

Geert Keil

Making Causal Counterfactuals More Singular, and More Appropriate for Use in Law

1 Introduction

Unlike any other monograph on legal liability, Michael S. Moore's book *Causation and Responsibility* contains a well-informed and in-depth discussion of the metaphysics of causation. Moore does not share the widespread view that legal scholars should not enter into metaphysical debates about causation.¹ He shows respect for the subtleties of philosophical debates on causal relata, identity conditions for events, the ontological distinctions between events, states of affairs, facts and tropes, and the counterfactual analysis of event causation, and he considers all these issues relevant to law.

Part V of the book deals with the metaphysics of causal relata. In the final analysis, Moore opts for fine-grained states of affairs as the "true" relata of the causal relation, while he concedes that legal liability should rather turn on coarse-grained Davidsonian events. In Part VI, Moore discusses the major philosophical theories of causation, including Humean regularity theories, nomic theories, probabilistic theories and various singularist and primitivist theories. Most attention is paid to David Lewis' influential counterfactual theory. Moore argues that the counterfactual theory founders for several reasons (Ch. 17). He argues, however, that counterfactual dependency is rightly regarded as a "desert determiner" both for moral responsibility and for legal liability, although it is neither a necessary nor a sufficient condition for causation (Ch. 18).² In cases of omissions, for example, the negligent person is often held responsible, both morally and legally, although omissions resist an event-causal analysis. Moore himself favours a singularist account while being non-committal about which

¹ Here is an oft-quoted assessment: "The lawyer cannot afford to adventure himself with philosophers in the logical and metaphysical controversies that beset the idea of cause." (Pollock 1901, 36)

² "[T]hose who mistakenly identify causation as counterfactual dependence can have most of the bottom-line conclusions about responsibility that they want, if they will but recognize counterfactual dependence to be a desert-determiner independent of causation." (CR, 426)

version of singularism is “closest to the legal presuppositions of what causation must be like” (CR, x–xi).

In this contribution, I shall argue that an amended version of the counterfactual theory of causation can be defended against Moore’s objections, that such a theory best captures our pretheoretical intuitions about causation, and that it is the most suitable theory for use in law. Moore maintains that all generalist theories of causation “founder on the same seven arguments that doom the counterfactual theory as well” (CR, x), “so that singularism wins by default” (CR, 507). I agree with Moore about many of the shortcomings of generalist theories. I shall however argue that rightly conceived, the counterfactual theory is a singularist theory, and that it enjoys all the advantages that Moore reserves for such theories.

2 Causation in Law, Science and Common Sense

To begin with, a number of Moore’s general assumptions about the connections between causation in law, in science and in common sense deserve attention. “The central idea that organizes the book”, says Moore, “is that causation as a prerequisite to legal liability is intimately related to causation as a natural relation lying at the heart of scientific explanation” (CR, vii). Hence, jurisprudence would be ill-advised to limit itself to its own homespun conceptions of causation. Now while the role that causation plays in advanced physical science has been a matter of controversy since Russell called it into question, the indispensability of the notion of cause in common sense explanations can hardly be denied. Moore approvingly quotes John Carroll’s dictum that “with regard to our total conceptual apparatus, causation is at the center of the center” (Carroll 1994, 118).³ As to the relation between causation in law and causation in common sense, Moore holds a second continuity thesis: “The law’s implicit concept of causation differs little from a concept of causation embedded in common sense explanations and evaluations of human behavior” (CR, viii).

The picture that emerges is that of a unitary and intuitive concept of causation that plays a pivotal role in common sense, in science and in law. This pic-

³ Davidson has even pleaded for a reversal of the received wisdom: “It is often thought that scientific explanations are causal, while explanations of actions and mental affairs are not. I think almost exactly the reverse is the case: ordinary explanations of action, perception, memory, and reasoning, as well as the attribution of thoughts, intentions, and desires, are riddled with causal concepts; whereas it is a sign of progress in a science that it rids itself of causal concepts.” (Davidson 1990, 22–3)

ture, albeit quite optimistic, has much to recommend it. One reason why the law's implicit concept of causation should be both intuitive and consistent with our descriptive metaphysics of causation, is that citizens must be able to know the conditions for legal liability. If legal liability presupposed a highly technical notion of causation, this constraint would be violated.

At first sight, there is a tension between the intuitiveness constraint and Moore's view that metaphysical subtleties are relevant for establishing the best theory of causation. This tension is relieved by the fact that mastery of a concept does not require being able to analyse it. It is one thing to be able to employ the concept of cause correctly, i.e., to have a good grasp of its extension in a vast number of cases. It is quite another thing to be able to spell out its logical structure, let alone provide necessary and jointly sufficient conditions. It takes a lot of hard and sophisticated work to make explicit the concept of causation implicit in our descriptive metaphysics.⁴

3 The Counterfactual Theory of Causation

As regards relevance to law, the counterfactual theory of causation is not just one theory among many. It underlies the dominant test for causation in many legal traditions, namely the *but for* test, also known as *sine qua non* test or *necessary condition* test. In Moore's words: "Such a test asks a counterfactual question: But for the defendant's action, would the victim have been harmed in the way the criminal law prohibits?" (CR, 83–4)

The idea that a cause is simply a necessary condition in the way the counterfactual account suggests is very intuitive. It is so intuitive that Hume took it to be equivalent with his regularity account, which is a blunder that no freshman student would get away with today:

We 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. (Hume 1748, sc. VII)

⁴ "Descriptive metaphysics", according to Peter Strawson, aims "to describe the actual structure of our thought about the world". It examines "the commonplaces of the least refined thinking", which "are yet the indispensable core of the conceptual equipment of the most sophisticated human beings. It is with these, their interconnexions, and the structure that they form, that a descriptive metaphysics will be primarily concerned" (Strawson 1959, 9–10).

Far from putting the first definition “in other words”, the second sentence introduces a distinctively new idea. Today’s standard version of the counterfactual theory was presented by David Lewis in 1973. Here is Lewis’ original definition:

If *c* and *e* are two actual events such that *e* would not have occurred without *c*, then *c* is the cause of *e*. (Lewis 1973b, 167).

I neglect Lewis’ distinction between causation and causal dependence, because I do not share the main motivation for this distinction, namely his assumption that the causal relation is transitive. Note that in this analysis the question of causal relata is already prejudged in the *analysandum*: The variables “*c*” and “*e*” range over events, and the statement to be analysed is the singular causal statement “event *c* has caused event *e*”. Further note that *c* and *e* have actually occurred, a fact that is implicit in the grammar of counterfactual statements. A statement can only be contrary to fact if the facts themselves obtain. Counterfactual dependence is always stated and evaluated *ex post factum*, in view of events that actually occurred. A future conditional such as “If the wrong wire is cut, the bomb may explode” is hypothetical, but not counterfactual. This trait of the counterfactual analysis completely changes the subject, compared with Hume’s way of putting the problem. This change of subject often goes unnoticed. Hume asked which properties an object or event must possess so that we can infer that another event will occur. He found none, drew his sceptical conclusions and sought a psychological explanation in terms of expectations, associations and habits. The counterfactual theory, on the other hand, asks of two events that actually obtained whether they were related as cause and effect or not. Hume, in asking two predictive questions at once – Will *e* occur? Will *e* be caused by *c*? –, tied the analysis of causation to the problem of induction and thus put the whole debate about the *metaphysics* of the causal relation on the wrong track. If we bracket Hume’s empiricist assumptions, there is no good reason to burden the theory of causality with the problem of induction.

4 The Case for Singularism

The primary task of a theory of causality, I submit, is to make explicit the truth conditions for uncontroversial cases of singular causal statements. In doing so, we start with a given *pair* of events and ask whether they were related as cause and effect. Incidentally, this is precisely the situation in the courtroom. Some damage or harm or other violation of rights has occurred, and the pivotal question is whether the defendant’s conduct was causally responsible for that harm.

Thankfully, what makes most sense in the courtroom makes most sense for a theory of causation in general. What philosophical and legal *theories* of causation add to the factual question of whether *e* was caused by *c* or not is the search for truth conditions for statements of the form “*c* caused *e*”.

The *sine qua non* test indicates the law’s leanings toward the counterfactual theory. The courtroom situation also brings into the open the law’s leanings toward singularism. The court’s primary task is not to find out whether fast driving in general causes accidents or whether smoking causes cancer. The cases to be decided are particular incidents, and the causal claims to be assessed are singular causal statements.

Quite a number of theories of causation pursue different projects, rather than specifying truth conditions for singular causal statements. Process theories and transfer theories try to sort out what distinguishes causal from non-causal processes. Interventionist theories look for recipes, i.e. for manipulations that work as means to bring about certain kinds of effects. The probabilistic theory either fails to give the correct truth conditions for singular causal statements, or does not address the issue in the first place. Lewis’ counterfactual theory is quite special in being unmistakably clear that the *analysandum* of a theory of causation are singular causal statements. Before going into the details we should pause to praise this insight.

