3

What Russell got Right

CHRISTOPHER HITCHCOCK

3.1 Introduction

FN:1 Bertrand Russell's 'On the Notion of Cause' is now[1] ninety years old, yet its central claim still provokes. To summarize his argument, I can do no better than to provide an excerpt from his oft-quoted introduction:

> … the word 'cause' is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable …
>
> All philosophers, of every school, imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced sciences such as gravitational astronomy, the word 'cause' never appears. Dr James Ward … makes this a ground of complaint against physics … To me, it seems that … the reason why physics has ceased to look for causes is that, in fact, there are no such things. The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm. (Russell 1913, p. 1).[2] FN:2

The most basic principles of 'advanced sciences' are not couched in the form '*A* causes *B*', but rather in the form of differential equations. Most philosophers have rejected Russell's main conclusion, mostly for good reasons. I wish to argue, however, that Russell's arguments carry important insights that can survive the rejection of his primary conclusion. While

For helpful comments I would like to thank audience members at the University of Queensland, the University of Helsinki, and the Causal Republicanism conference in Sydney, especially Peter Menzies, Huw Price, and David Spurrett. I would also like to thank Alan Hájek, Jonathan Schaffer, and Jim Woodward for comments on later drafts of the paper

[1] I.e., at the time of this writing, 2003. By the time you are reading the published version of this paper, it will be several years older than that.

[2] Henceforth, all page references are to this source unless stated otherwise.

there is no need for the word 'cause' to be completely extruded from the philosophical vocabulary, it should be used with much greater care than is usually exercised by philosophers. In developing my case, I will consider two of the most interesting critiques of Russell's argument, due to Nancy Cartwright (1983) and Patrick Suppes (1970). In the final section, I will pull the various threads together to paint a general picture of the nature of causation, and draw some conclusions for the use of causal concepts in philosophical projects.

3.2 Russell's Targets

Even in the short excerpt above, it is possible to identify no less than four targets of Russell's attack, in order (starting with the title) they are: (1) the notion of cause; (2) the word 'cause'; (3) the existence of causes; and (4) the 'law of causality'. Correspondingly, Russell seems to be make the following claims:

R1 The notion of cause is incoherent, or fundamentally confused.

R2 The word 'cause' has 'misleading associations', and should be eliminated from philosophical usage. (Presumably this is supposed to apply to cognates such as 'causation' as well.)

R3 There are no causes.

R4 The 'law of causality' is obsolete and misleading.

Russell does not seem to distinguish clearly between these claims, although they are far from being equivalent. R1 does seem to entail R2 through R4, but R2 does not clearly imply *any* of the others. R3 does not imply R1 or R2: there are no 'unicorns', but the notion of a unicorn is not incoherent (although see Hull 1992), nor should 'unicorn' be expunged from the philosophical vocabulary. (Indeed, precisely because there are no unicorns, the word 'unicorn' is useful in formulating philosophical examples like this one.) And R4 appears to be compatible with the denial of any of R1–R3: the notion of cause might be coherent and useful, the word 'cause' might be be useful, and causes might exist, even if it's wrong or misleading to say that *every* event is preceded by a cause from which it invariably follows.

Russell's third claim does not seem to reappear beyond the second paragraph, although as we have noted it would seem to follow as a natural consequence of R1. For this reason I will not consider R3 any further. In what follows, I will grant that R4 is basically correct, and argue that R1 is fundamentally mistaken. The most interesting claim is R2, which, I will argue, survives in a much modified form.

3.3 The Law of Causality

Following the time-honored technique used by impatient readers of mystery novels, let us turn to the last page of Russell's article, where Russell summarizes his major conclusions. He writes:

> We may now sum up our discussion of causality. We found first that the law of causality, as usually stated by philosophers, is false, and is not employed in science. We then considered the nature of scientific laws, and found that, instead of stating that one event A is always followed by another event B, they stated functional relations between certain events at certain times … We were unable to find any *a priori* category involved: the existence of scientific laws appeared as a purely empirical fact … Finally, we considered the problem of free will … The problem of free will *versus* determinism is … mainly illusory … (26)

I will return briefly to Russell's discussion of free will below. It is clear that at the end of the day, Russell took his critique of the law of causality to be fundamental. What is that law? Russell cites two different formulations (6):

> The Law of Causation, the recognition of which is the main pillar of inductive science, is but the familiar truth, that invariability of succession is found by observation to obtain between every fact in nature and some other fact which has preceded it.[3] (Mill 1843, book 3, ch. v, §2.)

FN:3

> … this law means that every phenomenon is determined by its conditions, or, in other words, that the same causes always produce the same effect. (Bergson 1910, p. 199.)[4]

FN:4

Russell's central argument is succinct and powerful:

[3] Given the sort of regularity view of causation endorsed by Mill, this statement implies that every event has a cause.

[4] It is clear from context that Bergson does not endorse this principle, but is only presenting what he takes to be the normal understanding of it. My colleague Alan Hájek points out, that Bergson's 'other words', repeated by Russell, do not express the same principle as that stated in the first clause. The principle 'same cause, same effect' would be trivially true if there were no causes; not so the principle that every phenomenon is determined by its conditions (at least if we count 'determining conditions' as causes).

