's promise to do the action he later omits to do. Here again, the defendant must pay for losses caused by the omission to perform, and here again, such causation is cashed out in counterfactual terms: The plaintiff is entitled to the benefit of his bargain, in other words, the position he would have been in if, contrary to fact, the defendant had done what he had promised to do.

5. Limitation on Liability in Tort for Misfeasance Recharacterized as Nonfeasance

Sometimes, we draw a distinction between doing something, such as killing Jones, and merely allowing something, such as allowing Jones to die.[10] This is not the act-omission distinction because cases of allowing an event to occur may well involve active misfeasance by a defendant, such as the act of turning off a respirator. Rather, the distinction is a causal one: In cases of mere allowing an event to occur, we, by our

---

8. *See, e.g.*, W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 56, at 373 (W. Page Keeton ed., 5th ed. 1984).
9. *Id*. at 376–82.
10. *See* Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, 5 OXFORD REV. 5 (1967), *reprinted in* PHILIPPA FOOT, VIRTUES AND VICES AND OTHER ESSAYS IN MORAL PHILOSOPHY 19 (1978); Warren S. Quinn, *Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing*, 98 PHIL. REV. 287, 287–312 (1989).

action, remove some impediment that allows nature to cause some harm.[11] For example, we turn off the respirator, and in the natural course of affairs, the patient dies for lack of oxygen.

This causal distinction is picked up by the law of torts mostly in euthanasia cases.[12] "Passive euthanasia" involves allowing nature to take its course, whereas "active euthanasia" consists of killing people, that is, doing acts causing their deaths. This causal distinction in turn is translated into a counterfactual question: If the later act of the defendant (such as turning off the respirator) did no more than return the victim to the state he would have been in if, contrary to fact, the defendant had not first intervened (for example, by hooking the victim up to the respirator), then the defendant is not liable for the harm. The action of the defendant is classified as nonfeasance, even though it plainly is an act and not an omission.[13]

6. Permissions for Self-Help in Remedying Situations of Unjust Enrichment

The notion of unjust enrichment cuts across torts, contracts, property, and even criminal law. The basic idea of unjust enrichment is that whenever we are strongly entitled to something, and another gains by using or possessing our entitlement, that other is "unjustly enriched" and must disgorge what is not his. Thus, if my personal property ends up in your possession through no fault of your own, you must give it back. If I justifiably destroy your dock in a storm in order to save my ship, I just pay you the value of your dock (since I cannot give it back).[14] If you justifiably rescind a contract with me, you must give back benefits conferred upon you by me pursuant to the contract prior to its rescission.

Of interest in the present context are unjust enrichment cases where the aggrieved party engages in self-help to retrieve what is hers. Judy Thomson's famous example in the context of the abortion debate is of this kind.[15] While you are a spectator at the performance of the local

---

11. *See* Moore, *Causation and Responsibility*, *supra* note *, at 19, 41–42.
12. Barber v. Superior Court, 195 Cal. Rptr. 484 (Ct. App. 1983) (holding the removal of both extraordinary and ordinary life sustaining machines to be passive euthanasia in a criminal case).
13. So analyzed in MICHAEL S. MOORE, ACT AND CRIME: THE PHILOSOPHY OF ACTION AND ITS IMPLICATIONS FOR CRIMINAL LAW 26–27 (1993).
14. Vincent v. Lake Erie Transp. Co., 124 N.W. 221, 221 (Minn. 1910).
15. Judith Jarvis Thomson, *A Defense of Abortion*, 1 PHIL. & PUB. AFF. 47, 47–66 (1971).

symphony, you pass out and awaken to find yourself medically hooked up to the lead violinist of the symphony; you cannot be unhooked before nine months have run without ending his life; you unhook, and the violinist dies. Thomson intends that you intuit that you are not responsible for his death. This libertarian intuition might be built on the assumptions that you are not killing the violinist when you unhook him, you are not killing him because you are not causing his death, and you are not causing his death because you are only allowing him to die. This is a mere allowing of death rather than a causing of death because you are merely returning him to the situation he would have been in if, contrary to fact, he had not been hooked up to you in the first place. The violinist merely dies the death after being unhooked that he would have died had he never been hooked up to start with.

This way of supporting the libertarian justification for self-help returns us to the doing-allowing distinction. An alternative way of supporting the libertarian's conclusion here is by way of an agent-relative permission: When you unhook the violinist, you are admittedly causing his death, but you are entitled to do so by your strong right to the control of your own body. This alternative justification seems the more plausible one in actual cases of abortion, where many abortion procedures are too active to be plausibly cast as merely allowing nature to take its course. Nevertheless, this alternative justification still depends quite obviously on a counterfactual judgment: You are permitted to return the violinist (fetus) to the state he (it) would have been in but for the misappropriation of your body. In the example given, that is a state where there is no use of your body by another, even though in that state there is death of that other. Again, the death he later suffers is no worse than the death he would have suffered if, contrary to fact, there had been no violation of your bodily integrity to start with.

7. Limitations on Negligence Liability in Tort for Harms Not "Within the Risk"

A celebrated issue in American negligence law is whether a defendant is liable for harm he causes that is not within the risk, the taking of which made the defendant's act negligent.[16] For example, *A* hands a loaded gun to a child, who drops it on *B*'s foot. The injury to *B*'s foot is commonly said not to be within the type of injuries (like shooting) the risk of which made it negligent to do what *A* did.[17] The still-prevailing

---

16. *See generally* Hurd & Moore, *supra* note *.
17. A simplified version of the illustration is exemplified in the *Restatement of Torts*. *See* RESTATEMENT OF TORTS § 281 (1934).

view in American tort law is that a negligent defendant is not liable for harms he causes that were outside the risks that made his act negligent.[18] This is often, but far from invariably, translated into a causal question: Did the aspect of his action that made him negligent, that is, the risk-creating aspect, cause the harm?[19] Or as courts sometimes put it, even where we grant that his negligent *action* caused the harm, nonetheless it is another question whether the *fact* that that action was negligent caused the harm. A chiropractor who has no license to practice medicine may, by his acts of treatment, cause injury to another. Yet, the fact that his action was done without a license may be causally irrelevant to the occurrence of the harm.[20]

This causally framed limitation of liability is then usually cashed out in counterfactual terms: Even if the *acts* of medical treatment were clearly a necessary condition of the harm, nonetheless it is a separate question whether the *fact* that those acts were unlicensed was a necessary condition of the harm. Furthermore, courts often answer this counterfactual in the negative: Even if the medical practitioner had been licensed, it would have made no difference, that is, he still would have injured the plaintiff, and so there is no liability.

8. Absence of Liability in Overdetermination, Concurrent Cause Cases Where the Other Cause Is a Natural Event

Overdetermination cases have long been a puzzle for legal systems. An overdetermination case is one where two putative causal factors are in play and either is sufficient, by itself, to cause some single injury. A concurrent overdetermination case is one where such causal factors operate simultaneously. In torts, for example, two fires are burning their way towards the plaintiff's house. Either fire, by itself, will be sufficient to both reach the plaintiff's house and burn it to the ground.[21] As it happens, the two fires join, and it is the larger, resultant fire that destroys the plaintiff's house. Suppose the defendant has negligently started one of the two fires, but the other fire is of natural origin caused, for example, by lightning, spontaneous combustion, or the like. A minority of

---

18. *See* Hurd & Moore, *supra* note *, at 339–47.
19. *See generally* ROBERT E. KEETON, LEGAL CAUSE IN THE LAW OF TORTS (1963); Richard W. Wright, *Causation in Tort Law*, 73 CAL. L. REV. 1735 (1985).
20. Brown v. Shyne, 151 N.E. 197 (N.Y. 1926).
21. Anderson v. Minneapolis, St. Paul & Sault Ste. Marie Ry. Co., 179 N.W. 45 (Minn. 1920); Kingston v. Chi. & N.W. Ry., 211 N.W. 913 (Wis. 1927).

American courts deny liability of the defendant to the plaintiff for his house.[22]

The rationale for this result is that the defendant's act of starting his fire didn't cause the destruction of the plaintiff's house. And the rationale for that causal conclusion lies in the truth of this counterfactual: If the defendant had not started his fire, the plaintiff's house would have been burnt to the ground regardless. The defendant's action, in other words, made no difference to what was going to happen anyway in the ordinary course of nature.

9. Absence of Liability in Asymmetrical, Concurrent, Overdetermination Cases

An asymmetrical, concurrent, overdetermination case is one where there is one big cause ("big" meaning sufficient, by itself, to cause the harm), joined by one or more little causes ("little" meaning not big enough to cause the harm alone, either individually or jointly with the other little causes).[23] Some examples include the following: the defendant's small fire joins a much bigger fire, and the resultant fire destroys the plaintiff's house; the defendant's acts stop up the drainage wickets in a flood control levy, but such a big flood occurs that the harm to the plaintiff would have occurred even with unstopped drainage wickets;[24] the defendant nicks a cable holding the plaintiff's cable car, a large force well in excess of the original carrying capacity of the cable without the nick causes the cable to break, but it nonetheless breaks at the nicked point, sending the plaintiff to his doom.[25]

The standard view is that there is no liability in these cases.[26] This is because it is held that the defendant did not cause the injuries, and he did not cause the injuries because these injuries would have happened anyway in the natural course of events. If, contrary to fact, the defendant had not started his small fire, stopped up the drainage wickets, or nicked the cable, these injuries would still have occurred. Therefore, the defendant did not cause these injuries, and the defendant cannot be made to pay for them.

---

22. Cook v. Minneapolis, St. Paul & Sault Ste. Marie Ry. Co., 74 N.W. 561 (Wis. 1898). This appears to be the majority rule in the English Commonwealth. *See* PETER CANE, RESPONSIBILITY IN LAW AND MORALITY 121–22 (2002).
23. *See generally* Moore, *Causation and Responsibility*, *supra* note *, at 10–11.
24. City of Piqua v. Morris, 120 N.E. 300 (Ohio 1918).
25. For a variation of the example, see Wright, *supra* note 19, at 1794, 1800.
26. *See, e.g.*, *Morris*, 120 N.E. at 303.

10. Limitations on Damages in Preemptive
Overdetermination Cases

A preemptive overdetermination case is one where once again there are two events, each of which could be sufficient to cause some injury, yet unlike the concurrent cause cases, here one event preempts the other from becoming a cause of such injury.[27] In the two fires example, suppose the fires do not join and that the defendant's fire burns the plaintiff's house to the ground before the other fire arrives. The first fire preempts the second, so the first fire is universally held to be the cause of the harm. However, often damages are limited to just those caused by the first fire.[28] In such a case, the damage limitation is temporal, not spatial. In other words, it is not that only (a spatial) part of the plaintiff's house was burned by the defendant's fire; admittedly, all of it *was* burned by that fire. Rather, the idea is that defendant's fire only caused the loss of the use of the house during the interval of time between its destruction and when it would have been destroyed by the second fire. We measure, in other words, the amount of loss caused by the defendant by comparing what did happen to what would have happened had the defendant not started his fire.

In criminal law, of course, there are no damages recoverable by the victim to be limited in this way. Still, even in the criminal context, preemptive causation limits the liability of the preempting causer in the following way: If the harm that the defendant caused was about to be caused anyway by some natural occurrence, then the defendant may have a justification defense for his behavior that would otherwise be unavailable. In the lifeboat cases, for example, many find the result in *Dudley & Stephens*[29] to be wrong. Many agree with Glanville Williams, and they do so for the reason that he pointed out: that the cabin boy, who was killed and eaten so the rest could survive until rescued, was about to die anyway of natural causes.[30] Really, the argument is, those who

---

27. *See generally* Moore, *Causation and Responsibility*, *supra* note *, at 9–13.
28. Dillon v. Twin State Gas & Elec. Co., 163 A. 111, 115 (N.H. 1932); Jobling v. Associated Dairies Ltd., [1982] A.C. 794 (H.L. 1981) (appeal taken from Eng.).
29. The Queen v. Dudley & Stephens, 14 Q.B.D. 273 (Eng. C.A. 1884).
30. GLANVILLE WILLIAMS, CRIMINAL LAW 739–41 (2d ed. 1961). Williams's conclusion is an old one, reflecting the considered judgments of Cicero, Kant, Bacon, Holmes, and the drafters of the Model Penal Code. For discussion and citations, see Michael S. Moore, *Torture and the Balance of Evils*, 23 ISR. L. REV. 280, 303 (1989),

stabbed the cabin boy did not fully cause his death. Rather, they only accelerated it. This supposedly lessening of causal contribution is then cashed out in counterfactual terms: If the defendants had not stabbed the cabin boy, he would have died shortly anyway.

I shall argue in what follows that none of these legal results can be justified in causal terms. This is because counterfactuals are not to be equated with causation. This does not mean that the results themselves are erroneous. In some of these cases I shall argue that the results are erroneous, but in others the results may be justified by what I shall call counterfactual baseline tests for liability, tests that have nothing to do with causation.

If one sees that counterfactual dependence and causation are quite distinct, and sees further that there are at least pockets of liability (for example, omissions) which are properly tested by counterfactuals, then one can see the issue of whether or not tort and contract law are simply mistaken by their general asking of causal questions; perhaps the correct general criterion of liability is counterfactual and not causal. I close with speculations about the desirability of generally limiting liability by counterfactuals and not by causation.[31] Before coming to any of these issues, however, I shall begin with the nature of counterfactual judgments themselves. Being at the center of this analysis, it would be helpful if we had a good understanding of what we will be talking about. Unfortunately,

---

31. This general relational question, about whether causal dependency or counterfactual dependency better captures the moral and legal basis for liability, is to be distinguished from the question that interests Stephen Perry in his submission to the symposium. Stephen Perry, *Harm, Counterfactuals, and Compensation*, 40 SAN DIEGO L. REV. 1283 (2003). Perry's question, like that of Joel Feinberg and Seana Shiffrin, regards what states of affairs are properly characterized as "harms." *See* JOEL FEINBERG, *Wrongful Life and the Counterfactual Element in Harming*, *in* FREEDOM AND FULFILLMENT 3 (1992); *see also* Seana Valentine Shiffrin, *Wrongful Life, Procreative Responsibility, and the Significance of Harm*, 5 LEGAL THEORY 117, 120 (1999). My relational question is different and asks what relation must exist between the defendant's action and some state of affairs for liability to attach *if* that state of affairs comes to be characterized as a harm. Notice that there are also two competing answers to Professor Perry's characterization question, one of which is also counterfactual. The counterfactual answer competes with an historical worsening answer, comparing, respectively, the victim's present condition to the condition he would have been in, on some baseline notion, or the condition he was in fact in prior to the defendant's wrongful action. Notice also that, given the independence of the questions, to reject the counterfactual analysis of the general relational question will not necessarily entail the rejection of the counterfactual analysis of the characterization-as-a-harm question. One might, for example, require causation of some state *S* for liability, and then assess whether, or to what extent, *S* is a harm by counterfactual judgments involving both *S* and other more beneficial states. This is the gist of the issue in the very interesting offsetting benefit cases discussed by both Perry and Feinberg.

such an understanding is much more elusive than it might seem.

II. COUNTERFACTUAL CONDITIONALS

A. *The Class of Counterfactual Statements*

In light of the thorny philosophical thickets through which we must tread,[32] one hardly knows where to start. Perhaps we should first isolate the class of counterfactual statements. Usually the very name elicits the following sort of example:

(1)   If the rocket were to have gone less than 18,000 miles per hour, it would not have escaped the Earth's gravitational pull.

It is tempting to think that we can define the class of such statements by their grammar, namely, by the fact that they are framed in the subjunctive mood. Yet such a grammatical demarcation of the class of counterfactual statements is an unreliable guide.[33] Some subjunctive conditionals are not counterfactual conditionals, and some indicative conditionals are.

More promising is to frame the class of statements in which we are interested in terms of their conditional form conjoined with the falsity of both their antecedent and their consequent clauses.[34] With regard to (1), the rocket in question did not go less than 18,000 miles per hour, and it did escape the Earth's gravitational pull. Therefore, when considered separately, both the antecedent and the consequent clauses of the conditional statement are false.[35] The statement is thus doubly contrary

---

32.   The preliminary thickets consist of the complexity of the issues, the amount of philosophical attention to such issues, and yet the lack of any consensus as to their resolution, on basic questions about conditionals. For example, do they have some unitary, general nature? Are there two basic kinds of conditionals, indicative and subjunctive? If so, how does one draw the line between them? See generally two recent monographs which build upon a quite extensive body of literature: WILLIAM G. LYCAN, REAL CONDITIONALS (2001); MICHAEL WOODS, CONDITIONALS (David Wiggins ed., 1997).

33.   See WOODS, *supra* note 32, at 5 ("The subjunctive exists at best only vestigially in English . . . .").

34*.   See, e.g*., J.L. MACKIE, TRUTH PROBABILITY AND PARADOX 65 (1973). More exactly, as Mackie also notes, *id.* at 71, counterfactual statements typically express the speaker's *belief* that the antecedent and consequent clauses of such statements are false.

35.   It takes some grammatical alteration to get determinate truth values to each clause of a counterfactual statement because such statements are usually put in the subjunctive mood and in the past tense. I follow convention here in translating the clauses into the indicative mood, and in rendering them tenselessly, thusly: "The rocket goes less than 18,000 miles per hour" (the antecedent clause); and, "The rocket does not

to fact and is the classic sort of counterfactual conditional statement.

Suppose another massive body came close to the Earth just where and when the rocket was speeding away from Earth. Then this statement could be true:

(2) Even if, contrary to fact, the rocket did go less than 18,000 miles per hour, it would have escaped the Earth's gravitational pull.

Unlike the classic example of counterfactuals, here the antecedent is false, but the consequent is true. Nelson Goodman calls these statements "semifactuals."[36] Notice that the truth of (2) makes the truth of (1) impossible. As Goodman puts it, "[A] semifactual conditional has the force of denying what is affirmed by the opposite, fully counterfactual conditional."[37] We thus cannot leave out semifactuals from our analysis of counterfactuals, for they seem to be one rendering of the contradictory, or at least the contrary, of counterfactuals.[38]

Nor can we leave out what Goodman calls factual conditionals,[39] which are conditional statements where both the antecedent and the consequent are true. For example:

(3) If the rocket escaped the Earth's gravitational pull, then it did not go less than 18,000 miles per hour.

Note that (3) is just the contrapositive of (1); it is true if (1) is true, and it is false if (1) is false. (It is a tricky question when contrapositives of true counterfactuals are themselves true, but this one is.) Since counterfactuals like (1) can often be transformed into a factual like (3), such factual conditionals can no more be left out of an analysis of counterfactuals than can semifactuals.

The statements in which we are interested will thus not necessarily appear in explicitly counterfactual form. Still, it is the counterfactual form that highlights what seems most puzzling about this class of statements: What is it in the world that makes true a conditional statement, both clauses of which are false? What connects rockets that did not fall to Earth with velocities that were never attained?

---

escape the Earth's gravitational pull" (the consequent clause). The convention is quite controversial. *See* WOODS, *supra* note 32, at 9–10. But my use of it here is harmless because the very point of reformulating these clauses is to see if, so construed, the entire statement can be taken truth functionally. The conclusion being that it cannot, the mood and tense shifts are merely by way of arguendo concession.

36. Nelson Goodman, *The Problem of Counterfactual Conditionals*, 44 J. PHIL. 113, 114 (1947), *reprinted in* NELSON GOODMAN, FACT, FICTION, AND FORECAST 3 (4th ed. 1983).

37. *Id.* at 5.

38. This is Goodman's assumption. *Id.* at 5–6.

39. *Id*. at 4.

B. The Covering Law Account of Counterfactuals

Very generally speaking, there are two answers in the considerable literature of postwar, Anglo-American philosophy on counterfactuals.[40] One is given by the covering law view. On this view, if statement (1) is true, it is made true at least in part by the truth of the general law that no object going less than 18,000 miles per hour escapes the Earth's gravitational pull. The exact role of this general law in contributing to the truth of the singular counterfactual statement in (1) is a matter of some dispute. F.H. Bradley, at least as Roderick Chisholm once interpreted him, took the whole point of asserting a singular counterfactual like (1) to be to draw attention to, to emphasize, or to describe the underlying covering law.[41] Yet Chisholm himself declined this exclusive a role for covering laws in explicating singular counterfactual statements. Rather, Chisholm urged that the assertion of a singular counterfactual statement presupposes the law statement without either describing it or having as the whole point of the assertion the drawing of attention to its existence.[42]

On the covering law view, the truth of the law statement is one item that makes the singular counterfactual statement in (1) true. However, also necessary to the latter's truth are the truths of a variety of other singular statements, statements that rule out conditions wherein the law does not hold.[43] One such condition to be ruled out, for example, is that states of affairs making the semifactual in (2) true, namely, that another body of large mass is near the rocket's path as it ascends from Earth. On this enriched covering law view, the point in making singular counterfactual statements like (1) is to apply the covering law to the imagined situation described in the antecedent clause in (1): "Supposing that this rocket is going less than 18,000 miles per hour, and supposing that certain other conditions C_1, C_2 . . . , C_n remain true, then we can infer that this rocket will not escape the Earth's gravitational pull." In other words, we use, but do not mention, the covering law which licenses the inference we wish to draw.

---

40. *See generally* DAVID H. SANFORD, IF P, THEN Q: CONDITIONALS AND THE FOUNDATIONS OF REASONING 76–86, 101–18 (1989); MICHAEL TOOLEY, CAUSATION: A REALIST APPROACH 64 (1987).

41. Roderick M. Chisholm, *Law Statements and Counterfactual Inference*, 15 ANALYSIS 97, 102 (1955), *reprinted in* CAUSATION AND CONDITIONALS 147, 151–52 (Ernest Sosa ed., 1975) [hereinafter Chisholm, *Law Statements*]. This is a revisiting of Chisholm's influential article, Roderick M. Chisholm, *The Contrary-to-Fact Conditional*, 55 MIND 289 (1946) [hereinafter Chisholm, *The Contrary-to-Fact Conditional*].

42. Chisholm, *Law Statements*, *supra* note 41, at 152.

43. Goodman, *supra* note 36, at 8.

The covering law analysis of counterfactuals should not be taken to imply that we can translate counterfactual statements like (1) into the covering law(s) on whom the truth of (1) depends. The suggested translation would be arrived at by taking the covering law to be expressed in truth functional form as:

(4) $(x)\,(Fx \supset \sim Gx)$

where F is the predicate, "is traveling less than 18,000 miles per hour," and G is the predicate, "escapes from the Earth's gravitational pull." The singular counterfactual statement would then be translated as an open sentence instantiating the relevant law, namely:

(5) $Fx \supset \sim Gx$
(If the rocket is traveling less than 18,000 miles per hour, then the rocket does not escape from the Earth's gravitational pull.)

What makes such a translation tempting is that it translates the troublesome counterfactual conditional into the familiar material conditional of modern logic.

Such a translation will not go through, however, because of the lack of connection between F and G in (4) and (5). The material conditional is a truth functional connective, meaning that the truth of $Fx \supset \sim Gx$ turns exclusively on the truth of Fx and of $\sim Gx$. This gives rise to the well-known "paradoxes of material conditionals," paradoxes making it difficult to translate any "if . . . then . . ." conditional in English into the material conditional of logic.[44] It is sufficient for the truth of $Fx \supset \sim Gx$, if either: (a) Fx is false or (b) Gx is false (meaning $\sim Gx$ is true). This leaves out what seems essential to counterfactual conditional statements, namely, that their truth depends on there being some connection between x being F and x being G. Only if such a connection exists is the counterfactual in (1) true, which is not at all captured by attempted translations like (5). Moreover, notice that because all counterfactual and semifactual statements like (1) and (2) have false antecedents, translating them into the material conditional would make all of them true. This is bad enough by itself, but it is particularly troublesome that (1) and (2) could both be true since they seem to be contradictories of one another.