What makes a theory of causation singularist, however, is not the nature of the *analysandum*, but that of the *analysans*. Generalist theories may well accept singular causal statements as starting points, but they attempt to analyze them in terms of general relations such as regularities, lawful connections or relations of necessitation between universals. Since there is no generally accepted view about the exact nature of the singularist/generalist distinction, I refer to Moore’s explanation. Moore distinguishes between an extreme and a moderate form of singularism, the “axis of difference between those forms” being “the degree of commitment to there being causal laws whenever there are singular causal relations” (CR, 496). Extreme singularism “holds that singular causal relations can (and even do) exist even if there were no true causal laws that connected types of events of which these causes and effects are instances” (CR, 496). Moderate singularism leaves open whether there are such covering laws. Singularists may grant that such laws exist “because the more pertinent axis here is not accompaniment, but reduction. [...] Singularism can be best seen as the account that does not reduce singular statements to statements of causal laws. Such a non-reductionist-defined singularism can then be agnostic about whether such laws are always present where there exist singular causal relations” (CR, 497).

The view that every singular causal statement implies a strict law that covers the case in question has been held by Kant, Mill, Popper, Hempel, Stegmüller

and Davidson, among others. Davidson has called this view, somewhat clumsily, “the principle of the nomological character of causality” (1970, 208), or, later and less clumsily, “the cause-law thesis” (1995). In Davidson’s version, the cause-law thesis says that if two events “*c* and *e* are related as cause and effect, there exist descriptions of *c* and *e* that instantiate a strict law” (1993, 312–3).⁵ We neither need to know this strict law nor the event descriptions, but they must exist. Strict laws, according to Davidson, are exceptionless generalizations that support counterfactuals and that do not contain *ceteris paribus* clauses or other open-ended phrases.⁶

The cause-law thesis is arguably false. The reason for rejecting it is simply that, for all we know, strict laws of the required kind do not exist. All candidates are falsified by counterexamples. As Cartwright (1983, 46) bluntly puts it: “there are no exceptionless quantitative laws in physics [...]. In fact our best candidates are known to fail”. Two decades before Cartwright’s book *How the Laws of Physics Lie* made the sceptical perspective on laws popular, Scriven had opened a paper with the remark: “The most interesting fact about laws of nature is that they are virtually all known to be in error” (Scriven 1961, 91).

This bold claim is in need of qualification, of course. For the present purpose, the claim should be restricted to those kinds of laws that the cause-law thesis must rely on. Not just any connection between laws and causation suffices for the cause-law thesis. A *causal* law must exist that actually covers the case, causal laws being “laws that do subsume cause-effect pairs” (Cummins 1983, 5). In this paper, I assume without argument that only events, rather than aspects thereof, or facts, or properties, or tropes, can be causally related. Under this assumption only laws about the succession of events can be causal laws, i.e., laws that state, crudely put: “Whenever an event of the type *c* occurs, it is followed by an event of the type *e*”. Such laws make a claim about what regularly happens. Most of the laws of physics do not. For example, the law of the pendulum says something about the relationship between the length of a pendulum and its period of swing. Boyle’s Law makes a statement about the ratio between the pressure, tempera-

5 Or, more precisely: “‘*A* caused *B*’ is true if and only if there are descriptions of *A* and *B* such that the sentence obtained by putting these descriptions for ‘*A*’ and ‘*B*’ in ‘*A* caused *B*’ follows from a true causal law.” (Davidson 1963, 16) Here is Moore’s version: “Singular causal statements are thus unpacked as referring to *instantiations* of causal laws by spatio-temporally located particulars” (CR, 471).

6 “Laws must be true universally quantified statements. They also must be lawlike: they must support counterfactuals, and be confirmed by their instances (these conditions are not independent). To qualify as *strictly* lawlike, they should contain no singular terms referring to particular objects, locations or times [...]. Strictly lawlike statements do not contain open-ended phrases like ‘other things being equal’, or ‘under normal conditions’” (Davidson 1995, 265).

ture and volume of a fixed amount of an ideal gas. These laws are laws of *coexistence*, not laws of succession. They say something about simultaneous, or rather timeless, relations between properties or physical quantities.⁷

Hence, rejecting the cause-law thesis does not amount to the bold claim that there are no laws of any kind with any role to play in determining the truth conditions of singular causal statements. The claim is that no *causal* laws of the required kind exist, i.e., no true, exceptionless generalizations that subsume cause-effect pairs. “Scepticism about laws” is an unfortunate expression of this claim, for no distinctly sceptical attitude is required in order to notice that laws of succession which state empirical regularities are plagued with exceptions. True, the philosophy of science is bristling with counter-reactions⁸, but far from disputing the diagnosis that series of events do not exhibit exceptionless regularities, most of these “Yes, but” reactions change the *concept* of law in the face of the problem of exceptions. Woodward, for instance, argues that “if we demand that all genuine laws must be exceptionless, it follows that we know very few laws” (Woodward 2000, 228). So, if we know quite a number of laws, the laws of physics must be conceived differently.

Another often overlooked *prima facie* argument against the cause-law thesis is this: If strict causal laws that subsume cause-effect pairs were available, then these would plausibly yield the truth conditions for singular causal statements. There would be no need to develop alternative theories of causation. It is precisely because such laws are not at hand that philosophers search for alternative theories that analyze causal relations in terms of energy transfer, counterfactual dependence, probabilities, experimentation, etc.⁹

Moore gives the no-causal-laws view short shrift. He addresses one of the weakest defences of this view, an article of Anscombe’s, who indeed fails to mount much of an argument.

⁷ The distinction between laws of succession and laws of coexistence goes back to Mill: “The phenomena of nature exist in two distinct relations to one another; that of simultaneity, and that of succession”, and accordingly they obey two kinds of laws (Mill 1843, 323 [III, v, § 1]). Thereupon, Mill makes a crucially important remark: “[U]nless there [are] laws of succession in our premises, there could be no truths of succession in our conclusions” (Mill 1843, 325). Mill correctly states that no truth about what actually happens can be derived from truths about simultaneous relations alone.

⁸ For an overview, see (Keil 2000, 182–240).

⁹ Lewis once rightly suspected that he did not exhaust the potential of the counterfactual analysis: “Perhaps by ignoring indeterminism I squander the most striking advantage of a counterfactual analysis over a regularity analysis: that it allows undetermined events to be caused.” (Lewis 1973b, 162–3). For the benefits of an indeterministic interpretation, see (Keil 2009, 22–25).

The cause-law thesis is only the strongest version of a generalist theory of causation. For a more moderate form of generalism, a number of options are still left. Moore frames the singularism/generalism distinction in terms of “degree of commitment to there being causal laws whenever there are singular causal relations” (CR, 496). I suggest to draw the distinction between extreme and moderate generalism in this way: *Extreme* generalism claims that *causal* laws exist that *subsume* the cause-effect pairs of ordinary singular causal statements. *Moderate* generalism is not committed to these restrictions. First, there are other kinds of general statements that may back singular causal statements, e.g., statements about dispositional properties, about non-nomic general facts, about non-nomic correlations or about symmetries. Secondly, physics knows other kinds of laws, apart from causal laws of succession. Thirdly, these laws may play other, less crude roles in the truth conditions of singular causal statements than the cause-law thesis imposes on them.

As for the counterfactual theory of causation, it is not clear whether according to Moore’s criterion it counts as a generalist or as a moderate singularist theory. In some places, Moore classifies it as singularist, because “causal laws play no essential role in the truth conditions of the relevant counterfactuals” (CR, 502f.). Irritatingly, however, he claims that the counterfactual theory and generalist theories founder on exactly the same arguments (CR, x). Mostly, he works with a tripartite classification of singularist, generalist and counterfactual theories.

In Lewis’ theory, laws do play an important role in evaluating the relevant counterfactuals. But since there is no consensus about *which kind* of laws must play exactly *what* role in order to make a theory of causality a generalist one, short of *causal* laws that *subsume* cause-effect pairs, it is best to leave the terminological issue open as to whether the counterfactual theory is a generalist theory.

5 Moore’s Criticism of the Counterfactual Theory

Moore raises numerous objections to the counterfactual theory of causation. His discussion consists of two parts. In chapter 16, he discusses the question of what makes counterfactuals themselves true. In chapter 17, he inquires whether causation can be analyzed in terms of counterfactuals. He calls both topics “two distinct parts” of the counterfactual theory (CR, 392). I am not so sure. Moore does not seem so either, for already chapter 16 deals primarily with causation. The reason for this amalgamation is that in Moore’s view, the counterfactual theory of causation makes an identity claim:

On the construal I favor, the counterfactual theory purports to tell us the nature of causation. Thus, it is best construed as an identity claim, namely causation is counterfactual dependence. (CR, 394)

Accordingly, the first group of objections¹⁰ turns on this identity claim:

(1) Overbreadth I

Counterfactual dependence is not sufficient for causation. Many counterfactuals have no causal interpretation. Counterfactual dependence relates things that are not causally related (394–6).

(2) Overbreadth II: Promiscuity

The counterfactual analysis multiplies the number of causes for an event beyond plausibility (396–403). In reaction to counterexamples the theory must admit as causes an immense number of finely individuated events, facts and properties. Furthermore, it must admit absences and omissions as causes, since effects are counterfactually dependent on them.

Two further objections concern basic tenets of Lewis' possible worlds metaphysics and semantics:

(3) Modal realism

The metaphysics of modal realism is “extravagant” and “difficult to accept” (385). Belief in the literal reality of possible worlds is counterintuitive.