> The principle 'same cause, same effect' ... is ... utterly otiose. As soon as the antecedents have been given sufficiently fully to enable the consequent to be calculated with some exactitude, the antecedents have become so complicated that it is very unlikely they will ever recur. (8–9)

By the time you pack enough material into A to guarantee that B will follow, A is so complex that we can't reasonably expect it to occur more than once. (Indeed, it was Mill himself who taught us just how detailed any such A would have to be.) But *any* A that occurs only once will invariably be succeeded by any B that just happens to occur on that occasion. Although rough generalizations may play a role in the early stages in the development of a science, science is not constructed upon invariable regularities. Note that Russell's critique also serves as an effective critique of regularity accounts of causation, such as that put forward by Mill. This may go some way toward explaining why Russell did not clearly distinguish between claim R4 and the others.

At various points Russell takes pains to attack not merely the truth or utility of the lw of causality, but also its *a priority*, as well as the *a priority* of any suitable replacement. This clearly shows that one of Russell's targets is the Kantian view that something like the law of causality is a necessary condition for us to gain knowledge of the empirical world. (See e.g. Kant 1781, Second Analogy, A 189–203, B 234–49; 1787, §§14–38.)

Note finally that Russell is not rejecting determinism *per se*. In fact he clearly embraces Laplacean determinism (pp. 13–14). So if anything, the situation is even worse for the law of causality than Russell claims, since there is strong evidence for indeterminism among quantum phenomena, and it is now known that even Russell's beloved classical gravitational astronomy does not entail determinism (Earman 1986, Norton this volume). The rejection of the law of causality is definitely something that Russell got right.

3.4 Misleading Associations

Claim R2 has two clauses, the first an observation about philosophy, the second a methodological recommendation. The first clause is phrased somewhat awkwardly in Russell's introduction: '... the word "cause" is ... inextricably bound up with misleading associations ...' It might seem natural to paraphrases this by saying 'the word "cause" has caused a great deal of confusion', although there are obvious reasons why this phrasing is undesirable. Russell's discussion of free will seems aimed in part to establish this first clause: it illustrates one area in which causal talk has led to confusion. Ironically, Russell's own paper also serves to illustrate the point. Russell's failure to distinguish between claims R1, R3, and R4 is no doubt due, in part, to ambiguities in words like 'cause' and 'causality'. The former can be a noun or a verb, the latter can describe a relation between events, or a principle governing such relationships. Indeed, David Lewis (1986a, p. 175) makes a point of avoiding the word 'causality' (he calls it a 'naughty word') precisely because it has such a double connotation.

The second clause of R2, that the extrusion of the word 'cause' (and its cognates) from the philosophical vocabulary is desirable, is clearly supposed to follow from the first clause; it also seems to follow from R1. We will discuss this methodological dictum in greater detail below.

3.5 On the Notion of Cause

R1 is not stated explicitly, although the title certainly suggests that the coherence of the notion of cause is to be scrutinized, and much of the argumentation on pp. 2 through 8 of Russell's paper is naturally construed as support for R1.

Russell challenged the coherence of the notion of cause in part by criticizing existing attempts to give a definition of this concept. In particular, he considers the following purported definition, one of three drawn from Baldwin's *Dictionary*:

> Cause and effect ... are correlative terms denoting any two distinguishable things, phases, or aspects of reality, which are so related that whenever the first ceases to exist the second comes into existence immediately after, and whenever the second comes into existence the first has ceased to exist immediately before. (Cited in Russell 1913, p. 2)

Russell then raises a Zenoesque paradox for this definition: If cause and effect are distinct and temporally contiguous, then they can't both be point-like; at least one of them must be temporally extended. Assume, for the sake of argument, that the cause has some finite duration, say from t_0 to t_1.

(There are tricky topological issues about whether the temporal endpoints are included within the duration of the cause. Since the cause and effect are contiguous, one, but not both, must be occurring at t_1. But let us put this issue aside.) Since the effect does not begin until t_1, the first half of the cause (with duration from t_0 to $t_{0.5}$), is not really efficacious: the pattern of invariable association can be disrupted by contingencies occurring in the interval between $t_{0.5}$ and $t_{1.}$. We can repeat this process *ad infinitum*, showing that at most one instantaneous time-slice of the cause can be efficacious, contradicting our assumption that the cause is extended. I will not examine this particular argument in detail, since the many difficulties with a simple regularity analysis of causation are widely known. Moreover, ninety years later we still lack a widely accepted analysis of causation.

In the twenty-first Century, however, we have become sufficiently cynical to reject arguments of the form:

> There is no adequate philosophical account of X, despite a rich history of attempts to provide one. Therefore, the concept X is incoherent or fundamentally confused.

This argument embraces an optimism about the power of philosophical analysis that is no longer warranted. With hindsight, then, we should reject Russell's argument for R1 on the basis of his critique of Baldwin's definition.