The obvious suggestion to repair these difficulties is to turn to strict conditionals for the translation. This would be to translate (1) into the following:

---

44.   *See, e.g.*, WESLEY C. SALMON, LOGIC 38 (2d ed. 1973).

(6) $\Box\,(Fx \supset \sim Gx)$

    (Necessarily, if the rocket is traveling less than 18,000 miles per hour, the rocket does not escape the Earth's gravitational pull.)

The strict conditional is almost as familiar to us as the material conditional,[45] so if counterfactual conditionals like (1) could be translated into strict conditionals like (6), that would lessen their uniqueness. Yet (6) suffers from the same defect as (5) in that it fails to capture the connectedness of *F* and *G*. The statement, $\Box\,(Fx \supset \sim Gx)$, will be true (1) if *x* being *F* is impossible, or (2) if *x* being *G* is impossible (making "$\sim Gx$" necessary). No connection between being *F* and being *G* is needed for the truth of the strict conditional in (6).

On the covering law account, one can use the strict conditional to show that a statement of the antecedent (*Fx*), a statement of other conditions (*S*), and a statement of the relevant law $((x)((Fx \bullet S) \supset \sim Gx))$ entail the statement of the consequent, as in:

(7) $\Box\,(((x)\,((Fx \bullet S) \supset \sim Gx) \bullet S \bullet Fx) \supset \sim Gx)$

Yet this is not to translate the counterfactual conditional in (1) into the strict conditional in (6). Statement (7) is no more than the logical truth known as modus poneus.

The covering law account of counterfactuals thus cannot avail itself of the two familiar conditionals we know for any direct translations. Counterfactuals like (1) license us to infer $\sim Gx$ from *Fx* (together with other conditions and the covering law), yet the counterfactual conditionals themselves resist reduction to some statement of that law.

One of the pluses of the covering law analysis of counterfactual statements is its ontological economy. Laws are ontologically no more mysterious than are universals generally—which is to say, a problem for nominalists perhaps, but not the kind of problem besetting the alternative, possible worlds account, to be discussed shortly. On the covering law model there are no ghostly particulars existing in some other world related to particulars in this world in some recherché manner. Rather, there are only properties, kinds, and relations, and the laws that connect them.

Despite the plausibility of the covering law account, both generally and in terms of its ontology, there are a number of objections to the

---

45. *See generally* SANFORD, *supra* note 40, at 87–100.

account. One is the objection that some counterfactuals seem to involve no covering laws in the background. Consider David Sanford's example:

(8) If I had waited four more days to take out my loan, I could have borrowed at a lower interest rate.[46]

As Sanford points out, there seems to be no law connecting the antecedent to the consequent. My taking the loan out early, for example, did not cause lower interest rates. It is true that a high volume of delayed loan applications should have the required effect by creating a lesser demand for loans, and thus, a lower price in terms of a lower interest rate. Yet this law is not invoked by (4). Rather, there simply were lower interest rates four days later, due to some causes no doubt, but not necessarily due to a drop in demand.

A defender of the covering law account of counterfactuals might reply that there are laws in such examples, even though they are not laws connecting the state of affairs referred to in the antecedent to those referred to in the consequent. Rather, the laws are those connecting the rate I will get to the rates generally prevailing in the relevant loan market, and the conditions include that prevailing rate together with those things that have to be true in order for me to receive the prevailing market rate. Yet these laws do not license us to infer the consequent from the antecedent in (8). Rather, this only allows us to infer the consequent from some fact independent of the antecedent, namely, the fact about the prevailing interest rate. Thus, these laws do not help us to make out the covering law analysis for counterfactuals like (8).

What Sanford is plainly after is an example of a counterfactual that uses one "just so" fact, such as the interest rate on a given date, as the basis for inferring the state of affairs described in the consequent clause, such as the interest rate on my loan. Yet to move from his example of a "just so" fact to the desired conclusion in examples like (8), some laws seem to be necessary, although not laws connecting consequent to antecedent. A better counterexample would be:

(9) If I had waited four more days to take out my loan, the interest rate would still have been 100 basis points lower than it was when I, in fact, took out my loan.

Now the consequent refers to a "just so" fact, the fact about the lower interest rate, so no laws are needed to infer a statement of the fact from statements of some other fact. Yet notice that (9) is a semifactual. That is, its antecedent is false, while its consequent is true. As we have seen, such

---

46. *Id.* at 86.

semifactuals are the contradictories of counterfactuals like the following:

(10) If I had waited four more days to take out my loan, the interest rate would not have dropped 100 basis points but would have remained where it was four days earlier.

If (10) were true, it would only be true because of some law connecting my loan application to the interest rate. Similarly, if (9) is true and (10) is false, that will only be because of the laws determining interest rates, laws under which my single loan application makes no difference.

Consider another of example of Sanford's:

(11) If we had bought one more artichoke this morning, we would have had one for everyone at dinner tonight.[47]

Suppose we bought eight artichokes this morning, but nine people were at dinner tonight. Then what (11) is really saying is that:

(12) If we had bought 1 plus 8 artichokes this morning, we would have had 9 artichokes for dinner tonight.

Note that (12) is just the arithmetic analogue of noncausal counterfactuals, such as:

(13) If this rectangle were a triangle, then it would have one less side;

or

(14) If this man were immortal, then he would live forever.

The laws behind such statements are not causal laws, to be sure. Yet, there are truths of a general sort at work in such statements.

There is a class of counterfactuals that Nicholas Rescher has dubbed "purely hypothetical counterfactuals" that do not seem to involve any laws or other generalizations.[48] For example (from Quine):

(15) If Bizet and Verdi were compatriots, Bizet would be an Italian.

But the law-less character of such counterfactuals seems to be at one with the complete indeterminacy of such counterfactuals. One way Bizet and Verdi could have been compatriots would be for Bizet to have been

---

47. *Id*. at 173.
48. Nicholas Rescher, *Belief-Contravening Suppositions*, 70 Phil. Rev. 176, 193–94 (1961), *reprinted in* Causation and Conditionals, *supra* note 41, at 156, 162 .

an Italian. However, another would be for Verdi to have been a Frenchman. Yet another would have been for both to have been Argentineans, etc. If we eliminate the indeterminacy by supposing that only Italy allowed two such famous composers to be citizens of the same country, then (15) is true, and the other suggestions are false. Yet the determinacy of such purely hypothetical counterfactuals is purchased by the law mentioned: Only Italy allowed such famous compatriot composers.

I conclude that it is more difficult than is commonly supposed to find nonanalytic, determinate counterfactuals that have no covering law as part of their truth conditions. This fact prepares the way for a second objection to the covering law view, however. The objection is that the very idea of a law has the idea of counterfactuals built into it.[49] The worry is that the covering law account of counterfactuals is circular, depending as it does on laws that themselves depend on counterfactuals.

The circularity worry begins by asking, how do we distinguish an accidental generalization from a covering law? An accidental generalization shares with true laws a logically general form. For example:

(16) All pieces of pure gold are less than one cubic mile in size.[50]

This is universal in form predicating of anything, if it is a piece of gold, then it is less than one cubic mile in size. Yet such a generalization, while true, is true by the happenstance that no pieces of gold that large have ever formed. However, there is no reason that such huge pieces of gold could never have formed. Rather, it just so happens they didn't. Contrast (16) to:

(17) All pieces of pure uranium-235 are less than one cubic mile in size.[51]

From what we know about the critical mass of uranium-235, there cannot be a cubic mile of the stuff. Note that (17) is a law of nature, whereas (16), while universal in form and true, is not.

The worry over circularity stems from the fact that the most obvious way to distinguish laws like (17) from accidentally true generalizations like (16) is via counterfactuals.[52] If, contrary to fact, geologic processes had occurred in certain ways, then there would be a cubic mile of gold. However, if analogous geologic processes had occurred with respect to

---

49. Goodman called this "the problem of law" for counterfactual conditionals. *See* Goodman, *supra* note 36, at 17–27.
50. This example is taken from D.M. ARMSTRONG, WHAT IS A LAW OF NATURE? 17–18 (1983).
51. *Id.*
52. *See* William Kneale, *Natural Laws and Contrary-to-Fact Conditionals*, 10 ANALYSIS 121 (1950), *reprinted in* PHILOSOPHY AND ANALYSIS 226 (Margaret MacDonald ed., 1954).

uranium, there would not be a cubic mile of uranium. The causal law seems in some sense to be necessarily true, whereas the accidental generalization is not. Furthermore, that necessity is tested by hypothetical cases that have not occurred.

To draw the sting from the circularity worry is to give an account of laws that does not depend on counterfactuals. If such an account is given, then we can use that notion of laws to explicate counterfactuals without circularity. Nelson Goodman's solution, to what may fairly be called his own problem, satisfies almost no one:

> What then does distinguish a law . . . from a true and general non-law . . . ? Primarily, I would like to suggest, the fact that the first is accepted as true while many cases of it remain to be determined, the further, unexamined cases being predicted to conform with it. The second . . . is accepted as a description of . . . all cases, no prediction of any of its instances being based upon it.[53]

This social fact of acceptance, in the face of less than total data, is in turn based on the predicates contained within a generalization: If the predicate is "projectable," the generalization is a law that can be used to predict unexamined cases. The projectability of a predicate, in turn, is a function of another social fact, namely, how "entrenched" in past practice is the predicate?[54]

To be blunt, this amounts to saying that a law differs from an accidentally true generalization because we treat it as such. Anyone with a less conventionalist or relativist metaphysics will find this unhelpful. Realists about universals will have a quite different solution. Realists will suggest that laws describe relations between universals, whereas accidental generalizations do not.[55] On this view, there is a relation between the size of a chunk of uranium and its existence, whereas there is no such relationship between the size of a chunk of gold and its existence. Of course, the realist has no more resources with which to verify the existence of a true law than does his relativist cousin. Nevertheless, at least he is not defining a law as simply the psycho-social fact that many

---

53. Goodman, *supra* note 36, at 20.
54*. Id*.
55*. E.g.*, ARMSTRONG, *supra* note 50, at 16. When discussing singular causal relations, realists will also have a non-question-begging answer to Goodman's "problem of law." The distinction between a genuine law and an accidental generalization is already built into the distinction between a genuine singular causal relationship and a mere coincidence of spatio-temporal succession; induction over the former produces a general law, whereas generalization over the latter produces an accidentally true generalization at best. I owe this point to Professor Richard Fumerton.

have a willingness to project because they accept past projections.

Laws do sustain counterfactual judgments, whereas accidental generalizations do not. Yet if laws describe relations between properties, then what a law is is not given by counterfactuals describing projections onto unexamined cases. Then one can say, as the covering law view of counterfactuals does, that a genuine law is part of the truth conditions for a counterfactual statement, without having the embarrassment that what makes the law in question a law is just its support of some counterfactual statements.

The third objection to the covering law view of counterfactuals is perhaps the most discussed. I shall call this the problem of "incomplete specification." Nelson Goodman's famous example is of a match being struck.[56] There is some law to the effect that a match, if dry, made a certain way, oxygen is present, etc., will light if scratched on an abrasive surface. The law supports the counterfactual:

(18) If the match were struck, it would light.

Yet, as Goodman points out, the law equally as well supports the counterfactual:[57]

(19) If the match were struck, it would be damp.

To see this, reformulate the "law of matches" as:

(20) ~ (struck • well made • oxygen present • dry • unlit).

If a match were struck, then the law is equally satisfied by it being dry and lighting (18), and by it being wet and not lighting (19). What is it, Goodman asked, that justifies us in placing dryness in S (the set of conditions statements) and litness in the consequent, rather than unlitness in S and wetness in the consequent?

In retrospect it is puzzling that a generation or two of philosophers puzzled about this problem, conceiving the problem to be one of finding some formal characteristics of statements that made them assignable to the statement of conditions or to the consequent clause of the counterfactual.[58] It seems obvious that there can be no such formal characteristics and that the solution is wholly pragmatic.[59] Formally, the law in (20), together with an antecedent clause supposing a match to be struck, supports

---

56. Goodman, *supra* note 36, at 8–17.
57. *Id.* at 14.
58. For an example, see Wilfrid Sellars, *Counterfactuals, Dispositions, and the Causal Modalities, in* 2 MINNESOTA STUDIES IN THE PHILOSOPHY OF SCIENCE 225, 227–48 (Herbert Feigl et al. eds., 1958), *reprinted in* CAUSATION AND CONDITIONALS, *supra* note 41, at 126 .
59. *See* Rescher, *supra* note 48.

equally as well (18) and (19), together with other counterfactuals like the following:

(21) If the match were struck, then there was no oxygen present.

If (18) seems like the counterfactual, that is because most times the question we ask about a struck match is whether it lit or not. If matches were used as wetting agents, then struck but unlit matches would be our assumptions, and wetness, the state of affairs we cared about, would be in the consequent. We would then most typically derive some counterfactual other than (18) from the match law. Perhaps not (19), but:

(22) If the match were unlit (when struck, well made, with oxygen present), then it would be wet.

What Goodman's puzzle really reveals is the indeterminateness of our question when we ask after the truth of some counterfactual. Note that *S*, the crucial set of statements about background conditions being held constant, is not explicitly specified. What goes into *S* depends on the asserter of the counterfactual statement. Given the law of matches, if "is dry" is put into *S*, then (18) is true. However, if "is wet" is put into *S*, then the contrary semifactual:

(23) If the match were struck, it would not have lit

is true and (18) is false. Only by making assumptions about what the speaker intends to place in *S* on the occasion of his utterance of (18) do we get determinate truth values to the statement uttered.

Even if only a problem in pragmatics, the problem thus revealed is a major one for counterfactuals. What goes into *S* to give determinateness to counterfactuals like:

(24) If Jefferson were alive today, he would disapprove of electronic eavesdropping.

or

(25) If Kaiser Wilhelm II were not so stupid, World War I would never have happened.

The truth of these counterfactuals is so uncertain because what they assert is so indeterminate. How are we to unpack what is changed, and what is left the same, when we move Jefferson to today or when we up

the Kaiser's IQ? For example, what are we picturing about Jefferson's education about the world of today? With no education, Jefferson's response to electronic eavesdropping is probably no more than a stupefied silence. However, with exactly our education, Jefferson's response is probably our own.

The problem of indeterminacy does not uniquely bedevil the covering law account of counterfactuals. The problem is endemic to counterfactuals and thus besets all accounts of them. Yet with what resources an account may attack the problem differs between accounts. On the covering law account, Goodman's point should be construed as this: Laws, as such, do little to resolve such indeterminacies.

C. The Possible Worlds Account of Counterfactuals

The second account of counterfactuals often begins with an analysis of conditional statements generally. According to this analysis, an assertion of a conditional is construed to be a conditional assertion.[60] The idea is to break apart all conditional statements, including counterfactual statements like (1), into a statement of the consequent by itself, with a condition attached to the assertion of the statement. As Von Wright described this approach, "I shall never speak of the conditional as a proposition which is being asserted, but only of propositions being asserted conditionally . . . ."[61] Likewise, as Quine has stated, "[a]n affirmation of the form 'if p then q' is commonly felt less as an affirmation of a conditional than as a conditional affirmation of the consequent."[62] In effect, "if p, then q" is taken to be "q," asserted whenever p is true.

One might well wonder how this general account of conditional statements could be applied to counterfactual conditional statements, that is, to statements where we know p is false. On its face, it would appear that such conditionals are all senseless because the speaker knows that the condition of asserting either the consequent or its negation, namely p, does not obtain. As Frank Ramsey put it:

> If two people are arguing "If p will q?" and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q . . . . [I]n a sense, "If p, q" and "If p, [~ q]" are contradictories. . . . If p turns out false, these degrees of belief [in q given p] are rendered *void*. If either party believes [~ p] for certain, the question ceases to mean anything to him except . . . what follows from certain laws or hypotheses.[63]

---

60. *See* WOODS, *supra* note 32, at 14.
61. GEORG HENRIK VON WRIGHT, LOGICAL STUDIES 131 (1957).
62. W.V. QUINE, METHODS OF LOGIC 21 (4th ed. 1982).
63. FRANK PLUMPTON RAMSEY, THE FOUNDATIONS OF MATHEMATICS 247 n.1 (R.B.

For Ramsey, then, the conditional-assertion-of-the-consequent view of conditionals generally reverts to the covering law view for counterfactual conditionals.[64]

In order to avoid this reverting back to the covering law view, what is needed is something in the world to which the consequent clause q can refer. If so, then q's truth value can be determined separately from its law-like connection to the antecedent clause p. Robert Stalnaker approached this problem by first asking after the justification for beliefs in counterfactuals.[65] Following this approach, one finds the justification conditions for belief in q by repairing to the total set of one's beliefs. As Stalnaker has stated, "First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is then true."[66] The totality of our beliefs, when revised in the way Stalnaker suggests, in effect creates a kind of world that Leibnitz called a "possible world." It is in that imagined world, created by our revised total beliefs, that we are to test the truth of the consequent of a counterfactual like (1).

If these are the conditions under which we are justified in believing in or asserting a counterfactual, then it is but a short step to the truth conditions for counterfactual statements. All we need is to move from the psychological notion of hypothetical beliefs to the ontological notion of the possible worlds such beliefs are about. As Stalnaker notes, "The concept of a *possible world* is just what we need . . . since a possible world is the ontological analogue of a stock of hypothetical beliefs."[67] The truth test is then analogous to the belief test quoted above. To paraphrase Stalnaker's earlier quoted statement, we must, first, add the state of affairs described in the antecedent of a counterfactual to the actually existing states of affairs of the world; second, make whatever adjustments are required to maintain consistency without denying the existence of the state of affairs added by the antecedent; and finally, look to see whether the consequent is true in that possible world. As Stalnaker adds, when adjusting the actual world into a possible world in

---

Braithwaite ed., 1954).

  64. For this interpretation of Ramsey, see WOODS, *supra* note 32, at 20.

  65. Robert C. Stalnaker, *A Theory of Conditionals*, *in* STUDIES IN LOGICAL THEORY 98 (Nicholas Rescher ed., 1968), *reprinted in* CAUSATION AND CONDITIONALS, *supra* note 41, at 165.

  66*.* *Id.* at 169.

  67*.* *Id.*

which the state of affairs described in the antecedent exists, we should look for a possible world "which otherwise differs minimally from the actual world."[68] It is in that minimally different possible world that we test the truth of "if p, then q," by seeing if q is true.

It is worth emphasizing how different this view is from the covering law view. On the covering law view, it is the lawful connection of p to q that governs the truth of, "if p, then q." On the possible world view, it is the truth of q alone that governs the truth of this conditional, with the caveat that it is the truth of q not in this world, but in a merely possible world. Consider for example how the two accounts differ in their treatment of semifactuals, where q is true in the actual world. On the possible world account, adding p, which is false, to the actual world results in a possible world, but in that world q is true as it is in the actual world, so that "if p, then q" is true. On the covering law account, "if p, then q" must depend on the connection between p and q. However, there is no connection. Those insisting on such a connection are forced to Goodman and Chisholm's assertion that the content of semifactuals is to *deny* a lawful connection between p and q.[69]

There are two very large problems with the possible world conception of counterfactuals. One is the ontology demanded by the theory. The theory defines the truth of some counterfactual, "if p, then q," in terms of the truth of q alone, in some possible world consistent with p. This seems to require that such possible worlds be quite real, real enough for us to ask whether q is true in them (even though false in the actual world). David Lewis has been quite up front about the "modal realism" required by the possible worlds interpretation of counterfactuals:

> When I profess realism about possible worlds, I mean to be taken literally. . . .
>   . . . Our actual world is only one world among others. We call it alone actual not because it differs in kind from all the rest but because it is the world we inhabit. The inhabitants of other worlds may truly call their own worlds actual, if they mean by "actual" what we do . . . .[70]

It is because such worlds are as real as the actual world which allows us to seek the truths of counterfactuals in those worlds.

As Lewis recognizes, many philosophers meet suggestions of such

---

68. *Id.*
69. Chisholm, *The Contrary-to-Fact Conditional*, *supra* note 41, at 298, *revised and reprinted in* READINGS IN PHILOSOPHICAL ANALYSIS 492 (H. Fiegl & W. Sellars eds., 1949); Goodman, *supra* note 36, at 6.
70. DAVID LEWIS, COUNTERFACTUALS 85–86 (1973) [hereinafter LEWIS, COUNTERFACTUALS]. Lewis's more sustained defense of modal realism will be found in DAVID LEWIS, ON THE PLURALITY OF WORLDS (1986) [hereinafter LEWIS, ON THE PLURALITY OF WORLDS]. Further defense is provided by JOHN DIVERS, POSSIBLE WORLDS (2002).

modal realism with "incredulous stares."[71]   Such stares reflect a judgment that the doctrine is so counterintuitive that it ought to be rejected out of hand. Hilary Putnam, for example, dismisses such realism as no more than a "dotty idea."[72] Admittedly, the idea is a difficult one to accept, even for those of us with quite strong stomachs for supposedly queer ontologies.

A natural thought is that perhaps the possible world interpretation of counterfactuals could make do with a less extravagant ontology. David Armstrong, for example, being unable to "believe in the literal reality of possible worlds" but knowing of no other way to deal with laws of nature except by considering possible worlds "cling[s] to the hope that an account of 'possible worlds' can be given which does not assume the existence of *possibilia*."[73]   While Lewis has spent some time dashing various versions of such hope,[74] perhaps some such, less expensive, account of possible worlds will be found. I leave that question open, with some skepticism that a nonconnectivist account of counterfactuals, one whereby we define the truth of the conditional solely in terms of the truth of its consequent clause in a possible world, can make do with anything less than Lewis's full-blown modal realism.

The second problem with the possible worlds view of counterfactuals is the problem of incomplete specification that we encountered before in our discussion of the covering law theory. To put this problem into the language of possible world semantics, it is the problem of specifying the possible world in which we are to test the truth of the consequent of some counterfactual like (1). Stalnaker, as we saw, required that the possible world selected "differ minimally" from the actual world.[75] We know that the possible world selected has to differ somewhat from the actual world, for in the actual world the antecedent is false, whereas we need a possible world in which it is true. Yet what changes in this selected world where we find, contrary to actual fact, that this rocket is going less than 18,000 miles per hour? Did it suddenly stop going 18,000 miles per hour? If so, was this an ad hoc violation of the laws of inertia? If not, did it slow gradually for some reason? What reason? Moreover, there is even vagueness in specifying just what it is in the

---

71.   LEWIS, COUNTERFACTUALS, *supra* note 70, at 86.
72.   HILARY PUTNAM, REALISM AND REASON 218 (1983).
73.   ARMSTRONG, *supra* note 50, at 163.
74.   LEWIS, ON THE PLURALITY OF WORLDS, *supra* note 70, at 136–42.
75.   Stalnaker, *supra* note 65, at 169.

antecedent that changes. When we say, "Contrary to fact, the rocket slowed to less than 18,000 miles per hour," how much did it slow? When we say, "If that hand-waving had not occurred, the auctioneer wouldn't have thought you were bidding,"[76] how much of that hand-wave are we removing in the possible world we are to use? All of it? Half of it horizontally? Two-thirds of it vertically? That part of it necessary for the hand-wave to have been recognized by the auctioneer as a bid? Or, when we say, "But for the hammer blow, the chestnut would not have been flattened,"[77] how much of the hammer blow is removed in the possible world we are to use? The whole blow? The actual contact only? The full force, or only the part of it necessary for the flattening of the chestnut?