(4) Similarity standard

Lewis' metric of overall similarity between possible worlds lacks an independent justification. Lewis gerrymanders his idea of a “closest possible world” to get the desired results, i.e., to make the causal conclusions come out right (401–2). The same holds for Lewis' claim, insufficiently argued for, that all backtracking counterfactuals are false (403–9).¹¹

The last two objections concern specific problems with the counterfactual analysis of causation:

(5) Counterexamples

The counterfactual theory cannot adequately handle many cases of overdetermination and pre-emption. It yields intuitively implausible results in these cases (410–25).

(6) Nontransitivity

“Counterfactual dependence across chains of causes does not, as causation seems to, weaken or peter out” over time (399).

¹⁰ I slightly rearranged and renumbered Moore's list (see CR, 508–12), which comprises objections to all generalist theories of causation, not just to the counterfactual theory.

¹¹ “Backtracking” counterfactuals are those where counterfactual dependence goes back through time, i.e., where earlier events allegedly depend on later events.

For the present purpose, not all of these challenges need to be addressed in detail. As to Lewis' modal realism, I claim that the counterfactual analysis of causation does not depend on this metaphysical foundation. In everyday life we constantly engage in counterfactual reasoning about causes and effects without being committed to the belief that possible worlds exist, or are real.¹² Neither common sense nor jurisprudence need take a stance on the metaphysics of modal realism.

Pre-emption cases are commonly regarded as a challenge to the counterfactual analysis. Moore discusses the historical case where two fires burned their way to the plaintiff's house. Both fires were big enough to burn down the house, but the first was faster, so that when the second arrived, there was no house to burn down (see CR, 412). In such a case, Moore holds, "our causal intuitions are very clear: a pre-emptive cause causes all of the damage, and a pre-empted 'cause' is no cause at all" (CR, 419). I agree. It is anything but clear, however, that a properly refined counterfactual analysis cannot handle these cases and yield the same result. In the vast recent literature on overdetermination, trumping and pre-emption, there are promising efforts to handle these alleged counterexamples. From a truly singularist perspective, the method of choice is the "fragility strategy" (see below, sct. 7).

My strategy for discussing the remaining objections follows a simple maxim: Do everything in the correct order!

6 Do Everything in the Correct Order!

Judging from Lewis' work, the correct order for developing a theory of causality seems to be this: First elaborate a general possible world semantics, analyse counterfactual dependence in terms of similarity between possible worlds, devise a metric for overall similarity, and then apply the results to the analysis of causation. Moore, too, thinks that the counterfactual theory of causation has to accomplish all these tasks. In thinking this he ignores the fact that Lewis' philosophical agenda is shaped by many motives that have nothing to do with causation, let alone with the role that causation plays in legal liability.

Let us take a step back. I suggested above that the primary task of a theory of causality is to make explicit the truth conditions for uncontroversial cases of sin-

¹² The issue is partly terminological. Lewis holds that possible worlds "exist" in an innocuous sense, though not "actually". The *actual* world is distinguished from other possible worlds by the fact that we inhabit it (see 1973a, 84–6). Roughly speaking, in Lewis' modal realism, "being actual" plays the role that "being real" or "exist" play in other metaphysical systems.

gular causal statements, such as “The cause of her death was poisoning” or “The explosion caused the bridge to collapse”, or “The cause of the capsizing was the shift of cargo”. If there are such uncontroversial cases, then their being uncontroversial should supervene on some fact. It is contentious, however, whether the factual basis of that consensus consists of anything over and above our descriptive metaphysics of causation as reflected in our linguistic intuitions and the causal judgements we actually make. As competent speakers of a natural language we share a number of implicit assumptions about causation. Among them there are bedrock intuitions that no philosophical analysis can overturn without changing the subject.¹³ The method of counterexamples is telling in this respect: The very fact that philosophical theories about causation are being tested by citing and trying to accommodate counterexamples shows that we expect philosophical theories to account for our intuitions, rather than the other way round. There seems to be a certain class of core concepts in our descriptive metaphysics, which may include “causation”, “knowledge”, and “truth”, where we are more confident of our linguistic intuitions than of the correctness of any philosophical definition or theory that purports to analyse those concepts. We are firmly convinced that Gettier cases are not cases of knowledge, and if our judgment that Jones, in Gettier’s original case, does not know who will get the job clashes with the best available philosophical definition of “knowledge”, then the definition has to go, not our judgment.¹⁴

Moore also assumes that there are “firm intuitions about causation” (CR, 507), even “bedrock intuitions” (403), though he does not enumerate them. I submit the following list:

- Causation is a relation between two items, called “the cause” and “the effect”.
- Causation is something that happens in space and time.
- Causation does not work backwards in time: Effects cannot precede their causes.¹⁵ (The additional issue as to whether cause and effect must be tem-

13 “The first step in providing a [...] definition of the causal relation is to set down the central tenets of our folk theory of causation – the platitudes about causation which are common knowledge among us.” (Menzies 1996, 97)

14 What holds for “knowledge” and “causation” does not seem hold for natural kind terms like “gold” or “tiger”. Chemical elements and biological species, it is said, have an essence, constituted by an underlying microstructure, that empirical science can study and that competent speakers may be wrong about. While linguistic intuitions trump philosophical theories, scientific essences trump linguistic customs. Moore resolutely sides with Kripke’s and Putnam’s scientific essentialism and even extends it to “social kinds” and “legal kinds”.

15 Note that the intuition against backward causation is neutral about the proper direction of explication. It does not rule out a causal theory of time.

- porally contiguous or whether they may be separated by a temporal gap is contentious.)
- The Humean intuition about *spatial* contiguity is not a bedrock intuition, but a little more fragile. Action at a distance has never been regarded as absurd as backward causation.
 - The causal relation is asymmetric and irreflexive. Of course, ordinary speakers need not know these technical terms. These logical properties of the causal relation are best seen as being abstracted from our actual causal judgments, which in turn are informed by the bedrock intuition of temporal succession. Nothing can be the cause of itself, and effects do not bring about their causes.
 - There is some kind of necessity involved. Mere co-occurrence does not suffice for causation. This modal intuition is firm, but notoriously hard to specify.
 - Finally, there is the commonplace that causal chains can weaken or thin out over time, in the same way that ripples from a stone dropped in a quiet lake fade out towards the lakefront. This commonplace, which implies that the causal relation is not transitive, is a particular challenge to philosophical theories of causality, which teach, in the majority, that the causal relation is transitive.

Given these commonsense intuitions, some of them bedrock, some of them soft rock, what can we reasonably expect from a philosophical theory of causation? According to Moore, the counterfactual theory aims at a “metaphysical reduction” of the causal relation to the relation of counterfactual dependence:

On the construal I favor, the counterfactual theory purports to tell us the nature of causation. Thus, it is best construed as an identity claim, namely causation is counterfactual dependence, either direct or linked down a chain. As an identity claim, the theory is committed to an equivalence corollary, namely, anything predicable of causation is equivalently predicable of counterfactual dependence, and vice versa. (CR, 394)

Now if both relations are meant to be identical and hence have to share all their properties, then the counterfactual theory is doomed to fail from the outset, for the simple reason that there are lots of non-causal counterfactual conditionals. Counterfactuals like “If today were not Tuesday, then yesterday would not have been Monday”, or “If the word ‘dog’ had one letter more, it would have four letters” have no causal reading. This argument from non-causal counterfactuals is well-known from Kim, and Moore restates it as the objection of overbreadth: “Counterfactual dependence relates things that causation does not” (CR, 396). The very expectation that the counterfactual theory must be able to rule out

cases of non-causal counterfactual dependence, however, rests on the identity claim that Moore has foisted on the counterfactual theory. The identity claim is hopeless, and Lewis does not make this claim, as his original definition witnesses: “If *c* and *e* are two actual events such that *e* would not have occurred without *c*, then *c* is the cause of *e*.” (Lewis 1973b, 167) The statement to be analyzed is the singular causal statement “*c* is the cause of *e*”. The variables *c* and *e* ranging over events in the analysandum, Lewis need not bother about Mondays and Tuesdays, or the spelling of words, or about Verdi and Bizet being compatriots¹⁶, or about kangaroos without tails.¹⁷ These troublemaker cases need not be considered in the first place, because they do not belong to the class of counterfactuals that could fit the bill, causation being a relation between distinct events.¹⁸ Admitting only events as causal relata rules out many alleged counterexamples, the distinctness condition rules out even more, Davidsonian individuation of events (see below) rules out the rest.

From Moore’s perspective, such a “stipulated narrowing down of the class or relevant counterfactuals” (CR, 395) is illegitimate. But the problem is self-created. Neither Lewis nor any other serious proponent of the counterfactual theory endorses the identity claim.

Moore also considers it illegitimate to add the constraint about temporal succession:

The persuasive power of the truism that you cannot change the past is already loaded with the asymmetry of the causal relation precisely because causation does not work backwards in time. (CR, 404)

Any additional restrictions on the counterfactual analysis “have to be shown to follow from the counterfactual nature of causation”, according to Moore (CR, 404). Again, they have to if we impute to the counterfactual theory the assumption that the causal relation is simply the relation of counterfactual dependence. Once we give up this unreasonable demand, we need not expect that the relation of counterfactual dependence alone yields all the commonplace features of causation. Once we give it up, we can grant that our intuitions do not and need not follow from the counterfactual analysis. If our preanalytic descriptive metaphy-

16 If Bizet and Verdi had been compatriots, would Bizet have been Italian or would Verdi have been French? (see Quine 1950, 15)

17 If kangaroos had no tails, would they topple over or would they walk around on crutches? (see Lewis 1973a, 8–9)

18 “Two events are distinct if they have nothing in common: they are not identical, neither is a proper part of the other, nor do they have a common part.” (Lewis 1973b, 212)

sics of causation is such that, among other things, effects cannot precede their causes, so that we know in advance that no analysis is correct that disregards this feature, then we do not need a general argument to the effect that all backtracking conditionals are false, which Moore accuses Lewis of not having provided.