There is, however, an interesting general point that emerges from Russell's critique. The problems for Baldwin's definition resulted in part from its combination of two distinct elements: cause and effect are supposed to stand in a relation of temporal contiguity, and also of invariable association. Many of the problems that face more recent attempts arise for the same general reason. Our notion of cause seems to involve (at least) two different dimensions: cause and effect *covary* in some way; and they stand in certain kinds of spatiotemporal relations to one another. Regularity, probabilistic and counterfactual theories (such as Mackie 1974, chapter 3, Cartwright 1983, and Lewis 1986b, respectively) attempt to capture the covariance aspect of causation, while process theories (e.g. Salmon 1984) attempt to capture the broadly spatiotemporal aspects. There have been some interesting attempts to combine these (Menzies 1989, Dowe 2000, Schaffer 2001), but none have received widespread acceptance. Thus Russell raises the possibility that our notion of cause may be incoherent because it attempts to combine incompatible elements. But as we noted before, this does not follow simply from the fact that no attempt to combine these elements has met with perfect success, especially in light of the partial success that has been achieved, and the insights that have been gained such attempts.

So far, however, we have not really touched on Russell's central argument for R1 (and by extension, for the others as well). This argument involves as a central premise the following claim:

P. The word 'cause' does not appear in the advanced sciences.

This premise seems to play a dual role. First, there is something like an inference to the best explanation from P to R1: R1 if true, would explain why advanced sciences do not explicitly employ causal concepts. (This explanation would presumably appeal to the tacit premise that advanced sciences do not make use of notions that are incoherent.) Second, P seems to undermine a primary motivation for thinking that there *must be* a coherent notion of cause. Any attack on R1, then, will have to address the premise P, and its dual role in supporting R1. The critiques of Cartwright (1983) and Suppes (1970) do just that. Cartwright attempts to provide us with an alternative rationale for having a notion of causation, one that does not appeal to the supposed role of causal concepts in science. Suppes directly attacks premise P itself. More subtly, I think Suppes' critique serves to undermine the abductive inference from P to R1: to the extent that P is true, it can be explained without appeal to the incoherence of causal notions. I now turn to a more detailed discussion of these two critiques.

3.6 Cartwright's Criticism

Cartwright reports having received the following letter:

> It simply wouldn't be true to say,
> 'Nancy L. D. Cartwright … if you own a TIAA life insurance policy, you'll live longer.'
> But it is a fact, nonetheless, that persons insured by TIAA do enjoy longer lifetimes, on the average, than persons insured by commercial insurance companies that serve the general public. (Cartwright 1983, p. 22)

TIAA—the Teacher's Insurance and Annuity Association—offers life insurance for university professors and other individuals involved in education. Owning such a policy is *correlated* with longevity because only educators qualify to purchase such a policy. Educators tend to live longer than average for a number of reasons: they are comparatively well-compensated and receive good health benefits, they tend to be well-informed about health care issues, their careers are not especially dangerous, and so on. Nonetheless, purchasing a TIAA life insurance policy is not an *effective strategy* for achieving longevity.

By contrast, quitting smoking, reducing consumption of saturated fat, and exercising are (to the best of our current knowledge) effective strategies for achieving longevity. Cartwright then claims that

> ... causal laws cannot be done away with, for they are needed to ground the distinction between effective strategies and ineffective ones. (Cartwright 1983, p. 22).

It's not clear to me that *laws* are needed, insofar as laws are normally thought to be exceptionless, fully general, and so on. Perhaps we can get by with regularities, or patterns of functional dependence. But apart from that, I believe that Cartwright is fundamentally correct. There is a crucial distinction between the *causal* laws, regularities, or what have you, and the *non-causal* ones. The former can, while the latter cannot, be exploited for the purposes of achieving desired outcomes. Ultimately, this distinction is what causation is all about.

This provides us with one explanation for the absence of overtly causal notions in gravitational astronomy: there is no real prospect for manipulating the positions or masses of the heavenly bodies. By contrast, the fields in which there is the greatest interest in causation today tend to be ones with a practical orientation, such as econometrics, epidemiology, agronomy, and education. But I also think that there is a deeper explanation for the absence of causal talk in gravitational astronomy.

Cartwright's pragmatic orientation provides us with a helpful way of thinking about causation. As our paradigmatic causal system, consider a collection of billiard balls bouncing around on a low friction billiard table. This is not literally a closed system, but for stretches of time it remains largely unaffected by influences other than those intrinsic to the system itself. We could write down 'laws' describing the motions of the billiard balls during these periods of isolation, although these will only be *ceteris paribus* generalizations that break down under certain conditions. These generalizations can be used to predict the motion of the balls so long as the relevant ceteris paribus conditions continue to hold.