   With regard to the last mentioned vagueness problem of specifying what in the antecedent we are supposing to change, J.L. Mackie urged that in ordinary speech:

> we regard the hammer-blow as a unit, and simply do not consider parts or subdivisions of it or quantitative alterations to it. The alternatives considered are that I strike the chestnut in the way described and that I do not. In constructing possible worlds, in considering what might or would have happened, we either plug in the hammer-blow as a whole or leave it out as a whole.[78]

Similarly, David Lewis thought that in ordinary uses of "c caused e," when supposing counterfactually that c does not occur, "we imagine that c is completely and cleanly excised from history, leaving behind no fragment or approximation of itself."[79]

---

76.   Jonathan Bennett, *Event Causation: The Counterfactual Analysis*, *in* PHILOSOPHICAL PERSPECTIVES, 1, METAPHYSICS 367, 369 (James E. Tomberlin ed., 1987), *reprinted in* CAUSATION 217, 219 (Ernest Sosa & Michael Tooley eds., 1993).
77.   J.L. MACKIE, THE CEMENT OF THE UNIVERSE 43–44 (1974).
78.   *Id.* at 44. Mackie recognizes that a scientific reconstruction of ordinary speech might require some abandonment of this "whole event" convention. Where strength of hammer-blows and flattening of chestnuts are matters of continuous variation that are functionally related (greater strength of blow, greater flattening, in some relation other than direct proportion), we could take "c causes e" to refer to this law of functional dependence; then there would be no worry about how precisely to specify how much of c to subtract in constructing a possible world in which to test the truth of the proposition "e occurs." For there would be a range of possible worlds equally relevant to the truth of the causal judgment, varying by how much of c is subtracted from the antecedent.
79.   David Lewis, *Causation as Influence*, 97 J. PHIL. 182, 190 (2000). Lewis, like Mackie, *see* MACKIE, *supra* note 77, would ultimately regiment ordinary usage of counterfactuals so as to excise this whole event convention. Also, like Mackie, Lewis came to urge that (at least for counterfactuals involved in causation) we should be taken to be specifying a range of variation in the antecedent. Lewis, *supra*, at 189–90. When verifying such counterfactuals, we are then looking at a range of events in the consequent, hoping to discover a pattern of counterfactual dependence between the range of properties of consequent-events and the range of properties of antecedent events. Unlike Mackie, Lewis's "pattern of counterfactual dependence" would be metaphysically prior to any causal law and not a product of it. This is because of the independence of

If we adopt this "whole event" convention for what we are subtracting when we frame the antecedent of a counterfactual, that leaves the larger problem of specifying the possible world in which we are to test the consequent. Stalnaker tells us that we select a possible world where "there are no differences between the actual world and the selected world except those that are required, implicitly or explicitly, by the antecedent."[80] Alternatively, as Stalnaker puts it, "among the alternative ways of making the required changes, one must choose one that does the least violence to the correct description and explanation of the actual world."[81] Such lesser violence is done when a possible world is selected that more closely resembles the actual world than does any other possible world in which the antecedent is true.

Lewis relies on a similar criterion for selecting the possible world(s) consistent with the antecedent of some counterfactual, in which we are to seek the truth value of the consequent of that counterfactual. We are to rely on an ordering of overall similarity between possible worlds consistent with the antecedent, and test the consequent in a possible world that is high up in the ordering of possible worlds. In constructing this ordering of overall comparative similarity to the actual world, we trade off (1) changes of particular fact against (2) ad hoc violations of laws, what Lewis refers to as "small miracles," and both against (3) wholesale changes in spatio-temporal regions of particular fact, and against (4) wholesale overturning of laws.[82]

It is an illusion to think that the possible world that is overall closest to the actual world is one where only the state of affairs referred to in the antecedent changes. In the example of (1), this would be to think that the rocket just is going much slower in this possible world, but in all other respects that world is the same as the actual world. For this to be the case, at the very least we need some laws not to hold in this possible world, namely, all the laws that connect the rocket going over 18,000

Lewis's possible worlds analysis of counterfactuals from Mackie's analysis of counterfactuals in terms of nomic sufficiency.

80. Stalnaker, *supra* note 65, at 171.
81. *Id.*
82. David Lewis, *Causation*, 70 J. Phil. 556, 556–67 (1973), *reprinted in* Causation and Conditionals, *supra* note 41, at 180, 184 [hereinafter Lewis, *Causation*]. Lewis more explicitly orders these four aspects of similarity in David Lewis, *Counterfactual Dependence and Time's Arrow*, 13 Noûs 455, 472 (1979), *reprinted in* 2 David Lewis, Philosophical Papers 32, 47–48 (1986) [hereinafter Lewis, *Counterfactual Dependence and Time's Arrow*].

miles per hour in the actual world to earlier events that caused it to go that fast in the actual world. For, remember, if those earlier events all occurred in this "closest" possible world too, then it must have been the laws that changed to produce the lesser speed of the rocket in this possible world. Once we grasp this fact, we will most likely withdraw the view that all other events being the same makes for a more similar world, given that so many laws are miraculously suspended in that world.

Heavy lifting is done on this account of counterfactuals by the notion of overall similarity. It is a real question whether the notion is up to the work assigned to it. To begin with, similarity, like qualitative identity, is not a primitive relation. It is not, that is, primitive like the relation "redder than," which is parasitic on no property save redness itself. There is no simple property of similarity. Rather, one item is similar to another with respect to certain other properties. One diamond may be similar to another in that they both have the same cut, clarity, and so on.

So, if the idea of overall similarity is to make sense, it must be done in terms of some combinatory function over the properties that two particulars share. There are two possibilities for such a function. One would be to utilize our ordinary notion of similarity, which is a highly contextualized notion. It takes a context of utterance to find the truth of statements of the form "x is similar to y" inasmuch as every particular is similar to any other particular in some respects and is dissimilar in others. It requires a context of utterance to isolate what the relevant respect(s) is for any given similarity statement.

Lewis recognizes the contextual nature of ordinary similarity judgments and at times urges that counterfactual utterances provide the needed clues about salient versus not-so-salient similarities between possible worlds.[83] Yet the utterance of counterfactuals like (1) does not seem context-dependent as do ordinary similarity statements. Consider the statement, "This diamond is similar to that diamond," and imagine different contexts of utterance, for example, by a husband replacing his wife's lost engagement stone, by an expert diamond cutter reporting on his copying of some cut on another diamond, or by an insurance executive assessing the value of a stolen diamond. The context may well reveal relevant respects by which similarity is to be judged, such as overall appearance to the naked eye, cut, value, respectively. By contrast, when one utters counterfactuals about rockets falling to Earth when traveling at less than escape velocity, matches lighting when struck, chestnuts flattening when struck by a hammer, and the like, the relevant likenesses all seem to be drawn from the relevant scientific law.

---

83. LEWIS, COUNTERFACTUALS, *supra* note 70, at 91.

In which case, the possible world approach just is the covering law approach, albeit with a heavy-duty ontology as the proffered solution to the law problem.

The other possibility for making sense of similarity judgments about possible worlds would be to take the "overall comparative similarity" of such worlds literally. Literally, overall similarity is a function of all the properties there are and a summing of how many of them are shared by the two particulars being compared for similarity. More properties shared, more similar, would be the metric here. One has to be a realist about universals to make much sense of this proposed similarity metric, but for some of us, that is not a problem.[84] Nevertheless, the calculability of such similarity is, to say the least, more than a bit mind-boggling. While recognizing this, Lewis nonetheless urges that "[s]omehow, we *do* have a familiar notion of comparative overall similarity"[85] and that we form this notion "by balancing off many respects of similarity and difference."[86]

I think it is problematic whether we ever employ such an overall notion of similarity. But even if we do not, perhaps Lewis is entitled to construct such a notion. Arguably, that is exactly what he does when he prioritizes likeness of large spatio-temporal regions over likeness of laws over likeness of isolated, particular events.[87] This, however, leads to two further questions. One is the question of motivation: What motivates the selection of a nonordinary notion of similarity? Is it just an intuitive sense of how counterfactuals should come out in their truth values that drives a notion of similarity to achieve that result? In this case "similarity" would be doing no epistemic work,[88] although it could denote the correct metaphysics of counterfactuals.

The second question is one of correctness. By Lewis's notion of similarity, what is the truth value of this counterfactual?

---

84. 2 D.M. ARMSTRONG, A THEORY OF UNIVERSALS: UNIVERSALS AND SCIENTIFIC REALISM 13 (1978). Realism about universals is presupposed because, for nominalists, every particular shares exactly the same number of properties with every other particular. *See* NELSON GOODMAN, *Seven Strictures on Similarity*, *in* PROBLEMS AND PROJECTS 437, 443 (1972).
85. LEWIS, COUNTERFACTUALS, *supra* note 70, at 92.
86. Lewis, *Causation*, *supra* note 82, at 184.
87. *Id.*
88. *See* PAUL HORWICH, ASYMMETRIES IN TIME: PROBLEMS IN THE PHILOSOPHY OF SCIENCE 167 (1987), *reprinted in* CAUSATION, *supra* note 76, at 208, 213.

(26) If the Soviets had pushed the button to unleash a nuclear response to the blockade of Cuba in 1962, the world would have experienced nuclear winter.[89]

What is the set of closest possible worlds in which we should see if the consequent is true? Is it one where the nuclear trigger was pulled but none of the systems worked—on the ground that this is closer to the actual world than one where the systems worked, and as a result, the world as we know it is gone forever? Doesn't Lewis's stipulated criteria of similarity favor maintaining whole swatches of space and time regions over the "small miracle" or two required for a nuclear trigger to fail?[90]

These are serious problems for the possible worlds account of counterfactual conditionals. I raise them, however, not to reject the theory but to clarify it. In fact, in what follows I shall use the possible worlds understanding of counterfactuals as we examine the counterfactual theory of causation. I do this partly because the counterfactual theory of causation is best worked out using the possible worlds conception of counterfactuals.[91] In addition, in the covering law conception, counterfactuals are part and parcel of a larger theory about laws, raising issues extraneous to our present concerns.[92] In any case,

---

89. *Id.* Like examples are provided by Kit Fine, *Critical Notice*, 84 MIND 451, 452 (1975) (reviewing LEWIS, COUNTERFACTUALS, *supra* note 70); *see also* Jonathan Bennett, *Counterfactuals and Possible Worlds*, 4 CANADIAN J. PHIL. 381, 395 (1974). The general form of such counterfactuals is where a small difference in what actually occurred would make a very big difference to the world.

90. It may seem that we do not have to take the changes from the actual world described in the consequent into account when judging whether that consequent would be true in the sufficiently similar possible world. Rather, the thought might be, we are testing the consequent in a possible world constructed from changes worked by the antecedent clause alone. The problem with this thought is that the states of affairs described in the antecedent in (26) are connected by causal laws to the many other states of affairs (nuclear winter, etc.) which make the world very different. Why wouldn't a minimally different possible world be one where such laws are suspended but no nuclear winter takes place? David Lewis invites precisely this tradeoff when he reformulates his test for counterfactuals this way: "In other words, a counterfactual is non-vacuously true iff it takes less of a departure from actuality to make the consequent true along with the antecedent than it does to make the antecedent true without the consequent." Lewis, *Causation*, *supra* note 82, at 184–85.

91. *Id.*

92. Counterfactuals, under the covering law view of them, are something of a mere tail on the dog insofar as causation is concerned. On the neo-Humean theory of causation, for example, *see* ARMSTRONG, *supra* note 50, MACKIE, *supra* note 77, Wright, *supra* note 19, singular causal statements are reduced to a combination of causal generalizations (laws) together with space-time locaters. Such laws are then analyzed in terms of sufficient conditions for the happening of certain events ("effects"). Counterfactuals enter because they are an implication of the idea of laws and sufficient conditions. Some neo-Humeans, such as Wright and Mackie, find a slightly larger role

much of what I shall say (in possible worlds talk) about counterfactuals and their relation to causation and responsibility can be applied mutatis mutandis to the covering law version of counterfactuals.

III. THE COUNTERFACTUAL THEORY OF CAUSATION

A. *The Nature of the Counterfactual Theory of Causation*

The counterfactual theory of causation consists of two distinct parts. The first is showing that causation can be analyzed in terms of counterfactuals. The second is showing what makes counterfactuals themselves true. We have already addressed the second of these topics. Now, I shall address the first.

The counterfactual theory of causation is one of a family of theories of causation that may broadly be described as Humean. The label is appropriate because of the following characterization of causation by Hume: "[W]e may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed."[93] In quoting this passage, David Lewis urged that "Hume defined causation twice over."[94] Whether this is so depends on which theory of counterfactuals one adopts.

On the covering law view of counterfactuals, Hume defined "causation" but once. A singular causal statement such as, "*a* caused *b*," is analyzed in terms of the statements, "*a* occurred," "*b* occurred," "*b* did not occur prior to *a*," and "*a* is an instance of some kind of event *A*, *b* is an instance of some kind of event *B*, where there is a covering law of the form, $(x)(Ax \supset Bx)$."[95] Because, under the covering law view, the truth of the *A/B* law requires the truth of the counterfactual, "If *a* had not occurred, *b* would not have occurred," Hume's purportedly second definition of cause would, in reality, be no more than a restatement of his first.

---

for counterfactuals in this theory of causation, by adding that a cause must be a necessary element in the set of sufficient conditions that together form the antecedent of causal laws. Still, the main thrust of the neo-Humean theory of causation is nomic sufficiency, and the main problems for the theory stem from that dominant aspect of it.

93. DAVID HUME, AN INQUIRY CONCERNING HUMAN UNDERSTANDING 87 (Charles W. Hendel ed., Bobbs-Merrill Co. 1955) (1748) (emphases omitted).

94. Lewis, *Causation*, *supra* note 82, at 556.

95. On the reductionist program of Hume, see Moore, *Causation and Responsibility*, *supra* note *, at 45–46.

Construing the counterfactual theory of causation in this way makes it less interesting because then the counterfactual implications of causal statements are but an aspect of the causal laws to which singular causal statements are to be reduced. These laws can be variously analyzed as mere regularities of nature (standard issue Hume), relations between universals (neo-Humeans), or relations of conditional probability (the probabilistic Hume). The problems that beset any of these reduction bases for singular causal statements then predominate in any discussion of the counterfactual theory of causation,[96] and problems unique to counterfactuals fade into the background. To avoid this, I shall focus on the possible worlds construal of counterfactual conditionals and the reduction of causation to these.

On the possible worlds construal of counterfactuals, Lewis is right to assert that Hume's second quoted statement gives a second and independent definition of causation. Again, "if the first object had not been, the second never had existed."[97] This is close to the law's dominant test for cause in fact in torts, criminal law, and contracts: But for the defendant's action, the harm would never have occurred.[98]

David Lewis approximates this legal test for causation through some preliminary moves.[99] The first is to relate a notion Lewis invents, causal dependence, to causation. Causal dependence operates in the reverse direction from causation, so that when y causally depends on x, x causes y. Causal dependence is sufficient for causation, although Lewis denies that it is necessary: There can be causation without causal dependence. Lewis then analyzes causal dependence in terms of counterfactual dependence, so that y causally depends on x if y counterfactually depends on x (with counterfactual dependence then being spelled out in the possible worlds manner before described). This allows Lewis to say that counterfactual dependence is sufficient for causation, although it is not necessary.

Lewis then introduces the notion of a causal chain.[100] Suppose x causes y, and y causes z. Lewis staunchly held to the view that the causal relation is transitive, so that in the situation suggested, x necessarily causes z. Lewis recognized that the relation of counterfactual dependence is *not* transitive. This is because the possible world in which one would test the counterfactual "but for x, would z have occurred?" is different from the possible world in which one would test

---

96. *See id.* at 45–48.
97. HUME, *supra* note 93.
98. *See, e.g.*, MODEL PENAL CODE § 2.03(1) (1962).
99. Lewis, *Causation*, *supra* note 82, at 185–87.
100. *Id*. at 187.

the linking counterfactuals, "but for *x*, would *y* have occurred?" and "but for *y* would *z* have occurred?" Thus, under Lewis's view of possible worlds, it may be that in the closest possible world in which *x* does not occur, *z* does occur. This could be the case because *z* could be caused other than by *y*. In this case, *z* does not counterfactually depend on *x* even though, on Lewis's transitive view of the causal relation, *x* causes *z*. So Lewis does not require for the truth of "*x* causes *z*" that there be direct counterfactual dependence of *z* on *x*. Rather, *z* causally depends on *x* if within each pair of events (*x*, *y*) and (*y*, *z*) there is counterfactual dependence. That is, so long as *z* counterfactually depends on *y*, and *y* counterfactually depends on *x*, then *z* causally depends on *x*. So it is also sufficient for *x* to cause *z* that there be stepwise (or indirect) counterfactual dependence of *z* on *x*, even though *z* does not counterfactually depend on *x* directly. Finally, Lewis holds that one of these kinds of counterfactual dependence (direct or stepwise) is necessary for causal dependency.

B. The Fit of the Counterfactual Theory of Causation to the Law's Uses of Causal Judgments

The counterfactual theory of causation prima facie fits very well with the ten demands made by the law on the concept of causation outlined earlier. That is, in using causation as one of the basic touchstones of responsibility for misfeasance, legal usages (1) and (2) above, the counterfactual theory seems to isolate the bad states of affairs we cause from those that merely follow upon our actions. On this theory, the ones we cause are, by and large, the ones that would not have occurred without our actions. The counterfactual theory of causation thus reinforces the idea that it is causation that measures the core of our moral and legal obligations not to make the world a worse place. For, on the counterfactual theory, it is causation that measures directly whether we have made the world a worse place by our actions.

Similarly, insofar as the law uses causation to identify those omissions for which we are liable from those which are immune from liability, legal usages (3) and (4) above, the counterfactual theory seems to fit the bill nicely. For counterfactual questions are indifferent to the valence of events. Whether positive or negative, the counterfactual question is equally well formed. We can ask whether a given harm *H* would have occurred if the defendant had not omitted to do some action *A* as easily as we can ask whether *H* would have occurred if the defendant had not done *A*. Insofar as "acts of omission" can be causes no less than "acts of

commission," and insofar as the law imposes liability on omissions that cause harms, the counterfactual theory of causation seems to be just what is needed by way of a theory of causation.

In treating the doing-allowing distinction as a causal distinction, legal usages (5) and (6) above, the law also presupposes the counterfactual theory of causation. For example, in cases of passive euthanasia we are said not to cause death when we turn off the respirator that we had turned on at some earlier point in time. If this is true, causation had better mean "not making the patient worse off than he would have been if we had never turned on the respirator to start with." In order to do the work demanded of it by such uses of the doing-allowing distinction, causation needs to be unpacked counterfactually.

As we have also seen, the law sometimes frames its harm-within-the-risk limitation of liability in causal terms. For there to be tort liability, it is often said that it is not enough for the act that is negligent to be a cause of the harm. Rather, in addition, the negligence itself must cause the harm. What courts mean by this last statement is that the aspect(s) of the action that made it negligent have to stand in a causal relationship to the harm. Not the acts of medical treatment by an unlicensed practitioner, but certain negligent-making aspects of those acts, such as that they were unlicensed acts. Although it is neither obvious nor uncontentious, the best notion of causation able to do the work demanded of it by the law here is the counterfactual notion.[101] This is because other notions of causation tend to take whole events as causal relata, whereas the counterfactual theory can easily raise the type of question (about aspects of events) asked by the law here. One can ask whether the harm would have occurred if the defendant's acts had possessed the property of being licensed acts as easily as one can ask whether the harm would have occurred if the defendant had not done the acts at all. Perhaps one can ask causal questions using facts as causal relata on other theories of causation, but on the counterfactual theory there is no doubt about it.

Although legal liability in the overdetermination cases, and the causal intuitions that underlie such liability, generally raise havoc for the counterfactual theory of causation, the three aspects of legal liability on which we earlier focused, legal usages (8), (9), and (10), fit the counterfactual theory nicely. If the defendant's fire does not cause the harm because it was joined by a fire of natural origin, as in the symmetrical concurrent cause cases in legal usage (8), then causation had better mean counterfactual dependence. After all, it is the truth of the

---

101. *See generally* HELEN STEWARD, THE ONTOLOGY OF MIND 168–69 (1997) (urging that "causal relevance" is counterfactual in nature and that it is only this counterfactual relation that can take facts, not events, as its relata).

counterfactual, that the harm would have happened anyway where and when it did, even without the defendant's fire, that alone supports the causal judgment on which the absence of liability depends. Similarly, in asymmetrical concurrent overdetermination cases, which was legal usage (9), the law's conclusion that smaller potential causes are not causes at all is most plausibly true on the counterfactual theory of causation. It is because the harm would have happened anyway that supports the causal judgment on which the absence of liability depends in such cases. Finally, a defendant can be said not to cause much of the harm in preemptive overdetermination cases, legal usage (10), only on a counterfactual understanding of causation. It is because the harm was about to happen anyway, even if the defendant had not done what he did, that supports the causal judgment on which this limitation on damage depends.

In these ten ways the law thus seemingly presupposes the counterfactual theory of causation. Whether causation is, in fact, the same as counterfactual dependence is the question we shall want to examine next. Just because the law wants it to be so does not make it so, for metaphysics does not answer to legal needs. Indeed, since I shall next argue that the theory is false root and branch as a theory of causation, we will be left in something of a quandary as to what sense to make of these legal doctrines.

C. Causation Is Not Counterfactual Dependence

The counterfactual theory of causation is subject to two sorts of objections. The first set of objections questions the sufficiency of direct or stepwise counterfactual dependence for causation; the second questions the necessity of either kind of counterfactual dependence for causal dependence. I shall consider each set of objections in turn.

1. The Sufficiency of Counterfactual Dependence for Causation

a. The Existence of Noncausal Counterfactuals

The counterfactual theory is met initially by the observation that "counterfactual dependency is too broad to pin down causal dependency."[102] We saw examples of this in considering the covering law account of

---

102. Jaegwon Kim, *Causes and Counterfactuals*, 70 J. PHIL. 570, 570–72 (1973), *reprinted in* CAUSATION, *supra* note 76, at 206.

counterfactuals. In statements (9) through (15) we concluded that there is counterfactual dependence between consequent events and antecedent events, but there are no causal laws connecting such events. We should now add that there is also no causal relationship between such events. In a phrase, counterfactual dependence cannot be sufficient for causation. Something must be added for the account to rid it of these troublesome counterexamples.