It is as clear as daylight that the relation of causation is not identical with the relation of counterfactual dependence. The existence of countless non-causal counterfactuals suffices to refute this claim. All there has to be is a partial overlap: On the counterfactual theory, the class of causally related pairs of events is a proper subset of the class of counterfactually dependent items. Causation is not counterfactual dependence *simpliciter*, but at most counterfactual dependence between distinct events that follow one another. This is how Moore's first charge of overbreadth of the counterfactual theory should be countered.

Not only do many counterfactuals lack a causal interpretation, many of them do not even seem to have a truth-value. Quine's Verdi-Bizet example was meant as a *reductio ad absurdum* of the assumption that counterfactual conditionals have a truth-value. But Quine was overstating the point. What the example shows is that *some* counterfactuals are too indefinite to have a truth-value. Now, if only a subset of counterfactual conditionals is truth-apt, it suggests itself to delineate and examine that set, i.e., to determine what distinguishes those cases from the others.¹⁹

It should have transpired by now what I mean by "doing everything in the correct order":

First step: Make up your mind what the primary task of a theory of causality is. I suggest that it is to provide truth conditions for uncontroversial cases of singular causal statements. If there are such cases, their being uncontroversial should supervene on commonsense intuitions about the nature of causality.

Second step: Collect those intuitions and commonplaces.

Third step: Turn them into constraints for the class of counterfactuals that a counterfactual analysis of causation has to consider.

There are remarkable concessions by Lewis that point in the same direction. In particular, he despairs of finding an independent justification for the appropriate similarity relation:

[W]e must use what we know about the truth and falsity of counterfactuals [...] to find out about the appropriate similarity relation – not the other way round. (Lewis 1979, 43)

¹⁹ Elsewhere I have tried to specify five measures to ensure the truth-aptness of counterfactual conditionals. See Keil 2000, 431–457.

It is fair to discover the appropriate standards of similarity from the counterfactuals they make true, rather than *vice versa*. (Lewis 1973b, 211)

Moore objects that Lewis lacks a general standard for the similarity of possible worlds that the counterfactual theory of causation can apply. Lewis, however, does not claim that we have independent resources to develop such a similarity metric. Rather he recommends what may be called the method of clear cases. He seems to think that we are more confident about, or that we have better epistemic access to, the truth of certain counterfactual judgements than to general similarity standards that purportedly fix their truth conditions. With regard to causal counterfactuals, the method of clear cases concerns the second of the three steps I suggested: In order to collect intuitions and commonplaces about causation, we first gather a sample of clear cases of causal counterfactuals (such as: “If I had not opened my hand, the pencil would not have fallen on the table”), then we figure out what these cases have in common, and thereby abstract the commonplace features of the causal relation from the sample.

“Collecting clear cases” of singular causal statements and causal counterfactuals does of course not mean collecting only *true* statements. This demand would indeed be circular. The counterfactual “Had Michael Jackson not been administered an overdose of Propofol, he would not have died that day” is a clear case of a statement amenable to causal interpretation, since it relates entities of the right kind. Whether it is true is another matter.

7 Singular Causes as Necessary Conditions?

Now I wish to defend the claims that counterfactual dependence between events is a singular relation, that the counterfactual theory of event causation is best construed as a singularist theory, and that only a singularist construal brings out its superiority over the INUS and NESS theories.

The biggest obstacle to the singularist construal is the myth that the antecedent of a causal counterfactual cites a necessary condition. Moore shares this myth: “On the counterfactual theory, for *c* to be the cause of *e* *c* must be necessary for *e*” (CR, 354). This cannot be true because strictly speaking, calling an event a necessary condition is a category mistake. The logical notions of necessary and sufficient conditions are not defined for dependency relations between events.²⁰

²⁰ “Obviously, given the usual definition of necessary conditions, namely through implication

On a less pedantic and more charitable reading, calling the striking of the match a necessary condition for its lighting is easily translated into the general statement that every time a match lights, it has been struck. Calling the striking a sufficient condition for the lighting is easily translated into the general statement that every time a match is struck, it lights. Now my claim is that the singular counterfactual “If this match had been struck, it would have lit” is anything but a singular instance of some general necessary relation.

In what follows I employ the standard notions of “necessary condition” and “sufficient condition”.²¹ One might object that the dependency relation to be analysed is *causal* necessity and/or sufficiency, and that these relations are different from *logical* necessity and sufficiency. As long as we are concerned with the logic of conditionals, this comment is out of place. The distinction between causal and logical modality concerns the non-logical question *by virtue of what* does the conditional become true. Take the textbook example: “If it rains, the street gets wet”. In calling rain a sufficient condition for wetness and wetness a necessary condition for rain, the logician *stipulates* that the conditional is true. If, on the other hand, we assert a *causal* connection between rain and wetness, the truth of the assertion depends on whether the regularity actually holds. Arguably, it does not. (Think of Wimbledon’s Centre Court that gets covered as soon as it starts raining.) What distinguishes “causal” from “logical” necessity is not the meaning of the modal terms, but rather the source of the alleged necessity. Analyses of the causal relation in terms of necessary and sufficient conditions do not invoke some special relation of dependency between particular events, rather they express the fact, if it is one, that events of a certain *kind* occur in constant conjunction. There is no second, independent, non-logical sense of “necessary and/or sufficient conditions” that would be explicable without the universally quantified conditional “Whenever *c*, then *e*”.

The distinction between causal and logical necessity tends to blur the all-important difference between the related items. Events can be causally related, but not logically. For statements, the converse holds. I do not only find fault, as Danto does, with the naïve application of the idiom of necessary and sufficient conditions to relations between events. There is a deeper issue related to that

and denial, it is strictly a mistake in category to speak of *events* as necessary conditions. I do so in fact only in a semantical respect, viz. as instantiating a description.” (Danto 1973, 94)

²¹ “If a conditional of the form ‘If *p*, then *q*’ is true, then the state of affairs expressed by *p* is said to be a sufficient condition for the state of affairs expressed by *q* and, correlatively, the state of affairs expressed by *q* is said to be a necessary condition of the state of affairs expressed by *p*.” Entry “necessary and sufficient conditions” in the *Oxford Companion to Philosophy*, ed. by Ted Honderich, Oxford 1995, 608.

fault, raised by Mill's objection that the striking of the match is not sufficient for its lighting. The logic textbook calls the striking a sufficient condition for the lighting, insofar as it assumes the conditional "every time a match is struck, it lights" as true. Mill, in denying that a particular striking was sufficient for a particular lighting, denies that the universally quantified conditional is empirically true. Further conditions have to be fulfilled: The match must have been dry, there must have been enough combustible air, etc., hence the striking cannot have been the cause of the lighting. No striking has ever caused a lighting, since, according to Mill, only "total causes" are causes proper: "The cause, then, philosophically speaking, is the sum total of the conditions [...] which being realized, the consequent invariably follows".²² Mill's comment is related to the sceptical view on the laws of nature discussed above. While the logician simply stipulates the truth of the conditional, the causal theorist makes an empirical claim about what always happens, and if no true, exceptionless generalizations that subsume cause-effect pairs should be available, then "nomic sufficiency theories", as Moore calls them, are in trouble. They are rightly "accused of parochialism in an indeterministic universe" (CR, 476).

Such theories are in trouble for another reason: On the one hand, no exceptionless causal laws seem to be available, on the other hand, the *sine qua non* test yields too many causes. Being a necessary condition does not seem to be sufficient for being a cause. Both Mill and the extensive discussion on causation in the law focus more on the second problem than on the first.

Sufficiency theories are in trouble, but not helpless. Mill distinguished between "total causes" and "partial causes", recent philosophy of science distinguishes "proximate" from "distal" causes, Hart/Honoré and Mackie refined the sufficiency theory to an ingenious combination of necessary and sufficient conditions. Causes are neither necessary nor sufficient for their effects, rather they are INUS conditions: insufficient but necessary parts of conditions which are themselves unnecessary but sufficient for the effect. The acronym NESS, introduced by Richard Wright and more widespread in legal theory, means the same: a cause is a necessary element of a sufficient set.²³ INUS theorists often introduce a special kind of "weak necessity", sometimes called "necessity-in-the-circumstances", and claim that "a cause is ordinarily taken to be necessary in the circumstances for its effect, and perhaps sufficient in the circumstances as well" (Mackie 1974, 50). I insist that this hybridization of the relation of causation

²² Mill 1843, 332 (Bk. III, Ch. V, § 3)

²³ Usually, Mackie (1965, 245) is credited for having invented the INUS analysis. Actually Hart and Honoré (1959, 112ff.) were the first to articulate it. Wright (2013) emphasizes the differences between the INUS account and his NESS account.

and the logical relation of being a necessary and/or sufficient condition does nothing to alleviate but rather confirms the worry about categorical confusion.