Now suppose that I intervene to change the motion of a particular ball. From the point of view of the 'laws' describing the isolated system, this intervention is 'miraculous', it is something that is not even countenanced within those generalizations. Now some of the original generalizations will be violated, while others remain valid (at least as applied to the motions of certain of the balls). The latter, but not the former, can be used to predict the effects of the intervention, and hence to support effective strategies. These are the genuinely causal generalizations. Looked at in a different way, the genuinely causal counterfactuals (what Lewis (1979) calls the 'non-backtracking counterfactuals') are the ones whose antecedents are (or rather, would be) made true by external interventions. This accords with Lewis' idea that the antecedents of non-backtracking counterfactuals are (typically) made true by 'miracles', except that in our case the antecedents are only miraculous relative to the 'laws' of the isolated system. From an external perspective, there is nothing miraculous about intervening on an otherwise isolated system. This picture is elaborated in much greater detail in Woodward and Hitchcock (2003), Hitchcock and Woodward (2003), and Woodward (2003). It also informs important recent work in causal modeling, such as Pearl (2000) and Spirtes, Glymour and Scheines (2000).

This picture runs into deep problems, however, when we try to expand the laws of the system so as to include the source of the intervention within the system itself. In particular, it's going to be extraordinarily difficult to find causation within a theory that purports to be universal—about *everything*. Russell's example of gravitational astronomy has just this character: the laws of classical gravitational astronomy were taken to apply to the universe as a whole (and presumably the 'final theory', if there is one, will also be universal in this sense). When the system under study is the universe as a whole, there is no external vantage point from which we can interfere with the system, and from which we can sensibly talk about miracles within the system. How to find or even understand causation from within the framework of a *universal* theory is one of the very deep problems of philosophy. In that regard, this problem is in good company. Nietzsche wrote that 'As the circle of science grows larger, it touches paradox at more places' (*The Birth of Tragedy*, cited in Putnam 1990).

This theme has been taken up in a paper by Hilary Putnam (1990). There are many places in science (broadly construed) where our theories work perfectly well so long as there is an external perspective which is not covered within the theory itself, but when the theory 'grows larger'—when it tries to encompass the external perspective as well—it becomes embroiled in paradox. Putnam discusses two central examples. The first is the measurement problem in quantum mechanics. Even in its classical (non-relativistic) form, that theory makes extraordinarily accurate predictions about the outcomes of various measurements that can be performed on a system. Once we include the measuring apparatus within the system, however, the normal rules of quantum mechanics (Schrödinger's equation and Born's rule) yield contradictory predictions for the behavior of the apparatus. The problem is sufficiently serious that otherwise sensible physicists have put forward outlandish metaphysical hypotheses (such as the existence of parallel universes, and a fundamental role for consciousness in physics) in hopes of resolving it.

Putnam's second example is the liar paradox. We can talk of sentences being true or false, and even construct powerful semantic systems, so long as the sentences in the target language do not contain the words 'true' and 'false'. These terms appear in the meta-language, the one used within the external perspective used to analyze the target language. But once we introduce the terms 'true' and 'false' into the target language, we are faced with the liar sentence: 'This sentence is false.' This sentence can be neither true nor false, on pain of contradiction, and our semantic theory 'touches paradox'.

The problem of free will and determinism also has something of this character. We are perfectly happy to apply our (hypothetical) deterministic theories to determine the consequences of our decisions, once made. Indeed, as compatibilists such as Hume (1748, §8) and Hobart (1934) have pointed out, freedom would be *undermined* if there were no systematic connection between our choices and their physical outcomes. Problems arise, however, when we try to bring the decision-making process itself within the scope of the deterministic theory. Some libertarians have postulated that *agents* can be causes that nonetheless lie outside of the causal order, thus guaranteeing the existence of an external perspective that cannot be brought under the scope of a deterministic theory, a constraint upon the diameter of the circle of science. This is yet another example of the extremes to which philosophers have been driven by the desire to avoid paradox.

3.7 Suppes' Criticism

The second main line of response that I want to pursue comes from Patrick Suppes' book *A Probabilistic Theory of Causality* (1970). In the introduction, Suppes launches an attack on Russell's key premise P. After citing the passage from Russell with which we began, Suppes writes:

> Contrary to the days when Russell wrote this essay, the words 'causality' and 'cause' are commonly and widely used by physicists in their most recent work. There is scarcely an issue of *Physical Review* that does not contain at least one article using either 'cause' or 'causality' in its title. A typical sort of title is that of a recent volume edited by the distinguished physicist, E. P. Wigner, 'Dispersion relations and their connection with causality' (1964). Another good example is the recent article by E. C. Zeeman (1964) 'Causality implies the Lorentz group.' The first point I want to establish, then, is that discussions of causality are now very much a part contemporary physics. (Suppes 1970, pp. 5–6)

Since Suppes wrote this almost forty years ago, I conducted a quick and unsystematic internet search of the *Physical Review* journals (a series of 9) from 2000 to 2003 and found 76 articles with 'cause', 'causes', 'causality', or some similar term in the title. Here are the first three examples listed:

> 'Tree Networks with Causal Structure' (Bialas et al. 2003)
> 'Specific-Heat Anomaly Caused by Ferroelectric Nanoregions in Pb(Mg[sub 1/3]Nb[sub 2/3])O[sub 3] and Pb(Mg[sub 1/3]Ta[sub 2/3])O[sub 3] Relaxors' (Moriya et al. 2003)
> 'Observables' in causal set cosmology' (Brightwell et al. 2003)

So Suppes' observation remains true in 2003.