   Or so it seems. Just as the covering law theorist may seek to discover hidden laws for all apparent counterexamples to his counterfactual theory, so the counterfactual theorist may seek to discover causal relations in such counterexamples. Consider this mereological counterexample of Jaegwon Kim:

   (27) If I had not written "r" twice in succession, I would not have written "Larry."[103]

Kim fails to "see a causal relation between these events,"[104] but a defender of the counterfactual theory might purport to find a "type of causation involved in mereological generation," what Ernst Sosa calls "material causation."[105]   Indeed, one might seek to accommodate all purported counterexamples (9) through (15) as instances of "a form of necessary causation to be distinguished from its more familiar contingent counterpart."[106]

   I am with Kim here. It is unhelpful to invent special kinds of causation in order to deal with otherwise troublesome counterexamples. That is an old game smacking of the same irrationality that Bentham railed against in his attack on legal fictions.[107] It is all too easy to save a generalization from troublesome counterexamples by changing the meaning of the terms used in the generalization. The strategy is as unjustified here as it is in the hands of other causal dualists like Antony Flew[108] or Herbert Hart and Tony Honore.[109] Nothing can be justified by such ad hoc "discovery" of special senses of terms that do not seem on

---

103.   *Id.*
104.   *Id.*
105.   Ernest Sosa, *Varieties of Causation*, 11 GRAZER PHILOSOPHISCHE STUDIEN 93, 99–100 (1980), *reprinted in* CAUSATION, *supra* note 76, at 240.
106.   *Id.*
107.   C.K. OGDEN, BENTHAM'S THEORY OF FICTIONS 7–12 (1959).
108*.   See generally* Antony Flew, *Psychiatry, Law, and Responsibility*, 35 PHIL. Q. 425 (1985) (reviewing MICHAEL MOORE, LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP (1984) (distinguishing "inclining" or "predisposing" causation from deterministic causation)).
109*.   See generally* H.L.A. HART & TONY HONORE, CAUSATION IN THE LAW 186–204 (2d ed. 1985) (finding there to be two types of weaker causal connection in addition to the "standard case" of causing harm).

1218

the face of their usage to be ambiguous.

It is better to do one's repair work at the other end. Instead of playing with what causation is, redefine the idea of counterfactual dependence. What is wanted is some exclusionary clause keeping out unwanted counterfactuals where what is driving them is an analytic necessity. Although not easy, perhaps the counterfactual theory can be supplemented by some requirement of nonanalytic necessity. If so, the unwanted counterexamples, while being examples of counterfactual dependence tout cour, would not be examples of nonanalytic counterfactual dependence. The danger, of course, is that "nonanalytic necessity" can only be cashed out in causal terms, rendering the counterfactual theory so viciously circular that it says no more than "causation is a causal kind of counterfactual dependence." I leave the issue open,[110] turning to more conclusive objections.

b. The Continued Overbreadth of Even Nonanalytic Counterfactual Dependence

A long recognized problem for the counterfactual theory of causation is the extremely nondiscriminating notion of causation it generates. This is so even though we rule out examples based on analytic necessity. I have earlier called this the "promiscuity" objection to the counterfactual theory.[111]

There are two versions of this objection, the at-a-time version and the over-time version. To illustrate the at-a-time version, reconsider Nelson Goodman's example of a match lighting because it was struck. As we saw, the causal law involved in such examples includes more than striking the match in the conditions necessary for the match to light. In addition, the match needs to be dry, well-made, and struck in the presence of oxygen. On the counterfactual theory of causation, these equally necessary conditions have equal title with the striking of the match to be called the cause of the match lighting. Some find it absurd to think that the cause of the match lighting is as much due to the presence of oxygen, or the dryness of the match, as it is the striking of the match.

Lewis joins John Stuart Mill in not finding this absurd at all. On

---

110. For further explanation of issue left open, see T. Yagisawa, *Counterfactual Analysis of Causation and Kim's Examples*, 39 ANALYSIS 100, 101 (1979).
111. MOORE, *supra* note 13, at 269.

Mill's view, the "whole cause" includes all of these equally necessary conditions.[112]   It is only pragmatic features of causal utterances that single out, say, the striking as "the cause" of the lighting.[113]   The striking is something we can perhaps control better, or it is the least expected and thus most surprising feature, or it is blameworthy because a human action, and so on.   Yet none of these pragmatic features of causal utterance go to the nature of the causal relationship itself.   That, Mill said, is a relation holding equally between each of these conditions and their common effect.   Indeed, as Lewis explicitly says, "[M]y analysis is meant to capture a broad and non-discriminatory concept of causation."[114]

Lewis and Mill are on solid ground in refusing to grant semantic status to the discriminatory principles by which we select amongst the various causes of the match lighting and thereby honoring one as "the cause." These are merely pragmatic features of causal utterances, dependant on the context of the utterance and the speaker's intent for their justifications.   As we recognized earlier with respect to the covering law account of counterfactuals, they are features that are irrelevant to causation itself.

More troublesome for the counterfactual theorist is the over-time version of the objection.   As we have seen, on Lewis's view causation is a fully transitive relation, so that if e counterfactually depends on d (so d causes e), and d counterfactually depends on c (so c caused d), then c causes e.[115]   This means that the number of causes for any given event e is staggeringly large, in each case reaching back to the big bang that apparently began this whole show.   This *does* strike me as absurd.

Our ordinary, and legal, conception of causation avoids this absurdity through three sorts of judgments, those of proximateness, those of breaks in causal chains, and those of coincidence.[116]   Caesar's crossing

112.   JOHN STUART MILL, A SYSTEM OF LOGIC: RATIOCINATIVE AND INDUCTIVE bk. 3, ch. 5, § 3, at 214–18 (Longman Group Ltd. 1970) (1843).

113.   Joel Feinberg nicely taxonomizes some of these pragmatic features in JOEL FEINBERG, DOING AND DESERVING 143–47 (1970).

114.   Lewis, *Causation*, *supra* note 82, at 183.

115.   However, this is not to say that e counterfactually depends on c. Counterfactual dependence is not transitive because the possible worlds in which one tests for e supposing the absence of c differ from the possible worlds in which one tests for e supposing the absence of d, and for d supposing the absence of c.   *See* WOODS, *supra* note 32, at 44.   Still, the *chains* of counterfactual dependence generate a fully transitive notion of causation, and that is the problem here.

116.   I discuss one of these in detail in Michael S. Moore, *The Metaphysics of Causal Intervention*, 88 CAL. L. REV. 827 (2000).   All these are discussed in Moore, *Causation and Responsibility*, *supra* note *, and in Hurd & Moore, *supra* note *.   My general take on these limitations on causation is that they are all trying to separate instances of de minimus causal contributions (by the remote, the preintervening, and the coincidental events) from instances of causal contributions substantial enough to be causes.   This allows the causal relation to be considered transitive if one counts de

of the Rubicon may well be a necessary condition for your loss of business on a given occasion, but that event will compete poorly with another, more proximate event, such as my breach of contract by revealing trade secrets, as the cause of that loss. The sheer degree of temporal removal disqualifies Caesar's act as a cause. Likewise, my setting steak knives for dinner may be a necessary condition for Jones's death, but it is cut off from being a cause of that death by a dinner guest picking up the steak knife and plunging it into Jones's heart. The intentional exploitation of the situation by the dinner guest cuts off my earlier act as a potential cause, necessary though it was. Likewise, negligently speeding in my car may well be necessary to my being just where and when a child darts out into the street, and thus, also a necessary condition of the child's being hit. Yet speeding, and like events, such as stopping for lunch, are only coincidentally connected to the child's injuries. They are not causes of the injury.

Lewis no doubt would wish to relegate the discriminatory principles by which we make such judgments to the pragmatics of causal utterances, just as he did for the principles involved in the at-a-time version of the objection.[117] Yet that move is implausible here. What makes the move so plausible versus the at-a-time version of the objection is that we can imagine contexts of utterance where the dryness of the match or the presence of oxygen might well be appropriate

---

minimus causal contributors to be in some sense causes. *See* Ned Hall, *Causation and the Price of Transitivity*, 97 J. PHIL. 198, 203 (2000).

117. Lewis takes this tack in *Causation as Influence*, *supra* note 79, at 196. To the suggestion that a cause of each of our deaths is our births, Lewis urges that even though this is indeed the case "it is understandable that we seldom say so. The counterfactual dependence of his death on his birth is just too obvious to be worth mentioning." *Id.* Often, Lewis takes another tack, urging that there is nothing odd in attributing causation in the cases mentioned in the text, in which event he has no need for discriminatory principles, be they semantic or pragmatic. *See id.* at 195 (urging that historians trace causation "over intervals of any length" and conclude "without more ado" that the beginning of the chain *caused* the end). Lewis earlier imagined such extended causal chains as, for example, linking my writing a letter of recommendation to every death that occurs because of it, concluding that "it would be strange to single out my acts as *the* cause of all those deaths. But it is *a* cause of them, under my analysis and also according to our common usage." DAVID LEWIS, *Postscripts to 'Causation,' in* 2 PHILOSOPHICAL PAPERS 172, 184 (1986). Lewis goes on to adopt such discriminatory principles as part of the semantics of causative action verbs like "killing." *Id.* According to Lewis's analysis, I thus can cause death but not kill in these situations of extended causal chains. *Id.* My own view is that there is an equivalence between sentences using such causative action verbs and sentences using causation and that therefore the discriminatory principles that form a part of the former are also a part of the latter. MOORE, *supra* note 13, at 227.

answers to the question, "What caused the match to light?" For example, when there was a fire in an Apollo test capsule years ago, the presence of oxygen was cited as a cause of the fire. This was appropriate because that presence was surprising, given the oxygen-free environment usually required in such settings. There seems no analogous context of utterance making Caesar's crossing the Rubicon, my setting of steak knives, or my negligent speeding appropriate answers to the question of what caused the loss of business, the death of Jones, or the injury to the child, respectively.

The many lawyers who relegate problems of proximateness, intervention, and coincidence to noncausal matters of "policy" are making essentially the same kind of defensive move as the move to pragmatics by David Lewis. The common lawyerly thought here is to concede that causation, on the counterfactual theory of it, is very promiscuous, but then to urge noncausal limitations on both moral responsibility and legal liability. These are the supposed policy limitations on liability. Yet what are these policies? Suppose one has a corrective justice theory of tort and contract law under which one owes a legal and moral duty to correct the harm one has caused. If Caesar really did cause your loss of business, why is his estate not liable for that loss? What policy, in general, would justify leaving losses where they fall even when a culpable harm-causer causes harm to an innocent harm-sufferer? To my mind, the policy arguments offered here, in terms of the inefficiency, unfairness, or unforeseeably large liability, are so poor as to be but substitutes for what we actually believe on other grounds. No one really thinks that Caesar caused your losses, no matter how clearly his acts may have been necessary for them.

The upshot is that counterfactual dependence, by itself, cannot account for features causation seems to possess. Counterfactual dependence across chains of causes does not, as causation seems to, weaken or peter out. Also, it does not break off suddenly, as causation seems to, and it seems to be robustly present even in coincidences. One might supplement the counterfactual account with some other aspects of causation. However, for this to be anything other than ad hoc, such supplementary aspects had better flow from the supposed counterfactual nature of causation. No such showing comes to mind as at all plausible.

c. More Overbreadth of the Counterfactual Theory: Omissions as Causes

Like the three generalist theories of causation (the Humean, neo-Humean, and probabilistic theories), the counterfactual theory is blind to the difference between acts and omissions, and more generally, to the

1222

difference between actual and absent events. There can be counterfactual dependence of a harm on my omission to do some act preventing that harm as easily as there can be counterfactual dependence of a harm on some positive action of mine. For example, "Had I not omitted to throw you the rope—had I thrown you the rope—you would not have drowned," is a perfectly well formed counterfactual. As good as the counterfactual, "Had I not pushed you in, you would not have drowned." More generally, there can be counterfactual dependence of some event on the absence, as well as on the presence, of certain other events. In this blindness, counterfactual dependence is different than causation.

At least, that is how it seems to me. In truth, many philosophers and legal theorists are receptive to the idea that omissions, and other "negative events," can be causes, in which case counterfactual dependence is not overbroad in this dimension. Yet this is a temptation to be resisted. The temptation to regard omissions as causes stems from a number of confusions. One is about the nature of omissions themselves. The thought is that omissions are a kind of particular event, a negative event. On this view there are acts of omission just as there are acts of commission. So there is nothing odd about either kind of event or action standing in causal relationships.

Yet omissions are not events or actions of any kind. My omitting to throw you the rope at time *t* is not an intentional choice by me not to throw the rope, nor is it any other kind of event, mental or physical. Nor is such an omission the state of my remaining motionless. I can omit to throw you the rope by dancing a jig, throwing a party, and so on, as well as by standing stock still. My omission to throw the rope at *t* is just the absence at *t* of any acts by me of the type, throwing you a rope. An omission is thus like an absent elephant—not a ghostly rope-throwing or ghostly elephant, but simply the absence of the relevant type of thing.[118]

Once one fully appreciates that there is no such thing as a negative event as a particular, that there are only uninstantiated *types* of events, most of the temptation to regard omissions as causes should evaporate. As Julie Andrews said in *The Sound of Music*, "Nothing comes from nothing, and nothing ever can."[119] Absent elephants trample no grass, and absent rope-throwings drown no one.

For those who persist, there is a second confusion motivating them.

---

118. I argue for this conception of an omission in MOORE, *supra* note 13, at 24–31, 276–77.
119. THE SOUND OF MUSIC (20th Century Fox 1965).

This is the confusion of singular causal relations, on the one hand, with causal generalizations and the explanations made possible by them, on the other. The statement, "His act of pushing *V* caused *V* to drown," describes a singular causal relationship existing between two event-tokens, my act of pushing and the death of *V* by drowning. By contrast, the statement, "Pushing those who cannot swim into deep water causes them to drown," describes a causal law; its relata are types of events, not tokens, and it describes a general, not singular, relationship between these universals.

It is well-known that explanations framed in terms of causal generalizations operate differently than do explanations framed in terms of singular causal statements. One difference is the presence of negative conditions in causal laws. Just as matches need not to be wet in order to light, so victims in the water need not to be thrown a rope if they are to drown. Roughly, the causal law is:

(28) ~ (*V* pushed into water • water deep • no rope thrown within reach of
      V • *V* cannot swim • *V* not drown).

Either *V* drowns or one of the other conditions is not fulfilled, and these conditions may as easily be negative as positive.

It is perfectly idiomatic to use such a generalization to explain why some particular victim drowned. "He wasn't thrown the rope on the dock" is as acceptable as "he couldn't swim," or "he was pushed in." Yet such appropriate use of omissions in explanations goes no distance towards establishing that some omission, taken singularly, caused *V* to drown—it didn't.

A third confusion making it difficult to see the obvious here is the confusion between facts and events.[120] On the Davidsonian view[121] that I have defended elsewhere,[122] an event is a particular consisting of a change in an object over an interval of time. A fact is an actual state of affairs, such as the possession by an event of a certain property or the absence of a certain type of event at a certain time. The killing of Smith by Jones yesterday is an event, whereas the fact that it was a particularly cruel killing, and the fact that Jones did not kill Smith quickly, are facts about that event. This distinction muddies the waters about omissions as causes in the following way: Statements using facts about omitted events to explain other events are perfectly idiomatic. For instance, "Smith

---

120. *See generally* Jonathan Bennett, Events and Their Names (1988) (clearly describing the distinction between facts and events). Bennett incidentally comes out the other way on facts, not events, being causal relata.

121. Donald Davidson, Essays on Actions and Events 135 (1980).

122. Moore, *supra* note 13.

1224

drowned because of the fact that James failed to throw him a rope." If one takes such facts to be the relata of singular causal relations, then the idea that omissions can be causes will seem unproblematic.

My own view is that the relata of singular causal relations are whole events, not those aspects of events we call facts. On this view, the perfectly idiomatic fact statements like the one quoted above are elliptically stated generalizations; the aspect of an event they pick out is the property making for a true causal law. It is thus no surprise that fact-based singular statements, like statements of causal laws, make use of omissions in explanations. The appropriateness of omissions in such explanations, however, goes no distance towards showing that omissions can be causes, that is, that omissions can stand in singular causal relations.

The fourth and last confusion is about the distinction discussed earlier, between doings and causings on the one hand, and mere allowings on the other.[123] Consider an earlier example of mine,[124] where some victim *V* is drowning, some rescuer *R* is about to throw *V* the rope that will surely save *V*, and defendant *D* tackles *R*, successfully preventing the rope from being thrown. "As a result," we might say, *V* drowns. Counterfactualists like Lewis argue that we must treat omissions as causes, on pain of the absurd conclusion that individuals like *D* do not cause the death of victims like *V*. How could they, the argument goes, if *D*'s causation of *V*'s death is only by virtue of causation through *R*'s omission?[125] Yet we can swallow the conclusion of *D* not causing *V*'s death, without absurdity, if we see two things. First, that *D*'s act of removing the only thing that could save *V* merely *allowed V* to die. Second, not all allowings lessen one's responsibility for the harm allowed. On the latter point, as I shall argue later,[126] whether *D* has some responsibility for *V*'s death depends on what we think the acceptable baseline is. If the baseline is *V* drowning with no rescue

---

123. *See supra* text accompanying notes 10–11.
124. MOORE, *supra* note 13, at 278 n.42.
125. *See* Lewis, *supra* note 79, at 196. John Collins calls such cases "causation by double prevention," concluding with Lewis that there must be causation by omissions in such cases. John Collins, *Preemptive Prevention*, 97 J. PHIL. 223, 226 (2000). To like effect, see also Michael McDermott, *Redundant Causation*, 46 BRITISH J. PHIL. SCI. 523, 529 (1995). Compare the extremely detailed analysis of Ned Hall, who concludes of such cases of "double prevention" that there is no causal connection because there is no causal process between *D*'s act and *V*'s death. Hall, *supra* note 116.
126. *See infra* text accompanying notes 196–99.

about to be undertaken, then D's act has made no difference. However, if the baseline is V being rescued by R, as seems plausible, then D has made V much worse off and is morally, but not causally, responsible for doing so. The trick is to see that that moral conclusion does not depend on granting causal efficacy to D's acts, and, thus, to R's omissions. Allowings for which we are morally blameworthy are just that, allowings, not to be confused with some supposed causing-by-omissions.

This lack of causal potency by omissions should be welcomed by anyone sharing the moral intuition that began this Article. That intuition was that the hard core of our most stringent moral and legal obligations is not to make the world worse, and that while we may have some obligation to make it better, such positive obligations are occasional and less stringent than their negative counterparts. If we adopt a theory of causation that is blind to the difference between acts and omissions, so that either can equally well be the cause of evils, how then are we to draw this distinction between negative and positive obligations? If my failing to prevent your death is just one way of causing it, no different in terms of causation from my shooting you, how can it be that I have a stringent obligation not to shoot you but only a less stringent, or no obligation to prevent someone else from shooting you? It would thus be no easier on our morality than on our metaphysics if causation were conceived so that omissions and other absent events could be causes.

Yet I see no escape route for the counterfactual theorist, other than to embrace omissions as causes.[127] His only other strategy would be to deny that there can be counterfactual dependence of any event on an omission. But how can he deny the obvious good sense of counterfactual conditional statements asserting precisely such a dependence? For instance, "If I had thrown the rope to Jones, Jones would not have drowned," is far too acceptable to be denied sense. I thus count it as a heavy mark against the counterfactual theory of causation that it cannot distinguish acts from omissions on causal grounds.

---

127. This is where Lewis appeared to rest, concluding that "it is not to be denied that there is causation by omission." LEWIS, *supra* note 117, at 191. Lewis distinguished three conceptions of omissions, one of which is "that there are no events of omission" and that therefore the relevant counterfactuals are about types of actions, not tokens. *Id.* at 191–92. Lewis later appeared to adopt this (correct) view of omissions, seeing clearly that "absences are not events" and that "absences are bogus entities." Despite this clearheaded avoidance of the first confusion referenced earlier in the text, Lewis nonetheless was driven by his analysis to the conclusion that "absences can be causes." Lewis, *supra* note 79, at 195.

d. *Still More Overbreadth: The Counterfactual Dependence Between Epiphenomenal Events*

The epiphenomenal objection is an old one, first developed against generalist theories of causation. Suppose some event c causes e at t_1, then c causes f at t_2, yet there is no causal relationship between e and f, which are merely epiphenomenal of one another. For example, I run in the morning with my dog; this quickly lowers my anxiety; it eventually makes my dog tired; lowering my anxiety did not tire my dog.

On the regularity theory of causation often attributed to Hume, events similar to e can regularly follow events similar to c, and events similar to f can regularly follow events similar to c. Unfortunately, events similar to f also can regularly follow events similar to e. Therefore, on the regularity theory, e *does* cause f, contrary to our supposition. Similarly, on the counterfactual theory, c causing f means that c was necessary to f. Suppose that c causing e on a given occasion means not only that c was necessary to e but also that, given the laws and circumstances other than c, c could not have failed to cause e. That means that c, together with the other circumstances, was sufficient for e. This means that e was necessary for c. But that means that e was also necessary for f—if e had not occurred, then c would not have occurred, and if c had not occurred, then f would not have occurred. Thus, on the counterfactual theory of causation, e did cause f, an embarrassing conclusion for that theory.

Lewis seeks to avoid this unwanted conclusion by getting rid of the sufficiency of c for e in the possible world(s) in which we are testing the counterfactual, "If e had not occurred, c would not have occurred." We rid ourselves of such sufficiency by giving up one or more of the other circumstances making c sufficient for e, or by giving up the law making c and these circumstances sufficient for e. Lewis argues that the possible world in which c occurs but e does not is closer to the actual world than is any possible world in which c occurs and e then occurs.[128] Lewis thus seeks to distinguish his account from generalist theories of causation and from the covering law account of counterfactuals. The regularity account of causation, for example, is stuck with there being a causal law, that is, regularity, between c-type events and e-type events.

---

128. Lewis, *Causation*, *supra* note 82, at 190.

Therefore, the troublesome sufficiency for *e* by *c* cannot be avoided, making *e* seemingly the cause of *f*. Lewis's possible worlds account of counterfactuals frees him from holding the *C/E* law invariant in his closest possible world. As Lewis writes, "the laws are not sacred."[129] Better to posit a minor miracle after *c* has occurred so that *e* does not occur, than to allow *f* to depend on *e* by stepwise counterfactual dependence.

The usual rejoinder to Lewis here is that he has gerrymandered his idea of "closest possible world" so as to get around the problem of epiphenomena. This may seem like a repeat of the "ad hoc" charge to Lewis's manipulation of "similarity" so as to get intuitively acceptable truth values for counterfactuals. As we saw there, Lewis was entitled to respond that he is free to construct a notion of similarity that squares with our pretheoretical intuitions about counterfactuals.

Here the charge is different, rendering that kind of answer much more problematic for Lewis. Notice that there is nothing counterintuitive about the counterfactuals "if *e* had not occurred, *c* would not have occurred" and "if *c* had not occurred, *f* would not have occurred" both being true. All we need is that *f* not occur in a possible world very much like the actual world, save that neither *e* nor *c* occurs. Lewis can stipulate that another possible world, one where *c* does not occur but *e* does, is closer to the actual world, because *e* occurs in both. However, what motivates the favoring of the retention of an event (*e*) over the retention of the *C/E* law? Why is Lewis's possible world closer to the actual world than my imagined possible world? Not because the earlier quoted counterfactuals are implausible by themselves, but only because they yield the wrong causal conclusion. Lewis may gerrymander his idea of closeness in order to get plausible counterfactuals, but surely he may not further gerrymander possible worlds in a way that eliminates independently plausible counterfactuals just to make the causal conclusion come out right. Such a secondary gerrymander of possible worlds would render the counterfactual analysis of causation vacuous. Indeed, one might as well define a cause as a cause and be done with it.