As to the problem that the events we call causes fall short of being sufficient conditions, Mill introduces a pragmatic element into his analysis. Often we cite something as “the cause”, he says, which is only a partial cause. Only “philosophically speaking”, the cause is “the sum total of the conditions [...] which being realized, the consequent invariably follows”.²⁴ Outside of philosophy, it is not. Mill mocks “the absence of any scientific ground” and “the capricious manner in which we select from among the conditions, that which we choose to denominate the cause”.²⁵ The *selection problem*, as it has been labelled, is notorious both in the philosophical debate as well as in the literature on causation in the law.

Again, Mill’s objection that alleged causes are not sufficient conditions rests on the confusion between the singular relation of causation and the general relation of sufficiency. Causation is a relation between events, sufficiency is not. Davidson got things straight:

The inevitable comment (since the time of Mill anyway) is that the striking may have been *part* of the cause, but it was hardly sufficient for the lighting since it was also necessary for the match to be dry, that there be enough oxygen, etc. This comment is, in my opinion, confused. For since this match was dry, and was struck in enough oxygen, etc., the striking of this match was identical with the striking of a dry match in enough oxygen. (Davidson 1969, 172)

The striking of the dry match in the circumstances did cause the lighting, notwithstanding the fact, or rather the counterfactual, that if the match had not been dry, or if the obtaining circumstances had been relevantly different, the striking would not have caused the lighting. The lighting would not even have occurred then, and *a fortiori*, it would not have been caused. Davidson makes clear:

The difficulty has nothing to do with Mill’s sweeping view of the cause, but attends any attempt of this kind to treat particular causes as necessary or sufficient conditions. (Davidson 1967, 150 – 1)

Davidson insists that the match did not have to be dry *in addition to* its being struck, rather its striking *was* the striking of a dry match. It should be noted that this way of putting it takes it for granted that causal relata are “coarse-

²⁴ Mill 1843, 332 (Bk. III, Ch. V, § 3).

²⁵ Mill 1843, 329 (Bk. III, Ch. V, § 3).

grained” events, rather than Kimian or Lewisian events.²⁶ What I suggest in order to make the counterfactual analysis more singular is basically to do the analysis with Davidsonian whole events rather than with poorly individuated *aspects* of events or property-exemplifications. The striking of the match that caused the lighting was a dated, unrepeatably particular with innumerable properties. Lewis calls such events “fragile”, but this way of putting it only obscures the plain fact that if events are spatiotemporal particulars, then no two events with ever so slightly different properties can be numerically identical. Wright argues for “causally relevant instantiated properties” (“abstract events”, “features of possible events”) as causes, and he adds that “the law is not ambivalent between event-causation and aspect-causation, but clearly requires the latter” (Wright 2013, 15). He underestimates the metaphysical complications of this view of causal relata, though. He maintains that abstract relata stand in “successional relations”, but neglects to explain how temporal relations between abstract entities should be possible.²⁷

A strong argument for Davidsonian “whole events” as causal relata is the linguistic fact that singular causal statements are extensional contexts, i.e., contexts that permit substitution of co-referring expressions *salva veritate*. If a blow with a cane was the cause of the wasp’s death, then by the same token, a blow with my grandfather’s heirloom caused the death of the second largest animal in my office. To be sure, not every arbitrarily chosen event description has the same *explanatory* value. Here is where causation and explanation come apart. Causal explanations are intensional while singular causal statements are extensional. Sometimes the mere citing of the causing event helps to explain why the effect occurred, sometimes it does not. The lethal blow with the cane had many unmentioned properties, but it was causally effective regardless of how it was described, and regardless of whether a different blow with different properties would have been lethal, too.²⁸

Linguistically, the inclusion of further properties and circumstances is achieved by the device of demonstrative reference. Davidson does not mention

26 Lewis argues that in order to be causally related, events must not be too “fragile”. A fragile event is one that “could not have occurred at a different time, or in a different manner. A fragile event has a rich essence.” (Lewis 1973b, 196)

27 “A causal law is an empirically derived statement that describes a successional relation between a set of abstract conditions (properties or features of possible events and states of affairs in our real world) that constitute the antecedent and one or more specified conditions of a distinct abstract event or state of affairs that constitute the consequent [...]” (Wright 2013, 19).

28 Moore’s perplexed question, “How can causal explanations be so intensional if causal relations are not?” (CR, 357), indicates that he has not fully grasped Davidson’s point.

this fact, so I must fill in. Singular causal statements refer to pairs of events either via definite descriptions, or via proper names, or demonstratively. In most real-life cases, the reference is demonstrative or even indexical: “The striking of this match caused its lighting”, “Opening my hand caused the fall of the pencil”. Often it is not necessary to use demonstrative articles, because already the definite article, as Frege, Russell and Quine have observed, works as a uniqueness operator.²⁹ In many contexts, the definite article is used in the function of a demonstrative article. And in referring to a particular event demonstratively, its countless properties and the circumstances in which it occurred are included without having to be mentioned or described. Goodman, in his early discussion of counterfactuals, got it right:

Notice especially that our assertion of the counterfactual [sc. “If that match had been scratched, it would have lighted”] is *not* conditioned upon these circumstances obtaining. We do not assert that the counterfactual is true *if* the circumstances obtain; rather, in asserting the counterfactual we commit ourselves to the actual truth of the statements describing the requisite relevant conditions. (Goodman 1954, 8)

It is only because a particular match in particular circumstances is referred to that we accept the counterfactual as true, for in general, strikings are not sufficient for lightings of matches. But in the case at hand, the match *was* dry enough, the atmosphere *did* contain enough oxygen, etc. These circumstances do not get described, but they obtained, and the demonstrative reference exploits their determinacy. Though Goodman does not lay emphasis on the fact that his example sentences contain demonstratives, this fact strikes me as essential. The reader understands that in the given situation the requisite conditions obtained, so that explicit conditioning upon certain circumstances is redundant.

Danto and Davidson are right to claim that particular events cannot be necessary or sufficient conditions. And if the causal relation relates events, then causes cannot be necessary or sufficient conditions for their effects to occur. To think otherwise is to commit a category mistake. Furthermore, a singular cause is not any combination of necessary and sufficient conditions; the NESS analysis as such obviously does not remedy the category mistake.

Let us apply this result to cases of preemptive overdetermination, which seem so difficult to handle for the counterfactual theory. Remember the case of the two fires: “Our causal intuitions”, says Moore, “are very clear” in this

²⁹ In appropriate contexts, terms like “the river” work as “degenerate demonstrative singular terms” (Quine 1960, 102). “Now *the*, when it is strictly used, involves uniqueness.” (Russell 1905, 481)

case: the first fire caused “all of the damage”, and the second fire, the “pre-empted ‘cause’ [was] no cause at all” (CR, 419 and 412). Agreed. Our intuitions are correct, and the counterfactual theory delivers the goods: Both *c* and *e* occurred, and had the first fire been absent, the effect would not have occurred. Those who object that the house would have been destroyed nonetheless forget that the variable “*e*”, both in the singular causal statement to be analyzed and in the counterfactual analysis, has a particular, concrete event as its value. No ever so slightly different event could have been numerically identical with the actual destruction. A particular causal counterfactual cannot be refuted by claiming that under slightly different circumstances, another destruction of the same or similar kind would have occurred instead. A different counterfactual would have come true then, but this does not eliminate the counterfactual dependence of the actual destruction on the actual fire. If the analysandum is about tokens, the analysis cannot relate types. Here I part company with Lewis. His reluctance to accept the fragility strategy (see 1973b, 193–9; 1981, 247–54) shows that he is not prepared to consider counterfactual dependence between events as a truly singular relation between spatiotemporal particulars.³⁰

The trouble begins if the true counterfactual that *e* would not have occurred absent *c* is rephrased in terms of necessary conditions. Only then the counterfactual provokes the confused comment that the first fire was not necessary, since the second stood in reserve. If the first fire had been absent, the second would not have done *this* job, i. e. the actual job that the effective fire performed, but a slightly different job. Hence, the first counterfactual is not rebutted by the counterfactual that if the first fire had been absent, the second would have stood in reserve to make another causal statement true.

There is an open problem with the fragility strategy, however. While the numerical non-identity of events that do not share all their properties follows directly from Leibniz’ law, *counterfactual reasoning* about fragile “whole events”

30 Lewis’ later account of “causation as influence”, however, contains moderate revisions. In the face of the hard problem of where to “draw the line between *C* not occurring at all and *C* occurring differently in time and manner” (Lewis 2000, 91), Lewis introduces the new technical term of an “alteration” of an event. Alterations are “different versions of the very same event” (90), which allow for new kinds of counterfactual dependence: “Then we look at the pattern of counterfactual dependence of alterations of the effect upon alterations of the cause” (91). While ordinary causation amounts to “whether-whether-dependence”, causation as influence gets analysed as “a pattern of counterfactual dependence of whether, when, and how on whether, when, and how” (91). Still, Lewis does not accept modally fragile Davidsonian events as causal relata. The new technical term precisely allows him to circumvent the problem of identity conditions and to remain “indecisive about the distinction between alterations that are different versions of the very same event and alterations that are different but similar events” (90).

can lead to unwanted results. Not just any relational property of c and e seems to be relevant to their counterfactual dependence, on pain of making the counterfactual conditional trivially true. Presence of oxygen in the surroundings is clearly relevant to the striking of a match, relations to events outside the “light cones” of c and e are clearly not. Generally speaking, it is not the whole history of an event that contributes to its individuation, but only its causal history. The challenge is to single out its causal history in a non-circular way.