Taking the above passage in isolation, it may be natural to construe Suppes as claiming that physics has advanced since Russell's time, and that we have now learned that causation *is*, after all, an essential notion in philosophical theorizing. But Suppes' point is really quite different. Continuing from the previous quote, Suppes writes:

> The reasons for this are, I think, very close to the reasons why notions of causality continue to be an important ingredient of ordinary talk ... At the end of the passage quoted from Russell, there is an emphasis on replacing talk about causes by talk about functional relationships, or more exactly, by talk about differential equations. This remark is very much in the spirit of classical physics, when the physical phenomena in question were felt to be much better understood at a fundamental level than they are today. One has the feeling that in contemporary

physics the situation is very similar to that of ordinary experience, namely, it is not possible to apply simple fundamental laws to derive exact relationships such as those expressed in differential equations. What we are able to get a grip on is a variety of heterogeneous, partial relationships. In the rough and ready sense of ordinary experience, these partial relationships often express causal relations, and it is only natural to talk about causes in very much the way that we do in everyday conversation. (Suppes 1970, 6)

Suppes' point, then, is not so much that physics has advanced, but that it has, in a certain sense, *regressed*. Physics, like other branches of science, seeks not only to provide a systematic treatment of the phenomena presently within its scope, but also to uncover new phenomena. At any given time, then, there are bound to be many phenomena that are deemed to be within the scope of a field such as physics, yet still await systematic treatment within that field. Russell was writing at a highly unusual time in the history of physics, or at least of gravitational astronomy, when existing physical theory was thought to have successfully encompassed (almost) all of the phenomena within its scope. This period was short-lived: Einstein turned gravitational astronomy on its head less than three years later.[5] In the meantime, physics has uncovered a slew of phenomena that have not yet succumbed to comprehensive mathematical analysis. Often, it is natural to describe these phenomena in causal terms.

FN:5

Russell's premise P, then, is strictly false. Suppes' attack is, nonetheless consistent with a more modest formulation along the following lines:

P* There are advanced stages in the study of certain phenomena when it becomes appropriate to eliminate causal talk in favor of mathematical relationships (or other more precise characterizations).

We can explain the truth of P* without appealing to the incoherence of causal notions. To say that one thing *causes* another is to state something very important about the relationship, as we have learned from Cartwright. Nonetheless, it is to characterize the relationship only in a qualitative way. When it becomes possible to state the relationship in more precise terms, we do so. But it does not follow that the relationship has ceased to be causal. To the extent that Russell's premise P is true, it does not support his conclusion RI.

[5] There is no evidence that Russell was aware of developments that were already underway within (a very small part of) the German physics community. Indeed, by 1912–1913, general relativity was not even on the radar screen of the British physics community (Andrew Warwick, personal communication).

3.8 Conclusions

So let us take stock. Russell was right that science in general is not grounded in some 'law of causality'. Moreover, he was right that the word 'cause' does not appear in sciences 'like' classical gravitational astronomy. There are, however, very few fields like this; classical gravitational astronomy is unusual in a number of respects: it is remote from practical applications, it purports to be a universal theory, and its principles can be captured in simple mathematical formulas. But it does not follow from any of this that our notion of causation is incoherent. Rather, following Cartwright (1983), we should recognize that the notion of causation is indispensable, not because it is fundamental to science, but because the distinction between causal relationships and non-causal relationships grounds the distinction between effective and ineffective strategies. This distinction is very real, and one that we are all aware of (at least implicitly).

Cartwright's argument gives rise to an interesting grammatical point. 'Cause' can be a verb, as when we say 'lightning storms cause fires'. 'Cause' can also be a noun, as in 'the lightning storm was the cause of the fire'. In addition, 'causation' is a noun, referring either to an abstract concept, or to a specific relation that holds between certain events (or facts, or properties, or what have you). 'Causality' is also a noun, often used as a synonym for 'causation', but also having a more archaic meaning akin to 'determinism', and associated with the 'law of causality' attacked by Russell. Cartwright's argument, however, focuses our attention on a less-discussed cognate: the adjective 'causal'. This adjective applies to relationships, such as laws, regularities, correlations, or patterns of functional dependence. It is this adjective that picks out the most central causal notion. The noun 'causation' is best thought of as referring not to a specific relation, but rather to a property that is possessed by some relationships and not by others.

This analysis meshes nicely with Suppes' critique. Central to Suppes' argument is the idea that claims of the form 'A causes B' can be *precisified*. For example, it may be possible to characterize functional relationship between A and B in mathematical terms. The different possible precisifications correspond to different causal relationships that might hold between A and B.