Lewis would, no doubt, deny that he is doubly gerrymandering his notion of similarity. Yet that is far from obvious. Recall that, as a criterion of similarity, third in importance was a violation of causal laws and fourth in importance, being of "little significance," was preservation of isolated events. Prima facie, that would favor retaining the *C/E* law over retaining the *c* event as we select a possible world consistent with *e*'s absence. Lewis's apparent view is that some law(s) will have to be violated anyway, so why not violate the *C/E* law, thus allowing retention

---

129. *Id.* at 191.

of *c*? Yet why is the amount of law violation a constant here? True enough, some law(s) will have to be violated when we select the possible world where *e* does not occur. Yet, why choose the *C/E* law, rather than the laws connecting *c* to some earlier events? Because, Lewis replies, later violations are to be preferred to earlier ones:

> To get rid of an actual event e with the least over-all departure from actuality, it will normally be best not to diverge at all from the actual course of events until just before the time of e. The longer we wait, the more we prolong the spatio-temporal region of perfect match between our actual world and the selected alternative. Why diverge sooner rather than later?[130]

Yet perhaps the *C/E* law is a basic law, so that suspending it is a bigger deal than suspending the law(s) that connect *c* to earlier events. If so, then it is not at all apparent that we should wait until the last possible moment to have a law violation.

At this point, one is entitled to feeling a bit at a loss. The vagaries in the notion of similarity between possible worlds are such that it is difficult to tell whether Lewis is smuggling causal conclusions into his selection of possible worlds. Examples such as this one certainly give the appearance of such illicit smuggling. And we shall see further examples of this later on, as we examine Lewis's response to other objections.

Perhaps we can bring the ad hoc nature of Lewis's constructions here into focus by imagining two contrasting scenarios. The first is the one with which we have been working: I run in the morning (*c*), this causes my anxiety to drop (*e*), and it (*c*) also causes (later) my dog to become tired (*f*). In the second scenario, my dog stays home, I run in the morning (*c*), and this causes my anxiety to drop (*e*). In the first scenario, Lewis, as we have seen, tests the supposed causal dependency of *f* on *e* by testing two counterfactual dependencies: "if not *e*, not *c*" and "if not *c*, not *f*." The first of these links he holds to be false because it is a closer possible world where *c* occurs (I run) and *e* does not (my anxiety does not go down) than one in which neither *c* nor *e* occurs (I neither run nor have decreased anxiety). Would Lewis say the same in the second scenario, where my dog stays home? That is, by hypothesis, *c* causes *e*, and does so in a way that given *c*, together with other circumstances, *e* had to occur. Now when we test the counterfactual "if not *e*, not *c*," is the closest world one that suspends the *C/E* law (in which event the counterfactual is false—*c* exists in such a possible world even without

---

130. *Id.* at 190.

e)? Or is the closest possible world one in which the law holds, and there is no *c* and no *e* (in which event the counterfactual is true)? If Lewis takes the latter option, then the fully ad hoc nature of these selections of possible worlds is naked for all to see.

Lewis has an alternative answer to the problem of epiphenomena, one that he in fact deploys, although not with any clear separation from the closeness machinations above described. This answer denies that the counterfactual, "if not *e*, not *c*," follows from the sufficiency of *c*, and other circumstances, for *e*. This answer denies that counterfactual dependence can backtrack through time, so that an earlier event (*c*) can counterfactually depend on a later event (*e*). This Lewis has to deny anyway in order to defuse the next objection. Thus, I shall consider this response in conjunction with a like response to the next objection.

e. Overbreadth Concluded: Backtracking Counterfactuals and the Temporal Asymmetry of Causation

Two of our bedrock intuitions about causation are that it is asymmetrical (if *c* causes *e*, *e* does not cause *c*) and that it has a one-way direction in time (causes must not succeed their effects in time). An embarrassment for the counterfactual theory of causation is that, at least prima facie, counterfactual dependence does not seem to be bound by these limitations.[131] If some cause *c* is not only necessary but sufficient for some effect *e*, then *e* seems equally sufficient and necessary for *c*. To be at all plausible, the counterfactual theory has to deny this possibility.

It must issue such denial, moreover, as something more than an ad hoc add-on to counterfactual dependence. That is, one cannot plausibly modify the counterfactual theory by saying that causation is counterfactual dependence plus the temporal condition that causes neither succeed their effects nor can they simultaneously be an effect of their effects. To be plausible, these additional restrictions have to be shown to follow from the counterfactual nature of causation, together with other plausible suppositions about the world.

David Lewis's initial answer to this problem was of a piece with his answer to the epiphenomenal problem. By managing the four criteria of closeness with care, he denied the truth of backtracking counterfactuals.[132] Later, however, he supplemented this showing with an argument purporting to show why, in general, the closest possible worlds in which

---

131. This is a problem noticed by P.B. Downing. *See generally* P.B. Downing, *Subjunctive Conditionals, Time Order, and Causation*, 59 PROCEEDINGS OF THE ARISTOTELIAN SOCIETY 125 (1959).

132. Lewis, *Causation*, *supra* note 82, at 190.

we should test backtracking counterfactuals render them all false. I shall now explore this argument.

The conclusion of Lewis's argument here might be taken to be a kind of deliverance of common sense:

> The literal truth is just that the future depends counterfactually on the present. It depends, partly, on what we do now.
> . . . The past would be the same, however we acted now. The past does not at all depend on what we do now. It is counterfactually independent of the present.[133]

This bit of common sense is not, and cannot, be the basis for Lewis's argument here. For the persuasive power of the truism that you cannot change the past is already loaded with the asymmetry of the causal relation: It is precisely because causation does not work backwards through time that we cannot change the past. It is in that causal sense of "depend" that the future *depends* on the past in a way that the past does not *depend* on the future. Lewis cannot just help himself to this cause-based deliverance of common sense. Rather, it is this causal conclusion that has to be matched by a like conclusion using "depend" in Lewis's counterfactual sense. That is what needs showing: that the past cannot counterfactually depend on the future.

Lewis begins the argument with an empirical hypothesis that the causal structure of the world is such that there is an asymmetry of overdetermination. An overdetermination case is one where there are two or more sets of conditions sufficient for the happening of some event. If two independent fires are headed for the same property, and either is sufficient (when conjoined with other circumstances) for the destruction of the property, the destruction of the property is said to be overdetermined. While Lewis acknowledges the existence of such cases of causal overdetermination (where the past overdetermines the future) he thinks the converse is much more common: The future typically overdetermines the past.[134]

The picture is that of a cone representing sufficient causes, each with many effects, but without overlap in the joint production of effects. Think of the proverbial stone dropped into a quiet pond, sufficient for many ripples and other effects (see Figure 30 below). Those effects are not typically overdetermined. That is, absent this stone being dropped,

---

133. Lewis, *Counterfactual Dependence and Time's Arrow*, *supra* note 82, at 455–76.
134. *Id.* at 49–51.

there would have been no such ripples, for there were no ducks (or anything else) about to land in the pond. Furthermore, given each ripple, e_1, e_2, e_3 ... e_n, there had to have been a stone dropped, for the ripples would have come about in no other way. Each effect, in other words, is sufficient for the common event that was its cause.

It is important to see that Lewis need not deny the common picture of causation as an *inverted* cone, where it takes many causes to produce a single effect. The picture is:

(29)
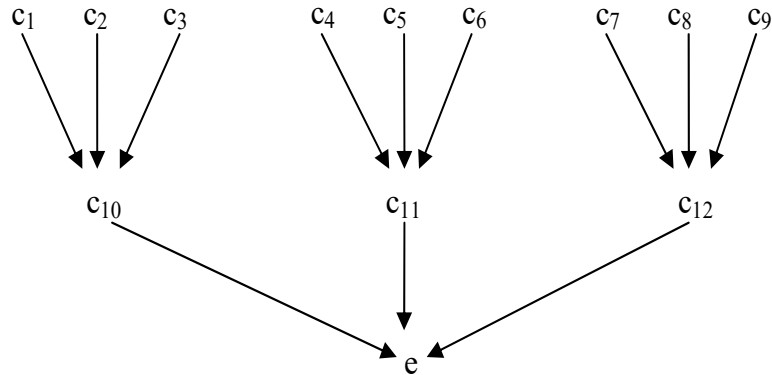
On the counterfactual conception of causation, there is no overdetermination by causes in this picture. There is one sufficient set for e (c_{10}, c_{11}, c_{12}), each member of which is necessary for e, and one sufficient set for each of the member of that set, and so on back through time.

Nor is Lewis's overdetermination asymmetry thesis incompatible with the common view that every event has many effects, pictured as a cone of causation.

(30)

1232
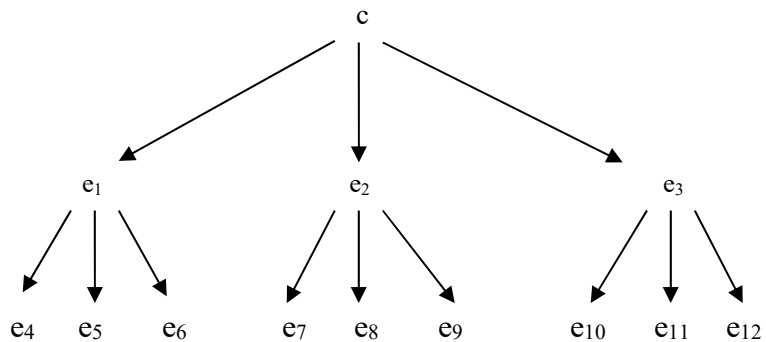
This is just the stone (c) causing close-in ripples (e_1, e_2, e_3), which in turn causes further and wider ripples (e_4 through e_{12}). Yet neither is Lewis's claim to be accepted just because one accepts this last view. Neither of these claims supports or refutes Lewis's claim here, which deals with an asymmetry of *sufficient* conditions.

Lewis's striking claim can be pictured by taking the arrows in Figures (29) and (31) to represent sufficient conditions rather than necessary conditions; the claim is that the world looks more like Figure (31) than like Figure (29):

(31)
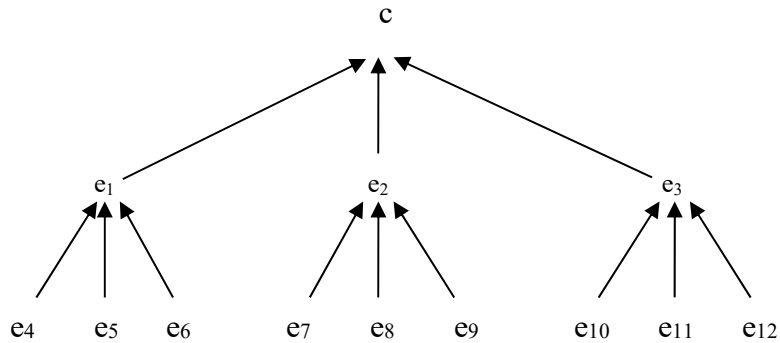
Symmetry of overdetermination might sometimes be pictured as:

(32)
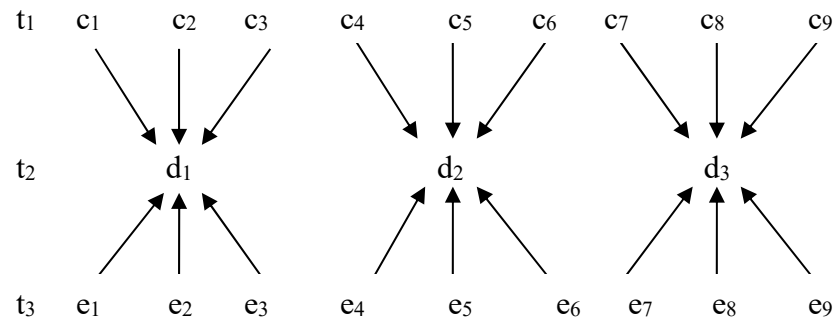
where causal overdetermination during the interval t_1–t_2 is matched by effect overdetermination during the interval t_2–t_3. Lewis's further claim is that effects are sufficient for their causes. These two claims result in Lewis's conclusion that there are numerous effects for any given cause, each of which is sufficient for that cause, while for each such effect there is usually but one sufficient cause.

If there is an asymmetry of overdetermination in the world, then the second step of Lewis's argument against backtracking counterfactuals is to use that asymmetry to argue for the typical falsity of backtracking counterfactuals. Such counterfactuals are usually false because in the possible world(s) in which they could be true, we have to allow for multiple violations of causal laws. This, because when we test the counterfactual:

(33) If the ripple had not occurred, the stone would not have been dropped in the pond,

we must find a possible world in which the stone is not dropped. However, for such a world to exist, many links of nomic sufficiency will have to be broken, namely, the sufficiency of each effect from that common cause. There will be possible world(s) closer to the actual world where most of those links are preserved, so that the stone is still dropped; and in those possible worlds, backtracking counterfactuals like (33) come out false. Overdetermination thus feeds directly into Lewis's criteria of closeness of possible worlds. As Lewis puts it, "The more overdetermination, the more links need breaking and the more widespread and diverse must be a miracle if it is to break them all."[135]

Before questioning Lewis's account, we should pause to praise it. Not only is it ingenuous, but it is also the kind of account needed here. Notice there is no ad hoc stipulation just tacked onto counterfactual dependence so as to artificially inject temporal asymmetry into the account. Rather, it is the alleged nature of causation—counterfactual dependence—that is used to explain the asymmetry of causation through time, together with an empirical assumption.

Still, there is room to doubt either step of Lewis's argument here. With regard to the first step, the question is whether the empirical assumption is true. Recalling that there are two parts to this step, one might first question why we should think, in general, that effects are sufficient for their causes. Paul Horwich, for example, urges that "Lewis does not . . . give grounds for a thesis of such generality, and I see no

---

135. *Id.* at 50.

reason to accept it."[136] Yet on standard logic, if *c* is necessary for *e*, then *e* is sufficient for *c*. If the counterfactual theory of causation is true, then every effect is sufficient for its cause. It may seem that Lewis cannot assume the truth of his theory in order to argue for it; yet notice that the asymmetry argument is a kind of reductio argument against the counterfactual theory. Assume the counterfactual theory is true; if so, then it has the absurd consequence that effects cause their causes. Lewis is thus here entitled to assume the truth of his theory in order to show that it does not generate this absurd consequence.

If, on the counterfactual theory, effects are generally sufficient for their causes, then one is still left with the other part of the first step of the argument, that is, with the counting question of whether causes have multiple effects (each of which is sufficient for its cause) more often than effects have multiple *sufficient* causes. Notice that this is not the question of whether there are more causes than there are effects; it is rather the question of whether there are more *sufficient* causes than there are effects.

My own sense is with Lewis on this one. It is plausible that every cause has multiple effects and equally plausible that every effect has multiple causes. Yet on the counterfactual theory of causation each of these multiple effects is sufficient for its cause, whereas, on the same theory, that is not necessarily true of each of these multiple causes. In fact, only rarely do we find redundancy mechanisms that ensure the existence of some effect even if what in fact causes it were to fail to cause it. On the assumed theory, causal overdetermination cases are not the norm, but effect overdetermination cases are.

More troublesome for Lewis is the second step of his argument, the step that moves from asymmetry of overdetermination to the falsity of backtracking conditionals. Lewis's argument was that overdetermined events cannot easily be removed in selecting possible worlds because to do so involves numerous law violations.[137] Therefore, backtracking counterfactuals are almost always false, because the event referred to in

---

136. HORWICH, *supra* note 88, at 215.
137. Law violation is the only sense I can make of Lewis's notion of the "miracle" involved in breaking the "link" between events, one of which is nomically sufficient for the other. Lewis, *Counterfactual Dependence and Time's Arrow*, *supra* note 82, at 50. I am thus at a loss to understand Lewis's later remark that in assessing the size of miracles (and thus the closeness of worlds), "[i]t's a blind alley to count violated laws." DAVID LEWIS, *Postscripts to "Counterfactual Dependence and Time's Arrow," in* PHILOSOPHICAL PAPERS, *supra* note 82, at 52, 55.

their consequents must exist (given the sufficiency of other effects for it), despite the nonexistence of the event referred to in their antecedents. For example, even without the occurrence of some ripple R_1, the stone still must have been dropped, given the sufficiency for such stone-dropping of each ripple $R_2, R_3 \ldots, R_n$.

If this is a good argument against backtracking counterfactuals ever being true, is it not an equally good argument against forward directed counterfactuals ever being true in causal overdetermination cases? In the concurrent overdetermination case of the two fires that join forces to burn down plaintiff's house, consider the counterfactual:

(34) If the defendant's fire had not been set, the plaintiff's house would not have been burned to the ground.

Since this is an overdetermination case, we should find the possible world where plaintiff's house would have been burned to the ground closer to the actual world than the possible world in which the house would not have been burned. Therefore, (34) is false, and the defendant's fire did not cause the destruction of the house (and neither did the other fire). This would stick Lewis with the extremely unacceptable conclusion that in all causal overdetermination cases there is no causation of the destruction by any single fire—not "too much causation," (that is, too many causes), but no causation at all. No theory of causation can accept this conclusion.

Worse, Lewis's theory also seems committed to backwards causation in the causally overdetermined cases. Imagine a possible world in which there is no burning of the plaintiff's house to the ground. Keeping that world as close as can be to the actual world means not having multiple law violations; therefore, in that world neither of the fires exist (if they did, given the sufficiency of each, there would have been a destruction). So the backtracking counterfactual:

(35) If plaintiff's house had not burned to the ground, there would not have existed either fire,

is true, and on the counterfactual theory of causation, the house burning caused each fire.[138]

The upshot is that Lewis needs a notion of closeness of possible worlds that does not take the minimization of law violations to make for a closer world per se. Perhaps there is a volume discount on multiple violations of the same law? If so, then perhaps Lewis can avoid these two absurd conclusions in the causal overdetermination cases. But by

---

138. This is a point nicely made by HORWICH, *supra* note 88, at 215.

the same token, he will not have solved the original worry about the temporal asymmetry of causation in nonoverdetermination cases.

What Lewis really needs is to argue that, appearances to the contrary notwithstanding, there are no true causal overdetermination cases. There is a very popular argument to this conclusion in both the legal and the philosophical literature.[139] It goes like this. If we look closely at supposed cases of causal overdetermination, we will see that there are never really duplicative sets of sufficient conditions for the same effect. There will always be differences in the effect if one or another of the sufficient conditions, or both together, cause "it" so that, really, "it" will not be numerically the same event at all.

In the two fires example, if fire one gets there first, that will be a different house destruction than if fire two gets there first, and both will be different destructions from the destruction that would occur from the fires joining and the resultant fire destroying the house. There will be qualitative differences in the destructions, for instance, different temporal locations or durations, different spatial occupations, different amounts of energy, and so on. These qualitative differences will then make for numerical differences in destruction events across the three possible worlds imagined.

As I shall examine in the next subsection, there are good reasons to believe that this argument is false, root and branch. The only point needed here is a more modest one: Lewis could not avail himself of such an argument without destroying the asymmetry of overdetermination on which his theory rests. For if fine-grained individuation of effects works to dissolve problems of causal overdetermination, then fine-grained individuation of causes will also dissolve the apparent overdetermination of causes by their multiple effects. A fire that destroyed but one house would be a different fire from one that destroyed two houses, three houses, or a whole town. Lewis must thus reject this supposed dissolution of causal overdetermination, and, in fact, he does so.[140]

---

139. Wright, *supra* note 19, at 1777–80, describes some of the legal literature on this strategy. The drafters of the Model Penal Code, for example, thought they could utilize this strategy to defuse the counterexamples presented by overdetermination cases. *See* MODEL PENAL CODE § 2.03(1) cmt. 2 (1985). In philosophy, this view is represented by MACKIE, *supra* note 77, at 46–47, and McDermott, *supra* note 125, at 539–44.

140. Sort of. *Compare* LEWIS, *supra* note 117, at 197–99, *with id.* at 205 n.26; *see also infra* notes 147, 149.

2. The Necessity of Direct or Stepwise Counterfactual Dependence for Causal Dependence

We have been considering the question of whether counterfactual dependence is sufficient for the existence of causation. I turn now to the question of whether counterfactual dependence is even necessary to causal dependence, the "ancestral" of causation. Is it, in other words, even a part of the story about the nature of causation?

The principal reason for thinking that direct or stepwise counterfactual dependence is not even necessary to causal dependence stems from the causal overdetermination cases. The general form of the worry is easily grasped. A causal overdetermination case is one where two or more sets of conditions are independently sufficient for the production of some harm. For example, if each of two fires is sufficient for the destruction of a house, then it follows that neither fire is independently necessary for the house's destruction. On the usual counterfactual theory of causation, where it is at least necessary that a cause be a necessary condition of its effects, that means that neither fire caused the destruction of the house! This is an unacceptable conclusion, both metaphysically, because it goes against firm causal intuitions, and morally and legally, because it exonerates two culpable fire-starters from liability for harm that they plainly caused.

It is also unacceptable because such a conclusion is hard to square with another implication of the counterfactual theory in these overdetermination cases, namely, the implication that because the two fires together are jointly necessary, they, considered together as a disjunctive unit, did cause the destruction.[141] If neither fire alone caused the destruction, how can the inclusive disjunction of them cause the destruction? Our normal logic is that if $fire_1$ does not cause the destruction, and if $fire_2$ does not cause the destruction, then it is not the case that either $fire_1$ or $fire_2$ (or both) caused the destruction. Formally, $(\sim Cx \cdot \sim Cy) \supset \sim (Cx \vee Cy)$.

---

141. Lewis, like many other philosophers, *see, e.g.*, D.M. ARMSTRONG, UNIVERSALS 82–83 (1989), finds disjunctive combinations of events with causal powers to be counterintuitive. LEWIS, *supra* note 117, at 212; DAVID LEWIS, *Events*, *in* PHILOSOPHICAL PAPERS, *supra* note 82, at 266–69. Lewis thus proposes that we regard the two sufficient fires as parts of one larger event, "the fires of April," say. LEWIS, *supra* note 117, at 212. Then one can say that this larger event was necessary for the destruction of the house. Still, even as a matter of mereology, there is a puzzle here. Some larger event c is a cause even though no part of c is a cause. More exactly: Event c consists of two parts, c_1 and c_2. There is no other part to c; c_1 does not cause e; c_2 does not cause e; yet c does cause e? Suppose a hammer blow (of more force than needed) causes a chestnut to be cracked. Surely the hammer blow causes the chestnut to be cracked only if some part of that blow causes the chestnut to be cracked. It is only in the limiting case where c is a minimally sufficient condition that there would be no mereological puzzle here. Yet in the cases here considered we know that c is not minimally sufficient.

This is no more than DeMorgan's Laws in logic. The counterfactual theory, however, contradicts this: The inclusive disjunction of the occurrence of fire$_1$ and the occurrence of fire$_2$ is a cause of the destruction (because that disjunction is necessary to that destruction), even though it is not the case that either fire caused the destruction (because neither fire was necessary to that destruction).

A common legal response to this well-known objection to the counterfactual theory of causation is by-and-large a puzzling response. Those many legal theorists who recognize that none of the defensive maneuvers, which we shall shortly examine, rescue the theory, nonetheless put aside the overdetermination cases as an unimportant and rare kind of aberrational side show. Yet the number of overdetermination cases that actually occur in real life is irrelevant to the problem they pose for the counterfactual theory. Unless appearances are deceiving, the overdetermination cases by themselves show that the counterfactual theory cannot be a theory of causation. At most, the theory could give a heuristic for the existence of causation. But what causation *is* would be untouched by the theory.