To sum up: It is not true, as Moore claims, that “on the counterfactual theory, for c to be the cause of e c must be necessary for e ” (CR, 354), at least not if “being necessary” means “being a necessary condition”. Asserting a counterfactual dependence between singular events does not amount to specifying a necessary and/or sufficient condition or any combination thereof. The counterfactual conditional $\neg c \square \rightarrow \neg e$ does not say or imply that in all cases where an e -like event occurs, a c -like event must have occurred, nor does it say or imply that in all cases where a c -like event occurs, an e -like event follows, nor does it say that c -like events are necessary elements of a sufficient set (NESS) for e -like events to occur. It does not state any general dependence between c -like events and e -like events whatever.

The widespread view that the antecedent of a counterfactual expresses a *conditio sine qua non* is untenable. The counterfactual (i) “Had the match not been struck, it would not have lit” states a necessary condition only if it is tacitly understood as implying the universally quantified indicative conditional (ii) “Every time a match lights, it has been struck beforehand”. But, singular counterfactuals do not carry such implications. (i) can be true without (ii) being true.

8 Causal Necessity without Necessary Conditions?

Now what about the modal intuition that the effect could not have failed to occur, given the cause? Causal connections are supposed to have some kind of modal force. Can a singularist counterfactual analysis account for the intuition that a cause *produces* the effect, rather than merely being followed by it, or that it makes the effect *inevitable*? Even if accounts in terms of necessary and/or sufficient *conditions* are untenable, the intuition will not go away that some kind of necessity is involved, perhaps a distinctive causal necessity.

The problem is where to get this necessity from, in the absence of strict causal laws. Armstrong’s necessitarian account of the laws of nature provides an instructive example. In Armstrong’s account, some physical properties F and G are

such that, as a contingent fact, the property of being F *necessitates* the property of being G . As Lewis and many others have observed, the necessitation relation N remains mysterious if not backed by something else:

Whatever N may be, I cannot see how it could be absolutely impossible to have $N(F,G)$ and Fa without Ga . (Unless N just is constant conjunction, or constant conjunction plus something else, in which case Armstrong's theory turns into a form of the regularity theory he rejects.) The mystery is somewhat hidden by Armstrong's terminology. He uses 'necessitates' as a name for the lawmaking universal N ; and who would be surprised to hear that if F 'necessitates' G and a has F , then a must have G ? But I say that N deserves the name of 'necessitation' only if, somehow, it really can enter into the requisite necessary connections. It can't enter into them just by bearing a name, any more than one can have mighty biceps just by being called 'Armstrong'. (Lewis 1983, 366)

Causal necessity does not come cheap, hence the view that "on the counterfactual theory, for c to be the cause of e c must be necessary for e " (CR, 354), faces a dilemma. Either the counterfactual theory succeeds in specifying causal laws that yield necessary conditions. In that case it collapses into a generalist theory. Or, in the absence of necessary conditions, some non-standard relation of causal necessitation is needed that no one has spelled out yet.

In a loose, preanalytical way of speaking, causal counterfactuals themselves express the sought-after modal force. The locution that a cause is necessary for its effects may be seen as an innocuous stylistic variant of the claim that e would not have occurred but for c . No harm is done unless this talk of necessity, which awaits analysis, is confused with the technical, logical notion of a necessary condition. Note again that without this confusion, the necessity claim cannot serve as a reason to deny the counterfactual in cases of overdetermination. In the circumstances, the second fire being too slow, the first fire caused the destruction. But as soon as it is argued that the first fire cannot have been necessary, because in its absence the other fire would have done the job, the logical notion of "necessary condition" is being employed that has no application to singular causal relations.

It has been suggested that the modal intuition about causality is fully captured by the counterfactual statement itself: "The 'causal glue' is the *truth* of the counterfactual statement, one could perhaps say" (von Wright 1976, 384). This move is unsatisfying, in leaving open what *makes* the counterfactual true. Many philosophers argue that counterfactual dependencies must be *grounded* in something factual, since they are not ground-floor properties of the physical world.

Well, they are grounded in facts. The *truth-conditions* for singular causal statements may be given by the counterfactual analysis, but the *truth-maker*

for the counterfactual conditional must include something factual. Truth-makers, if there are such things, are non-linguistic entities.³¹ Usually, they are conceived as facts, i. e., as states of affairs that obtain. The truth-makers for singular causal statements and for the causal counterfactuals that analyze them are the same, of course, if the analysis is correct. Here are some tentative and unoriginal suggestions: It is pivotal that the substances involved in the causing event have certain physical properties. As soon as we go beyond the metaphysics of causation and ask for a causal *explanation*, we will have to mention these properties. Claiming that the litmus tincture would not have turned red, had the acid not been poured into the test-tube, will fail to enlighten anyone who has witnessed the experiment. In science, we rarely ask for the triggering cause and leave it at that. Often it is obvious enough which event caused the effect. What we want to know, in addition, is why the effect produced was *F*, rather than *G*. For such a contrastive explanation, we must refer to the physical properties of the substances involved, to further facts about their arrangement and the physical forces present, and about contingent particular circumstances. These facts are what ground both the singular causal statement and its counterfactual analysis. To be sure, referring to those facts does not replace the counterfactual analysis, but supports it.

Plausibly, these facts are the source of the assumed necessity of causation. If a singular cause in some sense necessitates its effect, then this necessity does not derive from a strict causal law but from a number of general and particular facts which, taken together, account for the truth of the causal statement and its counterfactual analysis.

9 How to Explain the Petering Out of Causation

The last of Moore's six objections to the counterfactual theory was that "counterfactual dependence across chains of causes does not, as causation seems to, weaken or peter out" (CR, 399). The objection has force only if causation does peter out with time. In the philosophical debate, the majority view is that it

31 It is a matter of controversy whether the idea of truth-making is theoretically fruitful. Moore thinks so, Quine, Strawson and Davidson think otherwise. See Stepanians (2013) for an argument that Moore's "states of affairs" cannot do both jobs that Moore imposes on them, i. e., function as truth-makers for singular causal statements and as causal relata.

does not. Most theorists of causality, including Lewis, hold that the causal relation is transitive, and transitivity is incompatible with petering out.³²

On this matter, I side with Moore. Both in real life and in legal contexts we often judge that causal chains weaken over time, and with good reason. An adequate theory of causation should account for these judgments. This challenge is not easily met. It is much easier to explain transitivity. Most theories of causation – regularity theories, nomological theories and transfer theories, in particular – explain transitivity naturally, at the price of adopting a concept of causation that is neither part of our common sense picture nor appropriate to determine legal liability.

In the remainder of this paper, I will expound and defend the view that a singularist counterfactual theory can account for the nontransitivity both of counterfactual dependence and of causality. And only if the causal relation is not transitive, can causation weaken or peter out over time, as it plausibly does. My concern is limited, though. I shall not argue for the claim *that* the causal relation is not transitive – arguably, the burden of proof is on those who adhere to transitivity – but only that *if* it is not transitive, a singularist counterfactual theory has a nice explanation for this.³³

I once encouraged a student to spend an academic year in Norway. He heeded my advice. In Norway, he died from a heart attack during a swim in an icy mountain lake. When I was informed about the fatal accident, I felt sad, but not guilty. Now assume that the following counterfactuals are true:

- (A) If I had not encouraged him, he would not have gone to Norway.
- (B) If he had not gone to Norway, he would not have taken the swim in the icy mountain lake.
- (C) If he had not taken the swim, he would not have died.

Further assume that all three conditionals relate distinct events and have causal interpretations.³⁴ If we add transitivity, we get the result that my advice caused the student's death. But in real life, we seldom attribute causal or moral respon-

32 But remember that Lewis distinguishes between *causal dependence*, which is not transitive, and *causation*, which is transitive by stipulation: "If *c* and *e* are two actual events such that *e* would not have occurred without *c*, then *c* is the cause of *e*. But I reject the converse. Causation must always be transitive; causal dependence may not be; so there can be causation without causal dependence. Let *c*, *d*, and *e* be three actual events such that *d* would not have occurred without *c* and *e* would not have occurred without *d*. Then *c* is a cause of *e* even if *e* would still have occurred (otherwise caused) without *c*." (Lewis 1973b, 167)

33 Recently, the broad consensus that causation is transitive has started to crumble; see Hitchcock (2001) and Maslen (2004), who both refer to the counterfactual analysis.

34 One might well deny that (B) has a causal interpretation.

sibility over long and winding causal chains, even if the agent's contribution passes the *but for* test. As Moore puts it: "Caesar's crossing the Rubicon may well be a necessary condition for my writing this chapter, but so many other events have also contributed that Caesar's causal responsibility has long petered out" (CR, 102). The issue is all-important for theories of legal liability. The literature is vast, but manageable causal criteria that help to avoid implausible attributions of legal liability are hard to find. (There may be *normative* criteria, of course.) The standard theories of causation are deplorably unfit to yield such criteria. The *sine qua non* test fares particularly badly in this respect. Moore argues that "the sheer size of Caesar's contributions disqualifies Caesar's act as a cause" (CR, 398).³⁵ I agree, but Moore's quantitative consideration echoes rather than explains the intuition.