Following Chalmers (1996), who distinguishes between the 'easy' and 'hard' problems of consciousness, I will divide the problems of causation into the 'easy' ones and the 'hard' ones. The hard problems concern the origins of the distinction between the causal and the non-causal. These include finding the sources of causal asymmetry and causal modality (i.e., what makes causal relations non-accidental), and locating causation within a global scientific theory of the sort that physics aims to provide. We can of course expect to make some progress on these hard problems, just as we have made progress on the liar paradox and the measurement problem. Moreover, we may expect to learn a great deal about causation in the process of tackling these problems. But we should not expect these hard problems to go away any time soon.

The 'easy' problems, by contrast, concern distinctions among the different kinds of causal relationships. Like Chalmers' 'easy' problems of consciousness, these problems are actually pretty hard, but they should at least be tractable within existing philosophical methodology. Unfortunately, I think much of the energy here has been misdirected, by arguing over just which causal relation *is* causation *per se*. This is particularly prominent in the literature on counterfactual theories of causation, where there is a cottage industry of trying to define *causation* in terms of counterfactuals that are already assumed to be, in a broad sense, caus*al*.[6] Within this literature, authors have addressed such questions as: is *causation* transitive? Now I can't disparage this work in any general way, since I've engaged in precisely this debate myself. (The answer is 'no', by the way; see Hitchcock 2001a, 2001b). But there is a need to appreciate that the target of analysis here is not causation *per se*, but rather a specific kind of causal relation called 'token' causation (I find this terminology misleading, and prefer 'actual' causation, but I will not attempt to argue with entrenched usage). Many philosophers, especially metaphysicians, seem to think that actual causation just *is* causation. This belief is belied by the fact that there is an entire field devoted to causal representation and causal inference that barely mentions token causation. For example, token causation is never discussed in Spirtes, Glymour and Scheines (2000), and appears only in the final chapter of Pearl (2000). Presumably, these authors are nonetheless talking about *causal* inference and *causal* modeling. Token causation is in fact a rather narrowly circumscribed causal notion that is involved when we make retrospective assessments of responsibility. (See Hithcock MS for

[6] More specifically, the accounts are formulated in terms of 'non-backtracking' counterfactuals, without broaching the controversial issue of whether these can be analyzed in non-causal terms.

further discussion). It applies, for example, when attributing moral or legal responsibility, or when analyzing system failure in engineering.

The real work that I think needs to be done is that of providing useful taxonomies for causal relationships. Suppes' critique of Russell suggests one way in which we might classify causal relationships: by functional form. When we are dealing with precise quantitative variables, it may be possible to capture a causal relationship between variables in a mathematical equation. But even when a causal relation does not lend itself to precise mathematical characterization, we can still ask questions about the qualitative structure of the relationship: Do increases in one variable lead to increases or decreases in another? Is this relationship monotonic? Do the details of the effect depend upon the details of the cause in a fine-grained way, or is it more of an on-off affair? Do the various causal variables act independently, or do they interact with one another?

In addition to questions about the functional form of a causal relationship, we can also ask questions about the paths of causal influence: Does one factor cause another directly (relative to some level of graining), or via some third factor? Does the cause work by direct physical contact, or by removing an obstacle impeding the effect? Does one factor influence another in more than one way?

Let me say a little more about this last question, since it's a topic that I have explored in some detail (Hitchcock 2001b). Consider Hesslow's well-known example involving birth control pills. Thrombosis is a worrisome potential side effect of birth control pills. And of course, birth control pills are very effective at preventing pregnancy. Now it turns out that pregnancy itself is a significant risk factor for thrombosis. (Since birth control pills essentially mimic the hormonal effects of pregnancy, it is not surprising that the two have similar side effects.) These facts are represented in Figure 3.1, where the thickness of the line represents the strength of the influence, and the solidity of the line represents the direction of the influence. For many women—such as young, fertile, sexually active, non-smokers—birth control pills *lower* the overall chance of thrombosis. How should we describe this situation? Do birth control pills *cause* thrombosis, or *prevent* it? I describe this situation by saying that there are two different routes whereby the consumption of birth control pills affects thrombosis: one in which birth control pills act 'directly' (relative to the coarse level of grain used here), and one in which pills affect thrombosis by affecting pregnancy. Birth

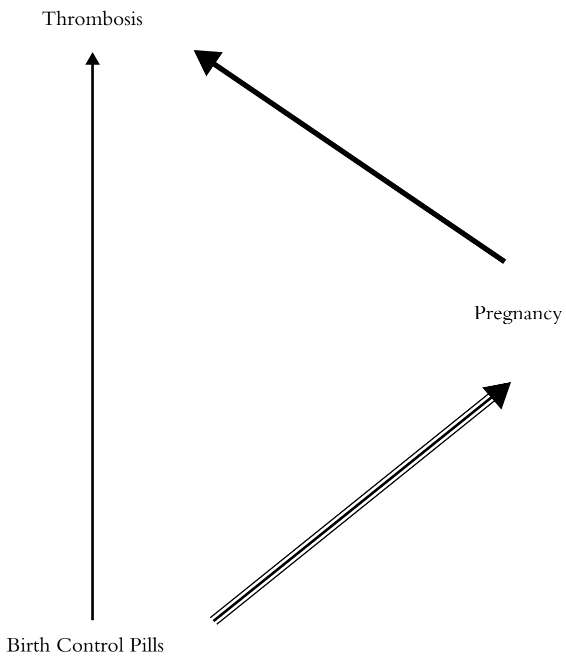

Figure 3.1

control pills have a *component effect* along each of these routes: along the direct route, the effect is positive, i.e. promoting of thrombosis; along the second route, the effect is preventative. The consumption of oral contraceptives also has a *net* effect on thrombosis, which is preventative. The terminology is intended to evoke an analogy with net and component forces in Newtonian Mechanics, although in the case of effects, there is nothing resembling the elegant additive rule for combining forces.