So I take such cases seriously as a challenge to the counterfactual theory. There are five kinds of cases that we should distinguish: ordinary garden-variety concurrent causation, symmetrically overdetermined concurrent, asymmetrically overdetermined concurrent, mixed concurrent overdetermination, and preemptive overdetermination. The ordinary, garden-variety concurrent cause cases are those numerous cases where several simultaneously present conditions are all individually necessary, and only jointly sufficient, for the production of some harm. Such cases are ordinary or garden-variety in that they are not overdetermination cases, and thus do not pose any problems for the counterfactual theory of causation. I mention them because they form a helpful background against which to see clearly the four kinds of overdetermination cases.

(1) A symmetrically overdetermined concurrent cause case is exemplified by the two fires, each sufficient by itself to burn the plaintiff's house, that join and together as one fire burn the house.[142] Simultaneously noisy motorcycles, the noise of each of which would scare the plaintiff's horse,[143] flood waters that join to flood a farm, and mortal wounds simultaneously inflicted, are further examples of

---

142. *See* cases cited *supra* note 21.
143. Corey v. Havener, 65 N.E. 69 (Mass. 1902).

such cases. Such cases are overdetermination cases because each fire, flood, and motorcycle, is individually sufficient to produce the effect in question. They are concurrent because they cause such an effect only commingled into one fire, one sound, one flood, one wound, or one blood loss, and they are symmetrical in that each fire, etc., is equally potent, that is, sufficient.

(2) In an asymmetrically overdetermined concurrent cause case, by contrast, one or more sufficient fires merges with one or more fires that are not individually sufficient. Or a small flood, insufficient to damage the plaintiff's land, is joined by a big flood that would have been sufficient by itself to damage the plaintiff's land.[144] Or, one defendant inflicts a small wound, another defendant inflicts a major wound that causes such rapid loss of blood that the victim would have died from that alone, but the victim in fact dies from loss of blood from both wounds.[145]

(3) A mixed concurrent cause case is one that stands halfway between the ordinary, garden-variety concurrent cause case and the symmetrically overdetermined concurrent cause case. For example, three equally sized fires join to burn the plaintiff's house, when any two of them would have been sufficient to do the job.

(4) Finally, a preemptive overdetermined cause case is one where there are two conditions poised to do some damage, each being sufficient to do so, but one operates first, preempting the ability of the other to cause the harm. For instance, two independently set fires, each sufficient to burn the plaintiff's house, do not join. One fire reaches the house first and burns it to the ground, and then the second fire arrives, but there is no house to burn.[146]

Notice that all of these cases pose the same general problem for the counterfactual theory of causation: The presence of a sufficient condition renders the other condition not necessary, and thus, not a cause under the counterfactual theory. Still, the cases differ enough to have generated

---

144. *See* City of Piqua v. Morris, 120 N.E. 300, 300 (Ohio 1918) (describing a situation where negligent maintenance of drainage wickets on a reservoir joined a flood so large that it would have overflowed the drainage wickets even if properly maintained).

145. People v. Lewis, 57 P. 470, 473 (Cal. 1899) (concluding that where a murder victim died from the loss of blood caused by a major knife wound and a minor bullet wound, both wounds caused the death because "[d]rop by drop the life current went out from both wounds, and at the very instant of death the gunshot wound was contributing to the event").

146. Perhaps the most famous preemptive cause case is the hypothetical one put by James Angell McLaughlin, *Proximate Cause*, 39 HARV. L. REV. 149, 155 n.25 (1925). *A* and *B* each independently intend to kill *V*, who is headed out into a desert; *A* drains *V*'s water keg, replacing the water with salt; *B* steals the keg not knowing of the substitution; *V* dies of thirst.

different attempts to wiggle around them, so I shall treat the four kinds of overdetermination cases separately below.

Before doing so, however, I shall first address the general defensive strategy commonly applied to them all. This is the fine-grained individuation of effects strategy mentioned before, a strategy that seeks to show that there really are no overdetermination cases of any kind. Earlier I addressed only the usability of such a strategy by someone, such as David Lewis, who relies on overdetermination of causes by their effects to answer the problem of asymmetry though time. Now, I shall address the correctness of the move, considered by itself.

There are three reasons to reject the move. The first is the extraordinary promiscuity introduced into the counterfactual theory by the fine-grained defense of it.[147] If any difference in properties of an event, such as a house destruction, makes for a different event across possible worlds, then the number of conditions necessary for that event, with exactly these properties, is staggering. To begin with, if one allows any difference in relational properties to make a difference in event identity, then every event is a necessary condition for every other event. Consider the house destruction by fire. If this occurred three years after Princess Diana's death, then a relational property of the event is that it occurred three years after Princess Diana died. Princess Diana's death was necessary for the event to have this property, so Princess Diana's death was one of the causes of the house destruction.

Even if one stipulates away changes in relational properties, the fine-grained move generates enormous promiscuity. Think of the conditions necessary for the house destruction that occurred to have had exactly the properties it did: the exact temporal duration of the burning of the house, the intensity of the heat generated, the sounds made by the destruction. Facts such as how the house was built, what age wood was selected, and

---

147. David Lewis gave his own examples of promiscuity in (at one time) rejecting the move for this reason. *See* LEWIS, *supra* note 117, at 197–98. Yet Lewis ultimately came to think that there was no determinate answer to the question of event individuation across possible worlds and that one could rescue the counterfactual theory from overdetermination counterexamples with the extremely fine-grained theory of "event alterations." An event alteration is any change, no matter how minute, in the time, place, and manner of event occurrence. Lewis then made causation depend on event alterations and not on event identity. *See* Lewis, *supra* note 79, at 186–88. As Lewis recognized, his adoption of the fine-grained strategy leaves the counterfactual theory open to the objection that "almost everything that precedes an event will be counted among its causes." *Id*. at 188. Lewis hoped one could deal with the flood by pragmatic considerations ruling out small differences in most contexts.

the type of insulation material selected, join a million other items to generate precisely the intensity of heat the house burning possessed. So a cause of the house burning, along with the defendant's negligent starting of a fire, was the degree of termite infestation? Surely not.

Even worse, many of the necessary conditions for the effect to have exactly the properties it in fact had are conditions that retarded and delayed the effect and thus seem in no sense to be the cause of that effect. For example, you throw gasoline on my house and light it up. I desperately try to put out the fire by throwing water on it. My water slows the fire somewhat but does not extinguish it, and my house is destroyed. On the fine-grained view, my water-throwing is as much a necessary condition of my house's destruction as your gas-throwing and lighting. For, according to this view, that destruction would have been a different destruction had I not thrown the water.[148]

Second, the fine-grained move adopts a theory of numerical identity of event tokens across possible worlds that seems wrong on its face. In the actual world, it is true that at any one time, if two putatively distinct particulars do not share all of their properties, then they cannot be one and the same particular. This Leibnizian criterion of identity says that a difference in properties makes for a lack of identity between the particulars possessing such properties, and the principle is surely true at a given time in the actual world. However, it is just as surely false as a criterion of numerical identity over time. On the most plausible theories of personal identity, you are the same person now as you were two years ago even though the properties you possess differ somewhat. You are

---

148.   I am assuming that the counterfactual theorist adopts Jonathan Bennett's proffered solution to the apparent asymmetry between hasteners (which generally seem to cause what they hasten) and delayers (which generally seem not to cause what they delay). Bennett's solution is to deny the asymmetry, so that events that either hasten or delay some event *c* are among its causes. JONATHAN BENNETT, EVENTS AND THEIR NAMES 70 (1988). This, of course, exacerbates the promiscuity problem for the counterfactual theory by treating delayers as causers when common sense clearly rejects them as such.

As to just why common sense is correct (and the counterfactual theory thus incorrect) here is a matter of some philosophical puzzlement. *See* Bennett, *supra* note 76 (who first raised the asymmetry problem against the counterfactual theory); Lawrence Brian Lombard, *Causes, Enablers, and the Counterfactual Analysis*, 59 PHIL. STUD. 195 (1990); Penelope Mackie, *Causing, Delaying, and Hastening: Do Rains Cause Fires?*, 101 MIND 483 (1992). Despite this extensive literature, I do not raise the asymmetry between delayers and hasteners as a separate problem for the counterfactual theory. In truth, the problem is part and parcel of the problem posed for the counterfactual theory by omissions. In a nutshell, delayers (*c*) are mere preventions of events that, had they not been prevented by *c*, would themselves have caused some event *e*. Event *c* does not cause event *e* in such cases, not because *c* is an omission (which it is not), but rather because *c* would have to cause *e* *through* an omission (namely, the omission of some putative cause of *e*). *See* discussion *supra* notes 124–27.

taller, shorter, heavier, smarter, and nicer than you were then, but none of these are your essential properties. Over time it is only your essential properties that must remain unchanged for you to remain you.

The same is true for event identity across possible worlds. Grant for the moment that there would be minute differences in duration, intensity, and sound, in the house destruction(s) in three different possible worlds: one where the defendant's fire burns the house by itself, another where the other fire does the job by itself, and a third possible world where both fires join to burn down the house. Unless these minute differences in house burnings affect the essential properties, there is but one house burning event here.[149]

The third reason takes away the arguendo concession just made: Why must it be the case that there is any qualitative difference in the three house burnings imagined? Surely it is possible that in every detail, save causal genesis, the "three" house burnings are qualitatively identical. In which case, even granting the erroneous extension of Leibniz to identity across possible worlds, there would be the unwanted conclusion of no causation in the overdetermination cases.

I conclude that the attempt to dissolve the problem here by showing that there are no overdetermination cases fails entirely. There are such cases, and we should now see how the counterfactual theorist proposes to deal with them.

a. Symmetrically Overdetermined Concurrent Cause Cases

There is an old saying in philosophy to the effect that one person's reductio ad absurdum is another person's valid inference. One response of the counterfactual theorist is to admit that there is no causation in the symmetrically overdetermined concurrent cause cases. David Lewis came close to this when he professed to lack any firm pretheoretical intuitions about such cases.[150] Even if this were true, still the counterfactual

---

149. Lewis, *supra* note 79, at 185–88. Lewis attempted to sidestep this objection by granting its truth vis-à-vis actual event identity but substituting an extremely fine-grained notion of event alteration. An event is altered by any change in its time, manner of occurrence, or spatial location, across possible worlds so that even though it is the *same* event it still would be an *altered* event if it would change to any degree in any of these dimensions. This is fine as far as it goes. The rub comes when Lewis then sought to alter the counterfactual analysis of causation so that event alteration is important and event identity is not. *See generally infra* text accompanying notes 176–91 (rejecting this altered counterfactual analysis).

150. Lewis, *Causation*, *supra* note 82, at 191 n.12. Lewis later disavowed the lack

theorist has to account for the fact that his theory does not lack firm conclusions here. On the counterfactual theory, it is plain that in these cases neither sufficient condition is necessary to the occurrence of the harm. Thus, on the counterfactual theory, a firm conclusion of no causation fails to match Lewis's unfirm intuitions, which should not happen if Lewis's theory is to match our pretheoretical intuitions about causation.[151] In any case, no one really believes such denials. Of course there is causation in such cases, and a theory that cannot account for that fact cannot be an acceptable theory of causation. The law of torts and of crimes joins common sense in finding causation in almost all such cases.

If we accept as we must two facts: (1) there is causation between c_1 and e, and between c_2 and e, in the symmetrically overdetermined concurrent cause cases, and (2) neither c_1 nor c_2 is a necessary condition of e (even though the inclusive disjunction of the occurrence of c_1 or c_2 is a necessary condition of e), then the counterfactual theorist needs some way of rescuing his theory. Very few strategies come to mind.

One possible strategy is to build on the intuition that surely independently sufficient conditions, when *considered together*, are the cause in the joint fire cases. Therefore, when considered separately, they must each be in some sense a cause as well. A number of lawyers and philosophers have thought that they can salvage the counterfactual theory by building on this intuition.[152] They insist that the counterfactual test works in these cases just so long as one restricts the causal question asked to the two fires, the two noises, and so on, considered together. And of course, but for either fire occurring, there would have been no destruction of the house, so the counterfactual test yields the intuitively correct result. Yet how or why one imposes liability on the independent (not in concert) defendant without finding individual causation by that defendant, remains a mystery.

Perhaps one can demystify this proposal in this way: Alter the counterfactual test so that c is a cause of e if Oc ("e occurs") is a disjunct of the disjunction, Oc or Od, and (Oc or Od) is a necessary condition of e? This proposal will not work. Notice that every event is the cause of e under this definition. For if d is a necessary condition of e, then the disjunction is necessary too, whatever event "c" names. Let e equal the house destruction, c equal Princess Diana's death, and d equal the fire that burned the house. Then the disjunction is necessary to e, so

---

of any common sense judgments about such cases. LEWIS, *supra* note 117, at 208. The last word on this point by Lewis returned him to his earlier sense that we have no firm intuitions about these cases. *See* Lewis, *supra* note 79, at 182.

151.   *See* HORWICH, *supra* note 88, at 210; McDermott, *supra* note 125, at 525–26.

152.   *See, e.g.*, MACKIE, *supra* note 77, at 46–47.

Princess Diana's death is a cause of the destruction of the house. To avoid such an absurdity, we could restrict d in the above definition to events not themselves independently necessary to e. Then the definition works in the symmetrically overdetermined concurrent cause cases. The problem is that the definition excludes the common, garden-variety concurrent cause cases from being cases of causation at all. Let c be the electric spark, e be the explosion it causes, and d_1, d_2 be the presence of hydrogen and oxygen, respectively. On the revised definition, the spark did not cause the explosion because the other conditions (d_1, d_2) were independently necessary. Of course, c could be the cause if one were allowed to disjoin c to Princess Diana's death, because in that disjunction the "d" is not necessary!

Another approach might be to adapt the strategy by which David Lewis seeks to rescue the theory in the preemptive cause cases. This strategy utilizes his idea of a causal chain. Even if c is not a necessary condition of e, perhaps c is a necessary condition of d, and d is a necessary condition of e. Then there would be counterfactual dependence of e on d, and of d on c, and thus, links of causal dependence, and thus, causation between c and e. Consider the two fires again. Let c_1, c_2, be the independent setting of the two fires, d be the burning of the joint fire just before it reaches plaintiff's house, and e be the destruction of plaintiff's house. Then neither c_1 nor c_2 is necessary to e, yet they cause it because they are individually necessary to d, and d is necessary to e.[153]

The trick is to make sense of c_1 and c_2 being individually necessary to d, while c_1 and c_2 are not individually necessary to e. One would have to revert to the fine-grained individuation move put aside earlier. That is, here, one would be finely individuating the intermediate event, d, rather than the ultimate effect, e, but the game is the same.[154] One has to say that the fire after joinder in the actual world is not the same fire as the fire that exists in the possible worlds when there is no c_1, or there is no c_2—the fire is bigger and stronger. The three reasons given for rejecting

---

153. In the terms of the trade, d is a "Bunzl event," an event that is nonredundantly but jointly caused by c_1 and c_2, and which itself then causes e. *See* Martin Bunzl, *Causal Overdetermination*, 76 J. PHIL. 134 (1979).

154. Thus, Bunzl (in a case where d is an electric shock that caused e, the death of a person) resorts to very fine-grained individuation of the microphysics of the shock produced by c_1 (versus that produced by c_2) in order to conclude that the shock (or is it shocks?) counterfactually depended on c_1 individually (as well as on c_2 individually). *Id.* at 141.

this move earlier suffice to reject it here as well.[155] In particular, there is no reason to suppose that there will always be any qualitative difference between the d that actually occurred (the joint fire), and the d that would have occurred in the absence of c_1 or in the absence of c_2.

In addition, there is no reason to think that there always will be an intermediate event d in such cases on which one can work this "Bunzling." Take Lewis's own example of two neurons, C_1 and C_2, both firing and causing another neuron, E, to fire where the threshold for E to fire is below the energy level of the firing of either C_1 or C_2 considered individually.[156] Such cases of direct overdetermination seem not only conceivable but plentiful.[157] Thus, even if, contrary to fact, it worked when it was available, the Bunzl rescue of the counterfactual theory seems unavailable.

A third strategy is to retreat to John Mackie's "INUS" analysis of causation,[158] under which each fire is an insufficient but necessary part of a set of conditions that is unnecessary but sufficient for the effect. An event can be a cause of some event e, on this view, if it is not necessary to e so long as it is a necessary element of a sufficient set of conditions. Each fire, the argument goes, is a necessary member of a set of conditions sufficient for destruction of the house, and the fact that there is more than one sufficient set does not matter because there is no requirement that each set be necessary as well as sufficient for e.

Notice that this is not a counterfactual theory of causation, for the account gives up the crucial idea that a cause must be necessary to its effect. Rather, a cause is essentially a sufficient set of conditions with the afterthought that causes are members of minimally sufficient sets of conditions, that is, each member must be necessary for the sufficiency of the set. The idea that causation is a matter of lawful sufficiency is the hallmark of the Humean and neo-Humean theories of causation. Such theories have problems of their own, particularly with respect to the asymmetrically overdetermined concurrent cause cases, the mixed concurrent cause cases, and the preemptive cause cases. Nevertheless, these are outside the domain of our present concern, which is the counterfactual theory of causation.

---

155. Lewis, puzzlingly enough, seems to favor a fine-grained solution to event identity here. *See* LEWIS, *supra* note 117, at 210–12.

156*. Id*. at 210. Lewis believed that we lack any clear intuitions in these cases. *See supra* note 150.

157. *See* McDermott, *supra* note 125, at 527–28. This point parallels the point to be made against Lewis's solution to cases of early preemption, *infra* text accompanying notes 166–69.

158. J.L. Mackie, *Causes and Conditions*, 2 AM. PHIL. Q. 245 (1965). Wright's NESS test is essentially the same as Mackie's INUS test. *See* Wright, *supra* note 19.

I conclude that the counterfactual theory of causation has no way of dealing adequately with the symmetrically overdetermined concurrent cause cases, as long as it remains a truly counterfactual theory of causation.

b. *Asymmetrically Overdetermined Concurrent Cause Cases*

There are two varieties of asymmetrically overdetermined concurrent cause cases. In one, the smaller causes that are not individually sufficient are nonetheless jointly sufficient. This variety is much like the symmetrically overdetermined concurrent cause cases, the only difference being that one of the two or more individually sufficient conditions is an aggregate of smaller causes. For example, three smaller fires are set, each of which is necessary to form a fire sufficient to burn the plaintiff's house. These fires then join another fire sufficient by itself to burn the plaintiff's house, and the house is destroyed.

More interesting because more distinct are the second variety of asymmetrically overdetermined concurrent cause cases. Here the smaller causes do not add up to a sufficient condition of the harm. For instance, a victim suffers two wounds from two blows by two defendants acting independently of one another. One wound is sufficient to kill the victim by loss of blood, while the other wound is not. The victim dies from loss of blood, both wounds bleeding.

Here the initial response of David Lewis to the symmetrically overdetermined concurrent cause cases may be more tempting. Lewis suggests that we deny that the smaller wound is a cause, so that the fact that it was not a necessary condition of the victim's death does not present a counterexample. Yet the California Supreme Court's causal intuitions in this actual case seem persuasive here: "Drop by drop" the blood flowed out of the victim "from both wounds."[159]

One way of buttressing this otherwise rather naked intuition is by likening these cases to the mixed concurrent cause cases we shall consider shortly. One might think the two wounds example to be like this five wounds example: The victim dies from the loss of blood from five independently inflicted wounds, the loss of blood from each wound being exactly the same; the blood loss from any three of the wounds would have been sufficient for death, but no lesser number of wounds

---

159.   People v. Lewis, 57 P. 470, 473 (Cal. 1899).

would have sufficed. When the victim dies from loss of blood, surely each wounding is a cause of that death. If that is so, then view the actual case as one in which the one big wound is equivalent in causal effect to three of the previous wounds, and the smaller wound is equivalent in causal effect to two of the previous wounds. Surely, the smaller wound in the actual case, which is twice as big as each of the five wounds in the previous case, is also a cause. It shouldn't matter whether the other contributor to the death is one big wound or three smaller wounds, given that the blood loss caused by each set is the same.

If the smaller factor is a cause in both varieties of the asymmetrically overdetermined concurrent cause cases, then these cases present the same problem for the counterfactual theory as does the symmetrical kind of case: These factors are not necessary conditions of some harm and thus should not be causes on the counterfactual theory. In fact, of course, they are causes. So much the worse for the counterfactual theory.[160]

c. Mixed Concurrent Cause Cases

The unique feature that makes mixed concurrent cause cases of some interest is that none of the causes of some event *e* are either necessary or sufficient conditions for *e*. Consider, for example, the five wounds case imagined earlier, where it requires a minimum of three wounds to cause death. Surely no one wishes to deny that each wound caused the death in such a case. Imagine a nonmixed, garden-variety concurrent cause case, where each of five wounds is individually necessary, only jointly sufficient, for the bleeding to death of the victim. Holding everything else constant, these will be smaller wounds causing less loss of blood than in the mixed case. If no one denies causation by each wound in the garden-variety case, it is hard to see how they could deny causation by each wound in the mixed case.

As an alternative example, consider the causal role of individual votes in generating the outcome of an election in which such votes are cast.[161] A garden-variety concurrent cause case is one in which the election's outcome was decided by just one vote, which makes each vote for the winning side necessary for that outcome. Contrast this rare case with the mixed concurrent cause case in which there are multiple votes cast for

---

160. Notice that the second variety of these cases poses equal problems for the Humean and neo-Humean theories because the smaller causes are not necessary elements in a set of conditions sufficient for the harm.

161. This is an example nicely explored in Alvin I. Goldman, *Why Citizens Should Vote: A Causal Responsibility Approach*, 16 SOC. PHIL. & POL'Y 201 (1999), *reprinted in* RESPONSIBILITY, *supra* note *, at 201.

the winning side in excess of what was needed for victory. In such a case no individual vote was necessary, or sufficient, to the election's outcome. Surely the causal role of each individual vote does not differ in these two kinds of cases. My vote is a cause of the outcome in both. The problem for the counterfactual theory of causation is that it discriminates between these pairs of cases in a way that causation does not.[162]

d. Preemptive Cause Cases

In preemptive cause cases our causal intuitions are very clear: A preemptive cause causes all of the damage, and a preempted "cause" is no cause at all. In the two fires cases, when both fires are independently sufficient to burn the plaintiff's house, if the fires do not join, and one fire destroys the plaintiff's house before the other arrives, the first fire is the cause of the harm and the second is not. To my knowledge, no one denies these causal conclusions. The problem such conclusions present to the counterfactual theory is that prima facie neither fire is a cause under the theory. But for the first fire, the second would have burned down the plaintiff's house, and but for the second fire, the house still would have been burned to the ground by the first fire.

Preemptive causation cases present a rather distinctive challenge to theories of causation. Unlike all three kinds of the concurrent cause overdetermination cases we examined before, here there is an asymmetry between the putative causal factors—one is clearly the cause and one is equally clearly not. The problem for counterfactual and other theories of causation is that the facts on which they rely, facts of necessary or sufficient conditions, are symmetrical between the putative causal factors. On its face, this makes it tricky to line up the causal facts with the underlying facts to which causation is to be reduced. Counterfactual theorists, for example, find it easy to characterize the second fire as not necessary to the harm, yet that also seems to make the first fire not necessary either. As another example, Humean theorists who trade in nomic sufficiency find it easy to characterize the first fire as sufficient, but then the same characterization seems to apply to the second fire as well.[163]

---

162. Incidentally, the Humean and neo-Humean theories fare no better with these cases because there is no nonarbitrary way to assign wounds or votes to sets of sufficient conditions such that any wound or vote is necessary to the sufficiency of that set.