In order to avoid holding agents with causally remote contributions responsible, German legal theory has devised the doctrine of "adequate causation", according to which not just any *conditio sine qua non* of a given harm establishes legal liability, but only those factors that caused the harm with some degree of foreseeability, hence "adequately".³⁶ This theory, however, has never been convincingly elaborated and lacks theoretical rigour. In the philosophical debate, the situation is not much better. Terminologically, causes have been distinguished from "boundary conditions", "proximal" or "direct" causes have been distinguished from "remote" or "distal" causes, "complete" from "partial" or "contributory" causes, but as soon as causal responsibility is to be judged in complicated cases, the matter is usually left to pragmatic considerations.³⁷

An easy way out is to dis sever moral and legal responsibility from causal responsibility. I follow Moore in dismissing this easy way out. The petering out of causal chains needs to be explained somehow, even if having caused harm does not suffice for being morally responsible or legally liable.

In order to demonstrate how chains of counterfactual dependence between events can break, although the events are pairwise linked, one further element needs to be added to the counterfactual analysis, or rather to be made explicit. Suppose that I throw a heavy stone against a window, the window breaks, and I maintain: "If I had not thrown the stone, the window would not have broken". An observer objects: "That's not true. I could have thrown the stone as well. Or a sudden earthquake could have shattered the window." The proper reply, to my

³⁵ "My own suggestion is that [...] when there are too many event-links in the causal chain, it becomes too attenuated to support judgments of transitivity." (CR, 122)

³⁶ Compare Moore's extensive discussion of the related "harm-within-the-risk" test in the Anglo-American legal tradition (CR, 178–225).

³⁷ Lewis is no exception (see 1973b, 162).

mind, would have to be: “You might have, but you didn’t. And no earthquake was imminent.” What is sought is an analysis that supports this reply.

The standard view is that only a possible world semantics can yield such an analysis. One must find out whether the closest world that we have to consider contains earthquakes or not, and which standards of similarity between worlds to apply. In my singularist view, the case is much simpler. Using the possible worlds jargon, in the nearest possible world in which I do not throw the stone there is no earthquake, but simply an unbroken window. The reason is that in evaluating the counterfactual, we do not consider a certain *kind* of situation, and then figure out which kinds of changes in such situations constitute necessary conditions for which kinds of further changes. Rather we refer to the situation that actually obtained. This is only possible *ex post factum*. In an indeterministic universe, it was not written in the stars that no earthquake was nigh, and if it was written, I did not know this when throwing the stone. But since the truth-value of the counterfactual is being assessed in retrospect, my prior epistemic situation is irrelevant. This holds quite generally: An incalculable number of interferences *might* have occurred, but when we evaluate a causal counterfactual, both events being present, we already know whether any of these interferences did occur.

The *ex post factum* character is not only built into the grammar of truly counterfactual conditionals (in contrast to hypothetical ones), as mentioned above: A statement can only be contrary to fact if those facts obtain. It follows already from the logical form of the singular causal statements to be analyzed. It is no accident that all the textbook examples of singular causal statements are in the past or perfect tense. Singular causal statements of the form “*c* (has) caused *e*” assert the holding of a relation, whereas they *presuppose* the existence of the relata. Compare the relational statement “Adam is Bill’s brother”, which presupposes that two persons, Adam and Bill, exist, while it asserts that both are related as brothers. Taking the logical form of singular causal statements seriously, it is definitely unwise to burden the business of giving truth conditions for such statements with the problem of justifying induction. It is unwise to ask two questions at once, as Hume did. If it remains to be seen whether *e* will occur in the first place, the question as to whether *c* and *e* will be related as cause and effect is premature. The situation in which singular causal statements are being evaluated is different: Both events have occurred, and the open question is whether they were related as cause and effect. Whether *e* was caused by *c* is being examined *ex post factum*, in view of events that actually occurred. The counterfactual analysis only brings the nature of this task into the open. (For this reason, it is the congenial theory for lawsuits, which are backward-looking, too.)

In defending the counterfactual “If I had not thrown the stone, the window would not have broken” against the objection from alternative causation or overdetermination, we refer to the situation that actually obtained. We compare the happenings to a counterfactual situation in which the stone’s throw is missing while preserving as much as possible from the actual situation. In order to make this strategy explicit, the counterfactual analysis must be supplemented with a *ceteris paribus* clause that was implicit in it all along (recall Goodman’s observation). The clause is neither open-ended (like “under normal conditions”) nor vacuous (like “provided nothing interferes”), but has a determinate content. The cp clause fixes the prehistory of the causing event and the circumstances that actually pertained when it occurred, so that the hedged counterfactual reads: “If *c* had not occurred, and if everything else had been as it was when *c* actually occurred, *e* would not have occurred”.

This move is not original, of course. Lewis himself suggests it, as one among many other suggestions he makes to maximize similarity: “To get rid of an actual event *e* with the least overall departure from reality, it will normally be best not to diverge at all from the actual course of events until just before the time of *e*” (Lewis 1973b, 171). This is the right line to take in causal contexts, but Lewis cannot explain why. He considers a number of competing respects of similarity between possible worlds, which have to be traded off against one another. In particular, he suggests a trade-off between “similarities in matters of particular fact” and “similarities of law” (163). This trading off of similarity respects, however, and hence the fixing of truth conditions for counterfactuals, he says, “are a highly volatile matter, varying with every shift of context and interest” (1973a, 92).³⁸ Far from developing a similarity standard that deserves the name of a “metric” (Moore), Lewis employs only a loosely knit network of similarity respects. No general, context-independent similarity standard being available, Lewis calls the relation of comparative overall similarity among possible worlds “primitive”.

At least in causal contexts, we can do better to motivate the demand to leave the past unaltered, I think. The cause is the factor that made the difference, and this factor is where the divergence from actuality must be located. In evaluating causal counterfactuals, we do not brood over standards of comparative overall similarity, rather we simply refer to the facts that actually obtained up to the moment described in the antecedent. When trying to identify singular causes, we ask what would have happened *in that very situation*, if only the putative

38 “Overall similarity consists of innumerable similarities and differences in innumerable respects of comparison, balanced against each other according to the relative importances we attach to those respects of comparison.” (Lewis 1973a, 91)

cause had been absent. That is, we do not make assumptions, but we refer to facts that obtained. What does the trick, linguistically, is the *indexical* reference of the *ceteris paribus* clause. The clause “and if everything else had been as it was when *c* actually occurred” refers directly to the actual situation, and not via some general description of it.

Notably, the indexical nature of the cp clause precludes its being used to restore the truth of causal *laws* in the face of counter-instances. Genuine laws must not contain singular terms referring to particular objects, locations or times. Hence, if cp clauses have an indexical content, there can be no such things as cp laws. Indexical cp clauses need singular propositions as their habitat, so that they cannot be used to defend the cause-law thesis in the face of counter-instances. But they can do something better for the theory of causality, as I just pointed out. They are perfectly intelligible if combined with singular counterfactuals.³⁹

Adding this indexical cp clause to the counterfactual analysis, an easy explanation for the nontransitivity of the causal relation is available. (Lewis is well aware that counterfactual dependence is not transitive, but he does not use this insight to abandon the transitivity of causation.) Three events *c*, *d* and *e*. Assume that the counterfactuals $\neg c \square \rightarrow \neg d$ and $\neg d \square \rightarrow \neg e$ are true. The cp clause in the wording giving above being added, it becomes clear that the clause does not fix the same circumstances in both cases. In the first conditional, it fixes the history of event *c*, in the second conditional, it fixes the history of event *d*. So it is not the case that *c* and *d*, and *d* and *e*, respectively, are connected by the same dependence relation, as they would have to be, by definition, if the relation is supposed to be transitive. Both pairs of events are related by counterfactual dependence, but not by the same *particular* dependence relation. On closer examination, counterfactual dependence between events is not an ordinary relation like “bigger than” or “older than”. It can be instantiated only once, because it is a relational property of a particular pair of events, which cannot be detached from that pair. In other words, it is not a universal, as ordinary relational properties are. Asserting a counterfactual dependence between events is making a direct reference to a particular situation, namely to the actual history and circumstances of an actual event. This reference to actual circumstances provides a nice explanation of *why* counterfactual dependence is not transitive.

Reasoning through a chain of counterfactual conditionals is to commit the so-called fallacy of the ambiguous middle term, or four-term fallacy (*quaternio*

³⁹ See Keil (2005) for a more thorough discussion and a defence of this clause against competing translations of the Latin phrase “*ceteris paribus*”.

terminorum). The presence of this fallacy has been noticed before,⁴⁰ but neither Mackie nor Lewis draw the conclusion that *causation* is not transitive, nor do they recognize the value of an indexical *cp* clause.