Net and component effects will be relevant in different circumstances. If a woman wants to know whether taking birth control pills is an effective way to prevent thrombosis, then we need to look at the net effect: if she wants to avoid thrombosis, she should take the pills.[7] On the other hand, if we want to know whether the manufacturer of the pills can be held liable for a particular woman's thrombosis, we look at the direct component effect, rather than the net effect.

FN:7

[7] Of course, you might think that she could gain the benefits without paying the costs by using an alternative form of birth control. In fact, however, even the relatively small difference in efficacy between birth control pills and other forms of contraception may be enough to produce a negative net effect.

This discussion illustrates just one dimension along which causal relationships can be classified; there are many others. And despite the flaws of various extant theories of causation, these theories often serve well enough to make the kinds of distinction that are needed. For example, within a probabilistic framework (such as that of Cartwright 1983, e.g.), it is relatively easy to account for the difference between net and component effects. In probabilistic theories of causation, one evaluates the causal relevance of *C* for *E* by comparing the probabilities of *E*, conditional on *C* and *not-C*, *while also conditioning on other relevant factors*. The difficulty, of course, lies in saying just what these other factors are; without providing an adequate answer to this question, we will not have an adequate theory of causation. Nonetheless, even this basic framework is helpful. Suppose that the set of background factors conditioned on does *not* include pregnancy; then, we would expect that conditioning on pill consumption would *lower* the probability of thrombosis. This captures the net effect of pregnancy on thrombosis. But now suppose that we do condition on pregnancy (or its absence). Then, we would expect that conditioning on oral contraception would increase the probability of thrombosis; that is, among women who do in fact become pregnant, those who take birth controls are more likely to be stricken by thrombosis, and likewise among women who do not in fact become pregnant. This probability comparison captures the component effect of pregnancy of thrombosis along the direct route. In effect, by conditioning on pregnancy, we eliminate any effect of birth control pills on thrombosis along the indirect route, leaving only the effect along the direct route.[8] Thus the probabilistic notion of conditionalization, while it may be inadequate for purposes of analyzing causation, can nonetheless be used to shed light on the distinction between net and component effects.

FN:8

Let us turn, finally, to Russell's second claim:

> ... the word 'cause' is so inextricably bound up with misleading associations as to make its complete extrusion from the philosophical vocabulary desirable ...

I agree with Russell that the use of the word 'cause' can be misleading. This is not because the notion of cause is fundamentally misguided, but rather because there are many types of causal relationship. When we say

[8] Capturing the component effect along the indirect route is trickier. It requires either graining more finely, so that there will be intermediate factors along the route that is direct in the coarse-grained picture; or using a special kind of chaining procedure. By far the best technical account of route-specific causal dependence is Pearl (2001).

that taking birth control pills causes thrombosis, we are saying something that is true in one sense, and false in another. Absent clarification, such usage has the potential to cause confusion. What is needed is not so much the elimination of the word 'cause', as the introduction of more precise terms for characterizing causal concepts.

There is in all of this some good news. Many philosophers are, of course, interested in the notion of cause for its own sake. But philosophers are also interested in causation because it seems to be an ingredient in many other philosophically interesting concepts: freedom, moral responsibility, perception, rational choice, explanation, and perhaps also knowledge and reference. There is an old-fashioned view, due in no small part to Russell's critique, that goes something like this:

> We cannot observe causal relations directly, therefore, unless causation can be analyzed in terms of empirically respectable notions, causation remains a part of spooky metaphysics. Until causal theorists are able to provide such an analysis, we should avoid the use of causal notions in our philosophical projects.

This view holds many philosophical projects hostage until causal theorists achieve the impossible. The alternate picture I am presenting offers a much more sanguine vision of how causal theorists may contribute to other philosophical projects. There are of course many hard philosophical problems that may well be held hostage to the hard problems of causation: the problem of free will and determinism within a physical world may well be of that sort. But there are many philosophical projects that stand to benefit from progress on the easy problems of causation.

Consider causal theories of reference, for example. The central idea here is (very roughly) that part of what it means for our word 'tiger' to refer to tigers, is for our current usage of that word to be caused by some appropriate baptismal event. How should I react to this proposal? My reaction (the right reaction, I hasten to add) is *not* the following:

How can it help to explain reference in terms of something as mysterious as causation? Until someone can explain to me what causation is in terms of something more respectable, like empirical regularities or even the laws of physics, I just don't understand it.