163. See the discussion of this in Moore, *Causation and Responsibility*, *supra* note

Lewis's initial response to preemptive cause cases was to rely on the stepwise counterfactual dependence giving rise to a causal chain.[164] This allows him to concede that the first fire was not necessary to the destruction of the house while concluding that the first fire nonetheless caused that destruction. To see this, let c_1 be the setting of the first fire, let c_2 be the setting of the second fire, and let e be the destruction of the plaintiff's house. In Lewis's theory, e depends counterfactually not on c_1, but on some intermediate events d_1, d_2, d_3; c_1 causes e if e counterfactually depends on d_3, d_3 on d_2, d_2 on d_1, and d_1 on c_1. The picture is then:
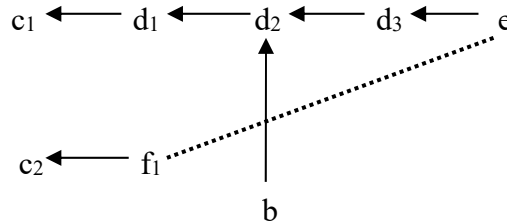
The direction of the arrows is that of counterfactual dependence, so the direction of causation is opposite. Notice that d_2 is a blocking event, because b is caused to occur by d_2, and b is the absence of what c_2 would need to cause e. For example, d_2 might be the burning of certain material located between c_1 and e, and it is this burning event (d_2) that both causes the material closest to the house to burn (d_3) and causes the destruction of material (b) needed by the second fire (c_2) in order for it to reach the house. Now, Lewis can concede that c_1 is not necessary to e, for absent c_1, c_2 would have caused e. Yet c_1 still causes e because causation is stepwise counterfactual dependence, which exists between c_1 and e. There being no such dependence between c_2 and e, c_2 does not cause e.

The trick to understanding this first stab at a solution of the preemption problem by Lewis is to see why d_3 is necessary for e. One might think that it is not because if there were no d_3, then e would still occur because caused by the chain initiated by c_2. Yet notice that by the time d_3 occurs, the c_2 chain is already blocked by d_2's causing of b. So without d_3, there

---

164. Lewis, *Causation*, *supra* note 82, at 191.

will be no e, and the chain of counterfactual dependence from e to c_1 is unbroken. This rather neat solution is undone if there can be backtracking counterfactual dependence. Look at what happens if d_3 turns out to be necessary for d_2, so that if there were no d_3, there would be no d_2. Since d_2 is necessary for b, the blocking event, without d_2 the causal efficacy of c_2 would not be blocked. This means that in the possible world where d_3 does not occur, neither does d_2 nor b, and in that world c_2 produces e after all. So e does not counterfactually depend on d_3, for e will come about even without d_3.[165] This yields the unwanted conclusion that c_1 does not cause e when we plainly know it does.

Lewis denies that d_3 is necessary to d_2. This is a "backtracking" counterfactual which Lewis holds is always false. "If d_3 had not occurred at t_5, d_2 would not have occurred at t_4" backtracks through time and is the converse of the counterfactual dependence Lewis needs to assert (of d_3 on d_2). His denial is based on the judgment that the possible world in which the truth of such backtracking counterfactuals is tested is one where the blocking-initiating event (d_2) occurs even if d_3 does not occur; then b occurs, c_2 does not cause e, and the counterfactual dependence of e on d_3 is preserved.

Notice that if Lewis allowed himself to test the counterfactual, "If d_1 had not occurred, d_2 would not have occurred" in the possible world in which he tests the backtracking counterfactuals, "If d_3 had not occurred, d_2 would not have occurred," then c_1 would not be the cause of e (because d_2 would not counterfactually depend on d_1, and the stepwise dependence of e on c_1 would be broken). But Lewis does not test the former counterfactual in the same possible world in which he tests the backtracking counterfactual. To see why not is to see, again, why Lewis thinks backtracking counterfactuals are different.

Lewis thinks that it is of a piece with the asymmetry through time of counterfactual dependence; d_2 counterfactually depends on a prior event like d_1, but d_2 does not counterfactually depend on a later event like d_3. This is due to the greater commonality of the overdetermination of causes by their effects compared to the overdetermination of effects by their causes. This again turns on the impact of multiple law violations on the similarity of possible worlds. It would take more law violations to have d_2 absent in a possible world where d_3 was absent than it would to have d_2 absent in a possible world where d_1 was absent. Again,

---

165.   *See id.*

however, the worry is whether sheer numbers of law violations can so consistently determine the similarity of possible worlds, without regard to the relative importance or centrality of the laws being violated.[166]

The killer problem for Lewis's stepwise solution to preemption stems from the fact that some preemptive overdetermination cases do not have the intermediary events, d_1, d_2, d_3 on which Lewis's stepwise dependence solution depends. In the two fires scenario, suppose that the fires approach the plaintiff's house from opposite directions so that there is no consumption by the first fire of the fuel necessary for the second fire to reach the house. There is, thus, no blocking events d_2 or b. Rather, the first fire preempts the ability of the second fire to burn the house simply by virtue of the fact that there is no house left to be burnt by the time the second fire arrives at its location. The picture is then a simple one:

$$c_1 \longleftarrow \qquad e$$
$$c_2 \cdots\cdots\cdots\cdots\cdots$$
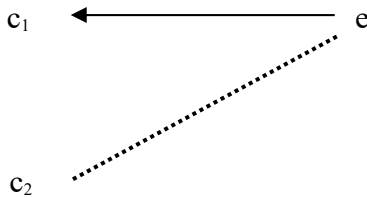
Now, e does need to depend counterfactually on c_1, while e should not depend counterfactually on c_2. However, nothing we have said thus far goes any distance towards showing how this is possible.

Lewis's initial response to this problem was to deny the existence of such cases of "late preemption."[167] The argument was that the destruction of the house by fire two would have necessarily been a later destruction than the destruction that actually occurred because of fire one; and therefore a numerically distinct destruction; and therefore c_1 was necessary to e, the actual destruction that occurred. Yet as Lewis later recognized, this fine-grained theory of event individuation across possible worlds is no better a solution in cases of late preemption than it is in cases of concurrent overdetermination.[168] In both cases such fine-grained individuation multiplies the promiscuity problem to the point that "they will give us lots of spurious causal dependence,"[169] as we saw before.

---

166. *See* discussion *supra* notes 130–31, 137–39.
167. LEWIS, *supra* note 117, at 204.
168. *Id*.
169. *Id*.

Lewis then tried a different tack. On what he called his "extended analysis,"[170] we are to imagine two spatio-temporal regions in which what goes on (the events) is exactly alike and where the laws of nature that govern the two regions are exactly the same. On our example, there is a fire just like fire one in both regions, and there is a destruction of a house by burning in both regions. Lewis's thought is that even though the regions differ in that in one of them there is a fire two that would burn the house down if fire one did not get there first, and in the other there is not, that difference should make no difference to our causal conclusions. If fire one caused the house destruction in the region *without* fire two, then the analogous fire one caused the analogous house destruction in the region *with* a fire two.[171]

This response, to my mind, abandons the counterfactual analysis of causation. Lewis himself conceded that his "extended analysis . . . is less purely a counterfactual analysis . . . ."[172] But it is a more radical departure than that. Now causation is not to be identified as stepwise counterfactual dependence. Rather, a causal relationship between *c* and *e* is to be identified with such stepwise counterfactual dependence as would exist in a region with events and laws relevantly like the actual region in which *c* and *e* exists. Lewis found it "intuitive" that the existence of a second, preempted causal factor is not a relevant difference between the actual world and this hypothetical world in which stepwise counterfactual dependence is to be sought. Yet if we look hard, surely the intuition about what makes the regions relevantly alike is wholly based on the laws in both regions, including the *C/E* law. It is these laws that make us so confident that if fire one caused the house destruction when there was no fire two, then fire one caused the house destruction when there was a (preempted) fire two.

This seems a wholesale abandonment of the possible world view of counterfactuals and a return to the covering law view. Moreover, by substituting the nomic sufficiency of the covering law model for the necessity of the counterfactual theory, Lewis's extended analysis suffers from the same problem as besets the neo-Humeans on preemption.[173]

---

170. *Id*. at 205–07.
171. Lewis also thinks that this extended analysis can also save him from concurrent overdetermination counterexamples. *Id*. at 211. It does not for the reasons I will address later.
172. *Id*. at 207.
173. *See supra* note 162.

Namely, if this analogy makes the preempting fire a cause, it equally makes the preempted fire a cause. After all, the preempted fire is just like a fire in another spatio-temporal region with the same relevant events and laws, save where there is no preempting fire, and in that region the preempted fire causes the destruction; therefore so does the preempted fire when there is a preempting fire? Can't be.

Lewis himself came to accept a version of this last criticism of his solution to cases of late preemption, although he thought the problem arose only in cases of "trumping preemption."[174] Trumping preemptive cause cases are those where the preempted causal factor runs its whole course—a course that normally produces some event *e*—yet it does not cause *e* because some preempting cause "trumps" it. Lewis adopted an example from Jonathan Schaffner: A major and a sergeant both simultaneously order troops under their command to advance, which they do. Only the major's order causes the troops to advance since they always obey the higher ranked officer; the sergeant's order would have caused them to advance if, contrary to fact, there had been no order from the major.[175]

Lewis accepted the idea that such trumping cases differ from cases of late preemption. If there is a difference here, it does not lie in the completeness of the causal chains (of the preempted factor) in the former cases and not the latter. In the two fires case where there is no house to burn because the preempting fire burned it first, this is supposed to be only a case of late, not trumping, preemption. Whether trumping preemption truly differs from late preemption need not detain us, however, since it is Lewis's proposed solution to supposed "trumping preemption" cases that should interest us. If the solution works for such cases, then the fact that there is no real difference between these and late preemption cases generally would be grist for the counterfactualist's mill.

Lewis's proposed solution begins with the fine-grained approach to event alteration discussed before.[176] Every slight difference in the temporal duration, manner, or location of the troops' movement makes for an alteration in the event to be explained, namely the movement of the troops.[177] It is not that these changes would make it a different event. While sometimes that may be true, even when it is not, there is still an alteration in the very same event. Then notice, Lewis continues, that

174. Lewis, *supra* note 79, at 183–84.
175. The example is from Jonathan Schaffer, *Trumping Preemption*, 97 J. PHIL. 165, 175 (2000).
176. *See supra* notes 147, 149.
177. For another example, see Lewis, *supra* note 79, at 185–86.

when, how, and where the troop movement occurs does depend counterfactually on the major's giving his order—for if he had not, the troops still would have moved (in response to the sergeant's order) but that movement would have been an altered version of the movement that actually took place. Thus, the alteration of the troop movement counterfactually depends on the major's order even if the existence of the troop movement itself does not so depend.[178] The final step in Lewis's solution is to urge the sufficiency of the former counterfactual dependence for causation. Specifically, Lewis insists that it is sometimes enough for the truth of the causal judgment, "The major's order caused the troop movement," that how, when, or where the troop movement occurred counterfactually depends on the giving of the order by the major.[179]

It will be recalled that I advanced three arguments against the fine-grained dissolution of the overdetermination problem.[180] Lewis's ideas about alterations of events, and about what he calls "when, how, and where counterfactual dependence," speak only to the second of these problems, the problem of event identity across possible worlds. As I noted earlier,[181] Lewis's ideas here leave untouched the first problem, that of extreme promiscuity. Under such an analysis, as Lewis admits, "almost everything that precedes an event will be counted among its causes."[182] That strikes me as absurd, even if it does not so strike Lewis.

Also untouched is the third problem: Why should we suppose that there would be any alteration in *e* if it had been caused by the preempted factor rather than by the preempting factor?[183] Consider again the troop movement. The movement that would have been caused by the sergeant's order could have been qualitatively identical to the movement actually caused by the major's order. The troops need have no greater readiness to obey one versus the other, the decreased audibility of the sergeant's order (when not accompanied by the major's simultaneous order) need make no difference to the troops (since, for example, either

---

178. *Id*. at 186–87. Lewis is building on a similar analysis proposed by L.A. Paul, *Keeping Track of the Time: Emending the Counterfactual Analysis of Causation*, 58 ANALYSIS 191 (1998).

179. Lewis, *supra* note 79, at 186–87.

180. *See supra* text accompanying notes 147–50.

181. *See supra* note 147.

182. Lewis, *supra* note 79, at 188.

183. As Ned Hall also concludes, "it is child's play to come up with counterexamples" where the substitution of the preempted factor for the preempting cause makes absolutely no difference to the event caused. Hall, *supra* note 116, at 221. Indeed, Hall presents quite a few in the course of his article. *See generally id.*

order was quite audible). On such suppositions, the counterfactual analysis still gives the wrong answer—neither order caused the troop movement—even when that analysis is amended in the way suggested.

The counterfactualist's only hope here is that there are no real world preemptive cause cases where there is no qualitative difference between the effect actually produced by the preempting cause and the effect that would have been produced by the preempted factor.[184] Yet even if this were so, surely there are many possible worlds where there are such cases, possible worlds very similar to the actual world in which we live. If "cause" names a real relation, say, of counterfactual dependency, then surely it does so in those possible worlds where the counterfactual theory comes out false.

Quite apart from inheriting these problems of fine-grained individuation, Lewis's final solution to the overdetermination cases does not solve such cases on its own terms. Lewis supposes, contrary to fact but assumed here arguendo, that there will always be some qualitative difference between the event caused by the preempting cause and the event that would have been caused by the preempted factor. Lewis then supposes that even in the actual case, the preempted factor made some qualitative difference to the event caused.[185] In the example given, the troop movement that would have been caused by the major's order had there been no accompanying sergeant's order would have been different from the troop movement caused by the major's order where both orders were given. The difference between a preempting cause, and a preempted causal factor then becomes one of degree:[186] The major's order is the cause, and the sergeant's is not, because the absence of the major's order would have made more of a qualitative difference to the troop movement

---

184. Lewis, at one time, thought that he could ignore such cases as being physically unrealistic because such cases invariably involved a temporal gap between the preempting causes and the effect. If Bob and Mary each throw a rock at a glass, Bob's rock arriving first to shatter the glass, Mary's rock thus harmlessly passing through the space formerly occupied by the glass, for the shattering that actually occurred (Bob's shattering) to have occurred at the same time as the shattering that would have occurred if Bob's rock had not been thrown (Mary's shattering) requires a delay in the actual shattering that is inexplicable; these have to be differently located (temporally) shatterings for Bob's rock to be a preemptive cause. *See* LEWIS, *supra* note 117, at 203. Lewis's view finds some support in those courts holding that there can be no causation of some event *e* by some other event *c* unless *c* hastens the occurrence of *c* (over and against when *e* would have occurred in the absence of *c*). *See, e.g.*, Oxendine v. State, 528 A.2d 870 (Del. 1987). Against this, notice the trumping example given in the text is not inexplicable in this way: The major's order can cause exactly the same troop movement as that which would have been caused by the sergeant's order. For further examples, see McDermott, *supra* note 125, at 526, 530–31.
185. Lewis, *supra* note 79, at 189.
186. *Id.*

1256

than would the absence of the sergeants' order.[187]

This surely will not do. To begin with, there is nothing impossible about there being perfect symmetry here between the two orders—the absence of either one of them could make no qualitative difference whatsoever to the resultant troop movement. But assuming they both make some difference, then on Lewis's own showing both are concurrent, and not preemptive, causes of the troop movement. If one of them makes a bigger qualitative difference than the other, then we may have an asymmetrical concurrent cause case; yet there is no reason to think, as does Lewis,[188] that only when they make roughly the same difference they should be considered concurrent causes.[189]

Finally, notice the lack of match between causation and counterfactual dependence opened up by this fine-grained defense of the counterfactual analysis. Two events, c and e, stand in a causal relationship only if some facts about such events stand in a relationship of counterfactual dependence. The relata of the two relationships now necessarily differs, inclining one to think that these must be quite distinct relationships. To restore the symmetry, one might urge that causation too relates to facts and not events.[190] Yet that is a high cost to pay to save the counterfactual theory from preemptive cause counterexamples.[191]

I conclude that the preemptive overdetermination cases join the three varieties of concurrent overdetermination cases in showing that counterfactual dependence is not necessary for there to be causation. Since such dependence is not sufficient for causation either, that reduces counterfactual dependence to something of a mere heuristic for causation: Often but not always, counterfactual dependence of some event e on some event c is good evidence that c caused e. But c can cause e without there being such dependence, either directly or in a stepwise fashion, and there can be such dependence without c causing e.

---

187. *Id.*
188. *Id.*
189. I ignore Lewis's further complication in terms of varying the time, place, and manner of the causes as well as of the effects, *see* discussion *supra* note 79, since it in no way diminishes the force of the objection. *See* Lewis, *supra* note 79, at 189–91.
190. L.A. Paul, *Aspect Causation*, 97 J. PHIL. 235 (2000) (defending the counterfactual theory by going in this direction).
191. Given their view that for a variety of reasons causal relata should be construed to be facts and not events, some would regard this not as a cost but as a virtue for the counterfactual analysis. *See, e.g.*, BENNETT, *supra* note 148, at 376–84; MACKIE, *supra* note 77, at 48, 265; Wright, *supra* note 19, at 1803–05, 1816.

It is time we return to the ten legal contexts with which we began. The counterfactual tests used by the law in these ten contexts cannot be justified on causal grounds. It remains to enquire whether these uses can be justified on any other ground. I consider each separately below.

A. The Basic Case of Misfeasance in Tort

I shall defer the general question of substituting a counterfactual baseline test across the board for causal tests of liability. Rather, here we should simply note the potentially far-reaching ramifications of realizing that the nature of causation is not counterfactual dependence. Much of the support of the law's traditional two-step dance dividing causal prerequisites of liability into cause in fact and proximate cause has been the thought that scientific causation is simply counterfactual dependence. Given the enormous promiscuity of that notion of cause in fact, as we have seen, it was natural to supplement it with some more discriminating test of (proximate) causation. Eliminate that thought, and the possibility of a unitary causal prerequisite to liability emerges, one with the discriminating power of the old two-part analysis except based wholly, rather than partially, on the science of causation. Such unitary analyses are, of course, not unknown in legal theory,[192] but when they start from a necessary condition analysis of causation, as Hart and Honore have done,[193] it is definitely an uphill struggle.

B. The Basic Case of Misfeasance in Contract

Contract law differs from torts here, at least for nonconsequential damages. Even if one throws out the counterfactual sine qua non test as a causal prerequisite of liability in contract, still one should analyze the expectancy measure of nonconsequentialist damage in explicitly counterfactual terms. For contract law never really needed causal notions to connect a defendant's wrongful breach of promise to the "benefit of the bargain" to which the promisee is entitled. If promisors truly have the normative power to change the state to which the promisee is entitled, then the promisor who does an act he promised not to do corrects his injustice, not by paying for the harm he has caused (the value of the promisee's reliance), but rather by paying for the "value of

---

192. *E.g.*, Richard A. Epstein, *A Theory of Strict Liability*, 2 J. Legal Stud. 151 (1973).
193. Hart & Honore, *supra* note 109.

his promise." The value of the promise is what it would have been worth if, contrary to fact, the act in breach had not been done. So long as there are good justifications for the expectancy measure of damages in contract law, either in terms of a promissory or an efficiency theory of contract, then contract law has all the justification it needs to continue to use counterfactual tests even while recognizing that they have nothing to do with causation.

C. Liability for Nonfeasance in Tort and Contract

Absent events cause nothing. Accept that truth, and the law's ascription of liability for omissions becomes problematic. Can we make sense of the law's requirement that omissions must cause harms for liability to attach, once we deny sense to omissions causing anything?

It is clear that some relationship needs to exist between omissions and harm before legal liability should attach. If nothing I can do can save you (to whom, ex hypothesis, I owe a duty) as you head over Niagara Falls, I cannot be made liable for your death because I did not try to save you. My omission has to make a difference for liability to attach. If this is not a causal difference, as it is not, it still seems well captured by counterfactual dependence: If I had thrown the rope or dove in, would you not have died? Even if I have the positive duty to save you, I must have the ability to save you to breach that duty, and that ability seems well captured by a counterfactual test.

Lewis worries that if we conceive of omissions as the absence of events, then "counterfactual dependence" has to change its meaning; for what is absent is any instance of a type of action, such as saving you, so that the relevant counterfactual is not about particulars but about universals. As Lewis puts it:

> I cannot analyze this in my usual way, in terms of counterfactual dependence between distinct events. Instead I have to switch to a different kind of counterfactual for the special case. The counterfactual is not: if event *c* (the omission) *had not* occurred. . . . It is rather: if some event of kind *K* (the omitted kind) *had* occurred. . . .[194]

I do not see the problem here, at least for liability purposes. If there is any particular act that I could have done that would have saved you, then I am liable for not doing that act. Surely we can frame a possible world

---

194. LEWIS, *supra* note 117, at 191–92.

differing from the actual world by the addition of a particular act no less than we can frame a possible world differing from the actual world by the subtraction of a particular act. (After all, is not Lewis's "modal realism" committed to the existence of such particular possibilia?)

It is true that, in framing counterfactuals for omissions, we will have an infinite number of them to examine in order to see if there was anything I could have done to have saved you. Absence of liability would thus be difficult to prove. So why not put the burden of proof where in fact it lies under our law: The plaintiff must prove one particular act I could have done that, if done, would have saved you?[195]

Why should it have any bearing on my liability whether I could have saved you? Why, that is, if I have a legal duty to save you, am I not liable whenever I do not try to do so, irrespective of whether or not I could have succeeded if I had tried? Consequentialists have a ready answer here: The incentives given by a liability rule stripped of any counterfactual limitation would produce too many deaths. Rescue attempts are not without risk to the rescuer. Bootless attempts, where there is no chance of success, risk life needlessly. On a straight minimization-of-loss-of-life basis, we should not want people attempting impossible rescues.

For those of us who are not exclusively consequentialists about moral or legal duties, the argument has to be different. Or so it seems. Yet ask, what risks does any plausible agent-relative theory demand be taken in order to effect the rescue of one to whom one owes a duty? A doctor uncontroversially owes a duty to rescue her patient—in life-threatening situations, is that a duty requiring actions risking the doctor's life up to the point that the chance of saving the patient is only very slightly larger than the chance of killing the doctor? Does even a parent have the obligation to dive in to save his drowning child, if the chance of saving the child only slightly exceeds the chance of drowning the parent along with the child? Agent-relative morality in fact seems more permissive of foregoing rescue attempts than does a consequentialist morality. If that is true, then it is out of the question for morality to demand rescue attempts where there is no chance of success.

We may well think it understandable, and perhaps even heroic, for a parent to elect to sink or swim together with his child, no matter what the odds. At a minimum, it shows a depth of emotional commitment that is admirable; indeed, such a parent might well prefer his own death

---

195. This is analogous to what the law does in requiring proof of negligence. It puts the burden on the plaintiff to isolate some cost-effective precaution that, if taken, would have prevented the injury. *See* Mark F. Grady, *Untaken Precautions*, 18 J. LEGAL STUD. 139 (1989).

to living with the sight or the sound of his child perishing in his memory. But surely the parent who elects otherwise does not breach any moral duty and is not the appropriate subject of legal sanctions.