As I said before, none of this amounts to compelling evidence *that* causation is not transitive. But given Lewis' theory, it strikes me that the burden to prove the contrary is on him. Instead of shouldering the burden, he simply *defines* that causation is transitive.⁴¹ That is a pity, given that the counterfactual analysis nicely explains non-transitivity. In the cases at hand, there *is* an unbroken chain of counterfactual dependencies, just as Lewis requires for causal chains: *c* and *d*, *d* and *e*, *e* and *f*, respectively, are pairwise linked. But since each antecedent in the series refers to different circumstances, the unbroken chain of pairwise counterfactual dependencies does not amount to transitive closure. Lewis does not claim that it does, but he cannot explain how such a chain of *non-transitive* counterfactual dependencies should generate, without further assumptions, the transitivity of causation.⁴²

10 Conclusion

The title of my paper, "Making causal counterfactuals more singular", is a misnomer. Counterfactual dependence between particular events *is* a singular relation. True causal counterfactuals do not require the existence of strict causal laws, i.e., laws of succession that subsume cause-effect pairs. All the counterfactual analysis can do is to bring the singularist character of causal relations into the open. In this contribution, I defended an amended version of the counterfactual theory of event causation both against Moore's criticism and against some traits of Lewis' version of the theory. The dialectical situation was somewhat complicated. Moore holds that all generalist theories fail, but wavers as to

40 "The non-transitivity of counterfactuals is nothing surprising: it is merely a variant of a very old form of fallacy, a syllogism with an ambiguous middle term." (Mackie 1980, 53) Similarly, Lewis explains the nontransitivity of counterfactual dependence by the fact that often "the closest worlds where the first antecedent holds are different from [...] the closest worlds where the second antecedent holds" (Lewis 1973a, 33). See also Moore (CR, 393).

41 "We extend causal dependence to a transitive relation in the usual way. Let *c*, *d*, *e*, . . . be a finite sequence of actual particular events such that *d* depends causally on *c*, *e* on *d*, and so on throughout. Then this sequence is a *causal chain*. Finally, one event is the *cause* of another iff there exists a causal chain leading from the first to the second." (Lewis 1973b, 167).

42 "But if all there is to direct causation is counterfactual dependence, and if all there is to transitive causation is the iteration of direct causation, there just is no room to add anything more to counterfactual dependence in order to generate transitivity." (Rosenberg 1992, 308)

whether the counterfactual theory is generalist or singularist or *sui generis*. I take him to think that the theory is ultimately generalist, since he claims that the counterfactual theory founders on the same arguments on which all generalist theories founder. It is not surprising that these arguments cut no ice against a singularist version of the counterfactual theory.

On a number of counts, I simply defended Lewis' theory against Moore's misdirected criticism. Moore's claim that according to the counterfactual theory, causation simply *is* counterfactual dependence (CR, 394) does not even do justice to Lewis' *analysandum*. Moore's unreasonable demand that all the commonplaces about causation have to *follow from* the counterfactual analysis rests on the identity thesis. His claim that on the counterfactual theory, the cause "must be necessary for the effect", suffers from an ambiguity: The logical notions of necessary and sufficient *conditions* are not defined for dependency relations between particular events. A distinctive kind of *causal* necessity might be involved, but is notoriously hard to pinpoint and does not seem to figure in the counterfactual analysis.

In other respects, I had to part company with Lewis in order to highlight some underrated strengths of the counterfactual approach. The common denominator of my revisions is that they make the counterfactual theory unambiguously singularist. The suggested revisions comprise: performing the counterfactual analysis with Davidsonian rather than with Lewisian events, taking seriously the ex post-character of singular causal statements, making explicit the indexical *ceteris paribus* clause that fixes the actual circumstances of the causing event, abandoning the alleged transitivity of causation, and, last but not least, doing everything in the correct order. The correct order is to start with bedrock intuitions about the nature of causality rather than with a general possible world semantics, and then to turn these intuitions into constraints for the class of counterfactuals that a counterfactual analysis of causation has to consider.

References

- Armstrong, David M. (1983). *What is a Law of Nature?*, Cambridge: Cambridge University Press.
- Carroll, John (1994). *Laws of Nature*, Cambridge: Cambridge University Press.
- Cartwright, Nancy (1983). *How the Laws of Physics Lie*, Oxford/New York: Oxford University Press.
- Collins, John D., Edward J. Hall and L. A. Paul (Eds.) (2004). *Causation and Counterfactuals*, Cambridge, MA: MIT Press.
- Cummins, Robert (1983). *The Nature of Psychological Explanation*, Cambridge, MA: MIT Press.

- Danto, Arthur C. (1973). *Analytical Philosophy of Action*, Cambridge: Cambridge University Press.
- Davidson, Donald (1963). Actions, Reasons and Causes. In his (1980), 3–19.
- Davidson, Donald (1967). Causal Relations. In his (1980), 149–162.
- Davidson, Donald (1970). Mental Events. In his (1980), 207–225.
- Davidson, Donald (1980). *Essays on Actions and Events*, Oxford: Clarendon Press.
- Davidson, Donald (1990). Representation and Interpretation. In K. A. Mohyeldin Said et al. (Eds.), *Modelling the Mind*, Oxford: Oxford University Press, 13–26.
- Davidson, Donald (1993). Replies. In: R. Stoecker (Ed.), *Reflecting Davidson*, Berlin/New York: de Gruyter.
- Davidson, Donald (1995). Laws and Cause. *Dialectica* 49, 263–279.
- Goodman, Nelson (1954). The Problem of Counterfactual Conditionals. Repr. in his *Fact, Fiction, and Forecast*, Cambridge, MA: Harvard University Press 1983.
- Hart, Herbert L. A. and Honoré, Anthony M. (1959). *Causation in the Law*, Oxford: Oxford University Press.
- Hitchcock, Christopher (2001). The Intransitivity of Causation Revealed in Equations and Graphs. *Journal of Philosophy* 98, 273–299.
- Hume, David (1748). *An Enquiry Concerning Human Understanding*, ed. by T. H. Green and T. H. Grose, Repr. Aalen: Scientia 1992.
- Keil, Geert (2000). *Handeln und Verursachen*, Frankfurt am Main: Klostermann.
- Keil, Geert (2005). How the Ceteris Paribus Laws of Physics Lie. In Jan Faye, Paul Needham, Uwe Scheffler and Max Urchs (Eds.), *Nature's Principle*, Berlin/Heidelberg/New York: Springer, 167–200.
- Keil, Geert (2009). Making Something Happen: Where Causation and Agency Meet. In Siri Granum Carson and Kjartan Koch Mikalsen (Eds.), *Nature and Rational Agency*, Frankfurt am Main: P. Lang, 9–28.
- Lewis, David (1973a). *Counterfactuals* (2nd ed.), Oxford: Oxford University Press 1986.
- Lewis, David (1973b). Causation. In his (1986), 159–213.
- Lewis, David (1979). Counterfactual Dependence and Time's Arrow. In his (1986), 32–66.
- Lewis, David (1981). Events. In his (1986), 241–269.
- Lewis, David (1983). New Work for a Theory of Universals. *Australasian Journal of Philosophy* 61, 343–377.
- Lewis, David (1986). *Philosophical Papers*, Vol. II, New York/Oxford: Oxford University Press.
- Lewis, David (2000). Causation as Influence. Repr. in Collins, Hall and Paul (Eds., 2004), 75–106.
- Mackie, John L. (1965). Causes and Conditions. *American Philosophical Quarterly* 2, 245–264.
- Mackie, John L. (1974). *The Cement of the Universe. A Study of Causation*, (2nd ed.) Oxford: Oxford University Press 1980.
- Mackie, John L. (1980). The Transitivity of Counterfactuals and Causation. *Analysis* 40, 53–54.
- Maslen, Cei (2004). Causes, Contrasts, and the Nontransitivity of Causation. In Collins, Hall, and Paul (Eds., 2004), 341–358.
- Menzies, Peter (1996). Probabilistic Causation and the Pre-Emption Problem. *Mind* 105, 85–117.
- Mill, John Stuart (1843). *A System of Logic, Ratiocinative and Inductive*, [= *Collected Works*, Vol. VII/VIII], ed. by J. M. Robson, Toronto: University of Toronto Press 1973/4.

- Moore, Michael S. (CR). *Causation and Responsibility. An Essay in Law, Morals, and Metaphysics*, Oxford: Oxford University Press 2009.
- Pollock, Frederick (1901). *Torts*, New York: Macmillan (6th ed.).
- Quine, W. v. O. (1950). *Methods of Logic*, New York: Holt.
- Quine, W. v. O. (1960). *Word and Object*, Cambridge, MA: MIT Press.
- Rosenberg, Alexander (1992). Causation, Probability and the Monarchy. *American Philosophical Quarterly* 29, 305–318.
- Russell, Bertrand (1905). On Denoting. *Mind* 14, 479–493.
- Scriven, Michael (1961). The Key Property of Physical Laws – Inaccuracy. In Herbert Feigl and Grover Maxwell (Eds.), *Current Issues in the Philosophy of Science*, New York: Holt, 91–101.
- Stepanians, Markus (2013). Moore on “the true metaphysics of causation”. In this volume.
- Strawson, Peter F. (1959). *Individuals. An Essay in Descriptive Metaphysics*, London: Routledge.
- von Wright, Georg Henrik (1976). Replies. In Juha Manninen and Raimo Tuomela (Eds.), *Essays on Explanation and Understanding. Studies in the Foundations of Humanities and Social Sciences*, Dordrecht: D. Reidel, 371–413.
- Woodward, James (2000). Explanation and Invariance in the Special Sciences. *British Journal for the Philosophy of Science* 51, 197–254.
- Wright, Richard W. (2013). The NESS Account of Natural Causation: A Response to Criticisms. In this volume.