My reaction is rather along the following lines:

> What, exactly is the causal relationship that's supposed to hold between the baptismal event, and my current usage? Will any old causal relationship do?

For example, suppose the relationship between a baptismal event and our current usage is as follows.

The warriors of a particularly bellicose tribe were on their way to massacre the members of a new upstart tribe, but were prevented from doing so when they were waylaid by tigers. They had never seen tigers before, and by sheer coincidence, named them 'lions' then and there. This tribe was soon wiped out completely by voracious tigers. As it happens, the upstart tribe that was saved by the tigers was the first to speak proto-indo-european. Thus if the baptismal event, the initial encounter with the tigers, had not occurred, then we would be speaking a very different language, and would not be using the word 'lion' today.

Is this the right sort of connection between a baptismal event and our current usage? It certainly seems to be causal. If this had really happened, would it follow that our word 'lion' refers to tigers? If not why not?

More generally, for just about any philosophical concept that has a causal dimension, not just any old causal relationship will be relevant. Analytic work thus stands to benefit a great deal from a better taxonomy of the ways in which things can be causally related to each other. It is in attacking these 'easy' taxonomical problems that causal theorists have the potential to provide the greatest benefit for philosophy at large.

References

Bergson, H. (1910). *Time and Free Will: An Essay on the Immediate Data of Consciousness* (F. L. Pogson trans.). London: George Allen and Unwin.

Bialas, P. Z., Burda, J., Jurkiewicz, Krzywicki, A. (2003). 'Tree Networks with Causal Structure'. *Physical Review, E* 67: 066106.

Brightwell, G., Dowker, F., García, R. S., Henson, J. and Sorkin R. D. (2003). '"Observables" in causal set cosmology', *Physical Review, D* 67: 084031

Cartwright, N. (1983). 'Causal Laws and Effective Strategies' in *How the Laws of Physics Lie*. Oxford: Oxford University Press, 21–43.

Chalmers, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.

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

Earman, J. (1986). *A Primer on Determinism*. Dordrecht: Reidel.

Hitchcock, C. (2001a). 'The Intransitivity of Causation Revealed in Equations and Graphs'. *Journal of Philosophy*, 98: 273–99.

____ (2001b). 'A Tale of Two Effects'. *The Philosophical Review*, 110: 361–96.

____ (MS) 'Token Causal Structure'.

Hitchcock, C. and Woodward, J. (2003). 'Explanatory Generalizations, Part II: Plumbing Explanatory Depth'. *Noû s*, 37: 181–99.

Hobart, R. E. (1934). 'Free Will as Involving Determinism and Inconceivable without It'. *Mind*, 43:.1–27.

Hull, D. (1992). 'That Just Don't Sound Right: A Plea for Real Examples', in J. Earman and J. D. Norton (Eds) *The Cosmos of Science*. Pittsburgh: University of Pittsburgh Press, 428–57.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*.

Kant, I. (1781). *A Critique of Pure Reason*.

____ (1787). *A Prolegomenon Toward Any Future Metaphysics*.

Lewis, D. (1979). 'Counterfactual Dependence and Time's Arrow'. *Noûs*, 13: 455–76. Reprinted in Lewis (1986b), 32–52.

____ (1986a). 'Postscripts to 'Causation',' in Lewis (1986b) 172–213.

____ (1986b). *Philosophical Papers*, *Volume II*. Oxford: Oxford University Press.

Mackie, J. L. (1974). *The Cement of the Universe*. Oxford: Clarendon Press.

Menzies, P. (1989). 'Probabilistic Causation and Causal Processes: A Critique of Lewis'. *Philosophy of Science*, 56: 642–63.

Mill, J. S. (1843). *A System of Logic*. London: Parker and Son.

Moriya, Y., Kawaji, H., Tojo, T. and Atake, T. (2003). 'Specific-Heat Anomaly Caused by Ferroelectric Nanoregions in Pb(Mg(sub 1/3]Nb(sub 2/3])O(sub 3] and Pb(Mg(sub 1/3]Ta(sub 2/3])O(sub 3] Relaxors'. *Physical Review Letters,* 90: 205901.

Norton, J. (this volume). 'Causation as Folk Science.'

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

____ (2001). 'Direct and Indirect Effects', in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. San Mateo CA: Morgan Kauffman, 411–20.

Putnam, H. (1990). 'Realism with a Human Face,'in J. Conant (Ed). *Realism with a Human Face*. Cambridge: Mass: Harvard University Press, 3–29.

Russell, B. (1913). 'On the Notion of Cause'. *Proceedings of the Aristotelian Society*, 13: 1–26.

Salmon, W. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Schaffer, J. (2001). 'Causes as Probability-Raisers of Processes'. *The Journal of Philosophy*, 98: 75–92.

Spirtes, P., Glymour, C. and Scheines, R. (2000). *Causation, Prediction, and Search,* Second Edition. Cambridge, MA: M.I.T University Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North-Holland.
Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
Woodward, J. and Hitchcock, C. (2003). 'Explanatory Generalizations, Part I: A Counterfactual Account'. *Noûs*, 37: 1–24.