If all this is so, then the law is on solid ground in placing counterfactual prerequisites to liability for omissions. The fact that these are not causal prerequisites does not matter at all. They are still justified features of our law.

D. Limitations on Liability for Mere Allowings

With the doing-allowing distinction, we enter onto much more contested, and contestable, territory, both as to the correct outcomes and as to the correct basis for any given outcome. For purposes of discussion, distinguish three cases. The first is Judy Thomson's symphony attender, who is wrongfully attached to the lead violinist as a kind of personal life-saving equipment; the second is the bystander who, while under no duty to do so, undertakes to save another drowning in the water but then abandons the rescue attempt because he thinks better of it; the third is the euthanasia case, where the doctor who initially plugs the patient in to the respirator then unplugs the patient after deciding that further use of the respirator on this patient is medically unjustified.

Once we separate causation from counterfactual dependence, we should see outcomes in these cases as having two quite distinct justificatory bases. One is a counterfactual basis. On this basis, we first isolate a "baseline," a status quo to which there is a strong entitlement by the actor. In Thomson's case, that baseline is clear: It is the symphony attender at her concert and still unplugged to the violinist; it is not the symphony attender plugged into the violinist as his personal life-saving equipment. The symphony attender's right to bodily integrity, and the wrongful violation of that right, makes this choice of baseline very clear in this case. Secondly, and still on the counterfactual approach, we then ask a counterfactual question about the action of the symphony attender's in unplugging herself from the violinist: If she unplugs herself, does she do no more than return herself and the violinist to the preplugging baseline? Granted, she caused the violinist's death by unplugging him, but the relevant question for liability is the counterfactual question, not the causal question.

The second approach to these cases is, of course, the causal approach. Sticking with Thomson's scenario, all the symphony attender did when she unplugged the violinist was let nature take its course. What killed

1261

him, or in other words, what caused his death, was the sickness with which he was originally inflicted. Her unplugging him only allowed that sickness to kill him. On this account, it is irrelevant how she came to be plugged into him. So long as she owes this perfect stranger no positive duties of rescue, she, like anyone, is free to go about her business even though this means the violinist succumbs to his disease.

We can test which of these accounts we find most plausible by isolating each basis from the other within the hypothetical. First, eliminate the causal basis for any justification: The only way the symphony attender can unhook the violinist is if he is dead, so she bludgeons him with blows that kill him just when he would have died from his disease, and with no more pain, had he never been plugged in to start with. Can the symphony attender return the violinist to the baseline by this more direct causing of his death? Second, eliminate the counterfactual basis: It is a cruel fact of nature that symphony attenders often end up plugged into musicians faced with life threatening diseases, and everyone knows this. (If this is too much for you, run this with Siamese twins.) May the symphony attender, or anyone else, still unplug the violinist knowing that he will succumb to his disease, because they are only letting nature take its course but not causing his death?

To my mind, the counterfactual argument is a stronger basis for justifying Thomson's result than is the causal argument (even though her result is perhaps justified only if we have both bases for reaching it). The reverse seems to be true in the euthanasia case I imagined. It surely does not matter which doctor plugged the patient in to the respirator. Surely, any doctor can unplug the patient with no more liability than attaches to the doctor who did plug in the patient. So what I have elsewhere called a "personalized baseline test" is not very intuitive.[196] Nor is a perfectly generalized baseline test intuitive; surely the proverbial stranger cannot unplug the patient and return him to his baseline. Restricting the baseline to "medical personnel but no other" makes plain what is really doing the work here: Medical personnel can be justified in doing things that bring about death in a way others are not, irrespective of whether they are returning patients to some baseline or not.

Yet this is not to say that doctors can directly kill their patients, by knives, lethal injections, or otherwise. The causal distinction does this much work here: Doctors can be justified in letting nature take its course more easily than they can be justified in causing death. The causal basis thus does some work here, even if the main workhorse is a justification that is independent of either counterfactuals or causation. The

---

196. MOORE, *supra* note 13, at 27.

counterfactual basis, by contrast, seems to do no work here at all.

Much the same modest contribution is made by the causal analysis to the resolution of the would-be rescuer, my second of the three cases earlier distinguished. If he undertakes to rescue while under no duty to do so, and then abandons the attempt, thereby letting the person in the water drown in the natural course of things, it is hard to see that there should be liability for the death. He merely allowed the victim to die. By contrast, if he reaches the victim, and then "abandons" the rescue by pushing the victim under (just where and when he would have gone down had there been no rescue attempt), he has caused the victim's death. The fact that he did not make the victim worse off—a counterfactual question—does not insulate the latter "abandoner" from liability.

On the other hand, there is also room for the counterfactual basis to do some independent work. As an independent basis of liability, if the would-be rescuer makes the victim worse off than he would have been had there been no rescue attempt, then the rescuer is liable no matter that he merely allowed, and did not (fully) cause, the death. The baseline, in such a case, is not victim-in-water-about-to-drown-anyway; rather, the baseline is, victim-in-water-about-to-be-rescued-by-someone-else.[197] Likewise, in cases of double prevention, such as D preventing a rescuer R from effecting the rescue of victim V,[198] D's prevention is a mere allowing of V to drown, not a drowning (or a causing of death) of V. Even so, D is responsible for V's death on a purely counterfactual basis, the baseline being V-about-to-be-rescued-by-R.

The upshot is that causation and counterfactual dependence compete evenly in the class of cases considered here. Sometimes, it is causation that attracts liability, sometimes it is "making a difference" in the way captured by counterfactual dependence, and sometimes it is either, or both together, that is needed for liability.

E. *Limitations on Negligence Liability for Harms Not "Within the Risk"*

If the causal relation takes whole events as its relata, and if

---

197.   Thus, in *Zelenko v Gimbel Bros.*, 287 N.Y.S. 134, 135 (Sup. Ct. 1935), the defendant was held liable for undertaking to rescue the plaintiff by isolating the plaintiff in an infirmary for six hours; the court's conclusion was based on the counterfactual, "If defendant had left plaintiff's intestate alone, beyond doubt some bystander . . . would have summoned an ambulance." *Id.*

198.   *See supra* text accompanying note 124.

counterfactual dependence takes aspects of whole events (or "facts") as its relata, then the legal doctrine of harm within the risk demands a counterfactual and not a causal test. That is indeed my temptation here. When lawyers say, "The act that was negligent may have caused the harm, but the negligence itself did not cause the harm," they are misspeaking. Acts like other events are the particulars that can stand in singular causal relations. Properties (or aspects) of acts, instances of properties of acts (or "tropes"), facts about acts, do not stand in such causal relations. So when we ask whether some aspect of an act that made the act negligent—such as practicing medicine without a license—caused the harm, we must be asking something else.

The something else we are asking is the counterfactual question: Would the absence of the property had made any difference to the existence of the harm? For example, would having a license to practice medicine have prevented the harm because, say, the surgeon lacked just those skills that medical training would have supplied? Under a properly formulated harm within the risk test, this counterfactual question would be asked on its own, without any pretense of being a causal question.

For me, however, this whole discussion is academic in a pejorative sense, for the entire harm within the risk limitation is a mistake, root and branch. Whether it is a causal doctrine, a counterfactual doctrine, or something else, thus does not much matter. It is a doctrine that should not exist, and may well be on the way out. Even if it were the case that it were equally sensible to ask either, "Did the unlicensed aspect of the doctor's act cause the harm," or "But for the fact that the act was unlicensed would there have been the harm to the patient?," yet defendants should be liable, assuming they were negligent, no matter what the answer to these questions might be. That counterfactuals have a role to play here is thus true only in a possible world where the doctrine makes more sense than it does in this world.

F. Symmetrically Overdetermined Concurrent Cause Cases Where One Cause Is Natural

Skeptics about the factual bases for the discriminations the law calls "causal" have long focused on these cases.[199] How can there be a causal distinction, they ask, between a fire of natural origin and a fire of culpable human origin such that when the defendant's fire joins the latter

---

199. *See, e.g.*, Henry W. Edgerton, *Legal Cause*, 72 U. PA. L. REV. 211, 214–16, 241–43 (1924). Contemporary economists such as Guido Calabresi, Steve Shavell, and Landes and Posner have followed Edgerton here. *See* MICHAEL MOORE, PLACING BLAME: A GENERAL THEORY OF THE CRIMINAL LAW 337 n.26 (1997).

he is liable, but when it joins the former he is not? Surely the physical processes are, or at least can be, the same no matter what the origin of the other fire. So how could the defendant's fire be a cause of the harm, or not, depending on nonphysical properties like moral culpability?

The skeptics are right. This is not, and cannot be, a causal distinction. What is plainly doing the work here is a noncausal, counterfactual test which asks whether the defendant's action made any difference vis-à-vis a baseline of natural circumstance. More specifically, the intuitions that guide these cases are (1) a sense that there is a morally significant baseline in what was going to occur naturally, and (2) that the counterfactual, "but for the defendant's action, the plaintiff would have lost his house to nature," is true. The baseline judgment in (1) is crucial here, because that is what distinguishes these cases from others where the second fire is of culpable human origin. In these latter cases, the plaintiff would have lost his house too, only not to natural circumstances, but to another, perhaps equally culpable, human choice.

It is only this kind of counterfactual baseline judgment that has a prayer of justifying nonliability in the cases where the other sufficient cause is of natural origin. Yet, like the harm within the risk doctrine, this doctrine too is not justified at the end of the day. The baseline on which it depends is not very persuasive. If it matters that the plaintiff was going to lose his house anyway, why should it matter whether that loss would have been due to nature, innocent human misadventure, or culpable human choice? Moreover, although this issues some promissory notes only partly paid at the conclusion of this Article, the defendants in these cases do cause the plaintiffs harm. On the corrective justice principle that we should pay for the harm we culpably cause to those who are themselves blameless, the defendants should be liable for these harms even though such harms would have happened anyway.

G. *Asymmetrically Overdetermined Concurrent Cause Cases*

These cases as well pit counterfactual rationales rather directly against cause-based rationales. When a court finds liability for a minor wound that together with a mortal wound produces death through loss of blood,[200] it is marching under the banner of causation. The minor wound

---

200. People v. Lewis, 57 P. 470, 471 (Cal. 1899).

was a cause of the death, even though that death would have happened anyway. When a court denies liability for negligently maintained drainage wickets on the ground that the flood was so large that it would have exceeded the capacity of the drainage wickets even if they had been properly maintained, it is marching under the colors of counterfactual dependency.[201] The split in legal authority is wholly due to a divergence in rationale along these lines. Once one sees that causation and counterfactual dependency are not the same thing, one can at least see this split in its proper light.

My own view, of course, is that the culpable causation of harm should be sufficient for liability. The fact that the harm would have happened anyway, even without the defendant's action, should not change this result. That again is a point to be argued for in the conclusion.

H. Limitations on Damages in Preemptive Overdetermination Cases

Courts that limit damages collectible from a preempting-harm-causer often proceed under false colors. Often they talk as if this were a cause-based limitation: When the defendant's fire arrives first and burns the house to the ground, and thus preempts the ability of a second fire to have done so, the defendant is often said to have caused only the loss of the use of the house in the interim between the two fires. Once one divorces causation from counterfactual dependence, one can dispense with this fiction. This damage limitation is not cause-based. It is wholly based on an independent, counterfactual judgment about what would have happened to the plaintiff's house had the defendant's fire not destroyed it. Again, the choice is starkly put: What matters more to our duties to compensate, harms we have caused or harms that would not have occurred without our action? And again, my choice is generally for the former with the caveat that sometimes the characterization of a state of affairs as a harm will involve some counterfactual judgments, particularly in the offsetting benefit cases.[202]

Puzzlingly, I find more work for counterfactual judgments to do in the criminal law analogy here, the necessity cases. In *Dudley & Stephens*,[203] the defendants in my view were justified in killing and eating the cabin boy as the only means to their own survival. The availability of this "balance of evils" justification does not turn on any lesser causal contribution to the cabin boy's death, as I once argued.[204] Rather, the justification is

---

201.   City of Piqua v. Morris, 120 N.E. 300, 300 (Ohio 1918).
202*.   See supra* note 31.
203.   The Queen v. Dudley & Stephens, 14 Q.B.D. 273 (Eng. C.A. 1884).
204.   Michael S. Moore, *Torture and the Balance of Evils*, 23 Isr. L. Rev. 280 (1989), *reprinted in* Moore, *supra* note 199, at 699, 700–01.

only available because of the truth of the counterfactual judgment that the cabin boy would have died soon anyway even if the defendants had not killed him. Such counterfactual judgments thus join certain causal and other judgments as the basis for allowing balance of evils justifications.[205]

V. CONCLUSION: THE DESIRABILITY OF GENERALLY SUBSTITUTING COUNTERFACTUAL BASELINE TESTS FOR CAUSAL TESTS OF LEGAL LIABILITY

There is a long-standing debate amongst moral and legal theorists as to the proper trigger for desert-based legal sanctions in contracts, torts, and criminal law. The debate is usually framed as a two-sided affair, between those who think one's blameworthiness is wholly a function of subjective features such as choice and intent and those who think that causation of harm to another adds to one's blameworthiness. Now we can see that the debate is at least a three-sided affair. Since counterfactual dependency differs from causation, we need to add the counterfactual criterion of having our actions make a difference, vis-à-vis some baseline, to the list of possible desert-determiners.

Some of the examples given in Part IV above may seem to be just the tip of an iceberg. The tip consists of the isolated instances where the law rightly imposes liability based not on causation of harm, but rather on the counterfactual question of whether one's actions did or could have made a difference. The iceberg itself would be a generalized counterfactual test, one generally supplanting any causal test for liability. Perhaps the sine qua non test, although lousy as a test of causation, can be justified explicitly as a general counterfactual limitation on that for which we are

---

205. Although why they do so is something of a puzzle. My earlier taxonomy of various causal discriminations, the act-omission distinction, and the doctrine of double effect, was done in terms of a construal of deontology that I still find plausible: Categorical moral prohibitions have as their objects certain actions like torture or killing, when intended by the actor. MOORE, *supra* note 199, at 698–705. Foreseen but not intended consequences, failures to prevent rather than cause, allowings rather than doings, etc., all take one out of deontology to the land of permissible consequential justification because in all such cases one is not intentionally doing the action morality categorically forbids. The permission for consequential calculation based on the counterfactual judgment that the victim was about to die anyway does not fit easily here. In such cases a defendant intends to kill and does an act fully causing death (i.e., he kills). Yet the fact that the victim was about to die anyway licenses an otherwise impermissible consequentialist justification. Such counterfactuals must enter deontology by way of an exception to the content of categorical norms, not by way of the scope or object of such norms. For the difference, see *id.* at 706–07.

morally responsible.

To make this more precise, we should distinguish four possible combinations of counterfactual dependence with causation as desert-determiners. The first is where the defendant does not cause some harmful state of affairs to come about, and yet the defendant did "make a difference" vis-à-vis the occurrence of that harmful state of affairs (where "making a difference" involves (1) the truth of the counterfactual, "if not c, then not e" and (2) e not being the same as some morally appropriate baseline state). Omissions and allowings are examples of this first combination of cause and counterfactuals, for in neither of these cases does the defendant's omission or action cause some harm. Yet if some act by the defendant would have made a difference (the omission cases), or if some act by the defendant did make a difference (the allowing cases), then the defendant is morally responsible for the harm and should be legally liable on any theory of liability based on moral responsibility. The presence of counterfactual dependence, in such cases where causation is absent, is sufficient for responsibility. If I fail to throw a rope to V, to whom I owe a duty, and if that throw would have saved V, I am responsible for V's death; if I tackle a lifeguard about to save V, and if, absent my tackling, the lifeguard would have saved V, I am responsible for V's death.

To be sure, moral responsibility based on the counterfactual dependence of some harm h on some act or omission will differ somewhat from moral responsibility based on causation of h by action. The moral duty we breach in the latter case may well be an agent-relative duty, whereas in the former case the duty breached will only be a consequentialist duty. This means that such counterfactual-based responsibility can be reduced or eliminated by sound consequentialist justifications—I may omit to throw the rope to save one in order to save two others, I may turn off the respirator allowing one to die in order to save the life of another with greater medical promise. Responsibility based on counterfactual dependence is thus both lesser and different from responsibility based on causation.

The second and third possible combinations of cause and counterfactuals are easy cases of responsibility. Where my act c causes some harm h, and h counterfactually depended on c, I am morally responsible for h. Where my act c does not cause h, and h does not depend counterfactually on c, I am not responsible for h (although I may be culpable and therefore punishable in criminal law).

The most interesting combination is the fourth possibility: My action c causes some harm h, but h did not counterfactually depend on c. As we have seen, there are some legal doctrines relieving an actor of responsibility for h in these circumstances. (These are the asymmetrical
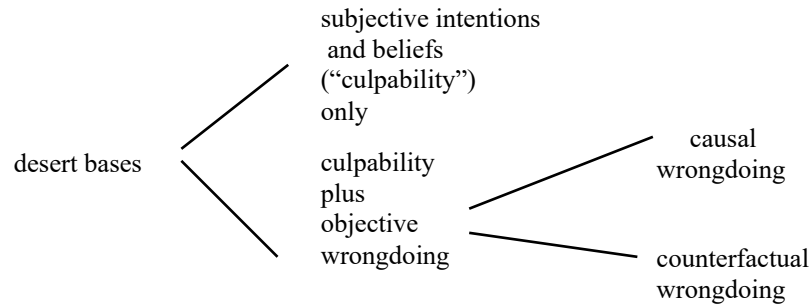
1268

concurrent cause cases, the symmetrical concurrent cause cases where one sufficient cause is a natural event, the damage limitation rule in cases of preemptive causation). If the general principle behind these doctrines were accepted, then we could say that the absence of counterfactual dependence was sufficient for nonresponsibility—or, equivalently, that counterfactual dependence was necessary to responsibility. Conjoined with the result in the first possible combination, this would be to say that counterfactual dependence of some harm h on c was both necessary and sufficient for responsibility for h—meaning that causation of h by c is morally irrelevant.

It is the overdetermination cases which most clearly focus this general issue, for in these cases there is plainly causation of the harm and equally clearly that causing makes no difference to how bad the world ends up becoming. As we have seen, the law generally makes the causer liable in such cases, but that hardly settles the question of moral correctness. The moral question is, what matters here, causing a harm or making a difference?

In the better known debate between causal theorists and subjectivists about responsibility, arguments for why causation matters are hard to come by. My survey of my predecessors on this topic made me wonder whether there are any reasons capable of supporting the judgment that causation matters.[206] That it does may be more basic than anything we can adduce in support of it. So I am not sanguine about what can be said here. Still, the relation between the two debates can be seen as two nodes on a decision tree:

---

206. *Id.* at 196–211.

(38)

```
                          subjective intentions
                           and beliefs
                           ("culpability")
                           only
                                                          causal
                                                          wrongdoing
desert bases              culpability
                          plus
                          objective
                          wrongdoing
                                                          counterfactual
                                                          wrongdoing
```

One gets to the second node—deciding between causation and counterfactual dependence as desert-determiners—only after one has rejected the purely subjectivist branch at the first node of the tree. So it is possible that whatever arguments there are that are capable of selecting objective wrongdoing as a desert-determiner (over subjectivism) are also capable of more particularly selecting the causal version of objective wrongdoing.

The main argument that I and others have used against subjectivists about responsibility has been a kind of reductio ad absurdum of an argument subjectivists deploy.[207] This subjectivist argument urges that we lack control over the results of our actions so that our blameworthiness cannot be increased by this factor. Increased blameworthiness for factors over which we lack control would be a kind of "moral luck," and morality, so the argument goes, cannot be so arbitrary as to admit the existence of such bad luck.

The reductio against this is based on the fact that we have no such control over what we intend, believe, or will either, so if this kind of control is necessary for blameworthiness, there is no such thing as moral blameworthiness. This kind of response, while effective against subjectivists, cuts no ice against "counterfactualists." For proponents of counterfactual dependency as a desert-determiner do not rely on some supposed control actors have over what difference their actions make in the real world, a control over results such actors could then be said to lack. Counterfactual theorists are, in this respect, in the same boat with causal theorists, for it is implausible that there is any more control of "what would have happened if . . ." than there is of what actually results from our actions.

The more positive argument that I have directed against the subjectivists

---

207. *Id*. at 233–46.

is based on the epistemic power possessed by the twin emotions of guilt and moral hatred.[208] The general idea is that our emotional reactions, when they are virtuous, are good but not infallible guides to the truth of the moral judgments that such reactions cause.[209] Feeling guilty, for example, can be a good indicator that one is guilty. Against subjectivists, the argument is that there is a large difference in the emotional reactions to failed attempts and uninstantiated riskings, on the one hand, compared to successful attempts and realized riskings, on the other. It is the latter that gets the blood to the eyes, the former generating usually no more than relief at a "near miss."

This argument can be deployed against counterfactualists if we hone in more precisely on what it is that makes us feel so guilty for ourselves and so angry at others. Consider the preemptive overdetermination cases. If you have culpably caused a serious harm to an innocent, does it diminish your responsibility in the slightest that another person stood ready to cause that harm if you did not? True enough, as a preemptive cause your action made no difference, yet that fact seems to make no moral difference. Think how ill it lies in the mouth of a wrongdoer to try to lessen his responsibility by saying, "If I had not done it, someone else would have." Similarly, if you were the preempted cause, should you feel the guilt of the actual doer of the deed? Or, isn't the reaction still one of relief at a near miss: "I almost did a great wrong, but as luck would have it, I didn't—someone (or something) else did."

The principle to which these emotions and judgments point is a principle of "ownership"—in some suitably extended sense, we own the results of our actions. Such results become a part of our history. They write an entry in our moral ledgers, for the good if they are good, otherwise if they are bad. Some, such as Peter Cane[210] and Tony Honore,[211] wish to go further, arguing that our very identity depends on our being responsible for what we cause. Yet if we keep personal identity over time to truly essential properties, surely we could be the person we each are even if, contrary to fact, we hadn't caused some harm; what we cause is too contingent a feature of our lives to be plausibly listed as essential to personal identity.

It is true that often those who speak of "personal identity" do not mean

---

208. *Id.* at 229–32.
209. *Id*. at 127–38.
210. CANE, *supra* note 22, at 57, 106, 117, 185.
211. *See generally* TONY HONORÉ, RESPONSIBILITY AND FAULT (1999).

it literally. Rather, they mean what I have called elsewhere the sense of identity that we each possess.[212] We each do have a sense of the kind of person that we are, what psychoanalysts used to call our ego-ideal. That ego-ideal is impacted by what we culpably cause, for owning up to those items is what does and should shape our sense of who we are. I see this as another way of putting the "ownership" metaphor mentioned above.

However this is put, it is not yet much of an argument. It will appeal only to those who have been horrified, ashamed, or numbingly distressed by some awfulness of which they were the author. Such people know that causing things matters to responsibility in a way that requires no other argument. Those with either better characters or more fortunate opportunity sets will lack the relevant experience that makes this intuitively so plain to the rest of us.

---

212.   MICHAEL S. MOORE, LAW AND PSYCHIATRY: RETHINKING THE RELATIONSHIP 407–09 (1984).