Mind Mach
DOI 10.1007/s11023-006-9033-3

Dispositional versus epistemic causality

Jon Williamson

Received: 13 July 2006 / Accepted: 13 July 2006

Abstract I put forward several desiderata that a philosophical theory of causality should satisfy: it should account for the objectivity of causality, it should underpin formalisms for causal reasoning, it should admit a viable epistemology, it should be able to cope with the great variety of causal claims that are made, and it should be ontologically parsimonious. I argue that Nancy Cartwright's dispositional account of causality goes part way towards meeting these criteria but is lacking in important respects. I go on to argue that my epistemic account, which ties causal relationships to an agent's knowledge and ignorance, performs well in the light of the desiderata. Such an account, I claim, is all we require from a theory of causality.

Keywords Causality · Causation · Dispositions · Capacities

Introduction

In this paper I shall compare two philosophical theories of causality, Nancy Cartwright's dispositional account of causality and my own epistemic view. These two accounts proffer two different reactions to David Hume's scepticism about causal connections:

> It appears that, in single instances of the operation of bodies, we never can, by our utmost scrutiny, discover any thing but one event following another; without being able to comprehend any force or power by which the cause operates, or any connexion between it and its supposed effect. ... One event follows another; but we never can observe any tie between them. They seem *conjoined*, but never *connected* (Hume, 1748, paragraph 58).

J. Williamson (✉)
Philosophy Department, SECL, University of Kent, Canterbury CT2 7NF, UK
e-mail: j.williamson@kent.ac.uk

Nancy Cartwright claims that causal connections do exist, in the shape of *capacities*, which are dispositions of causes to produce their effects (Cartwright, 2003):

> The generic causal claims of science are not reports of regularities but rather ascriptions of capacities, capacities to make things happen, case by case. ... aspirins have the capacity to relieve headaches, a relatively enduring and stable capacity that they carry with them from situation to situation; a capacity which may if circumstances are right reveal itself by producing a regularity, but which is just as surely seen in one good single case (Carwright, 1989, pp. 2–3).

Cartwright is a realist about these dispositions, arguing that capacities are required to make sense of science and that consequently they must be presumed to exist as mind-independent entities, as real as tables and chairs (Cartwright, 1989, p. 1).

Cartwright may well be right about the pervasiveness and importance of causal claims in science. But, I'd argue, that should not force us towards realism about dispositions. Instead I advocate *epistemic causality*. This position develops Hume's view that causal connections are a type of mental connection. Accordingly, causality is a feature of the way we represent the world, rather than a direct feature of the world itself.

The plan is first to outline a number of desiderata, properties that I would expect any satisfactory philosophical theory of causality to have (§2). We shall see that many contemporary theories stumble at one or other of these hurdles. In §3 I shall evaluate Cartwright's dispositional account, arguing that while it has attractive features, some limitations are brought to light by the desiderata. I shall then, in §4, describe my epistemic account. This view, I shall argue, copes well with the desiderata (§5). Finally in §6 I shall close with some very brief remarks about the two accounts of causality and the connection between causality and probability.

Desiderata

I suggest that a philosophical theory of causality should

Objectivity: account for the objectivity of causality,
Calculi: underpin calculi for causal reasoning,
Epistemology: motivate a viable epistemology,
Variety: cope with the great variety of causal claims that are made,
Parsimony: be ontologically parsimonious.

We shall run through these desiderata in this section.

Objectivity

There are two main senses in which causal claims may be described as objective. First, there appears—at least in most domains—to be a fact of the matter as to what causes what; a causal relationship is not a matter of subjective opinion. Thus if two agents disagree as to a causal relationship, at least one of them must be wrong. Second, one might think that causal relations are objective in the sense that they exist as a part of agent-independent physical reality, disjoint from a subject's mental or epistemic activity.

I take it that a philosophical theory should render causality objective in the former, logical, sense. By and large, causal claims are objectively true or false, and a philosophical theory that fails to explain this objectivity is lacking in an important respect.[1]

I do not take objectivity in the latter, existential, sense to be a requirement. Different philosophical theories may take different stances on this issue, but it would be other considerations that would decide between them. One might disagree and consider objectivity in this second sense to be independently desirable, presumably by appealing to intuition or to linguistic practice. But those of us who do have some intuition about the location of causal relationships will surely admit that this sort of intuition is rather unreliable: ontological intuitions have routinely been overturned in the history of physics, let alone philosophy. Moreover, linguistic practice may be no better an ontological indicator. We do of course talk of causal relationships as if they are out there: we say 'the wind caused the tree to fall', apparently attributing the causing to the wind, with no mention of our role as knowing subjects. However, this linguistic practice may indicate logical objectivity rather than ontological objectivity. Perhaps we talk that way, not because our talk somehow reflects the nature of reality, but because it's simple and because we can: the truth of the claim does not vary from agent to agent, so any mention of agents would introduce an unnecessary complication.[2] If this is plausible then to infer from talk to the nature of reality is straightforwardly fallacious (Mill, 1843, §V.III.6). As Edwin Jaynes warns,

> Common language—or, a least, the English language—has an almost universal tendency to disguise epistemological statements by putting them into a grammatical form which suggests to the unwary an ontological statement. ... To interpret the first kind of statement in the ontological sense is to assert that one's own private thoughts and sensations are realities existing externally in Nature. We call this the 'mind projection fallacy' (Jaynes, 2003, p. 22).

Some accounts of causality apparently fall at this first hurdle. *Perspectival* or *agency* accounts of causality which deem causal relationships to vary according to the situation or capabilities of different agents yield a relativistic notion of causality—a notion that is not objective in the logical sense (Menzies & Price, 1993; Price, 2005; Williamson, 2006, §6).

[1] By suggesting that there is a fact of the matter about what causes what, I do not intend to commit to the existence of facts. Causal claims are objective in the logical sense if they are true or false; this says nothing about the nature of the truth-makers of these claims. It is up to particular theories of causality to say something about truth-makers. I accept that there may be exceptions to logical objectivity. It may be subjective as to whether growing vines on west-facing slopes causes the resulting wine to taste of chestnuts. One might also claim that causal relationships are not fully objective in the logical sense because they are relative to population: e.g., having a degree in philosophy may be a preventative of poverty in one population, but a cause of poverty in another. Moreover, some maintain that causal claims are relative to context or to contrasting claims—see e.g., Schaffer (2005) and references therein. I leave such questions open here. I require only that a theory of causality should generate causal relationships that are 'by and large' objective in the logical sense: different theories may understand this nebulous qualification in different ways, which is fine as long as they do not clash with clear cut intuitions of objectivity.

[2] Moreover we often try to tailor language to instil an air of objectivity: academic writing is increasingly depersonalised, with the passive voice to be heard above that of any human subject; similarly in science there is a tendency to depersonalise and objectify. The objectivity we are striving for is surely logical rather than ontological.

Calculi

A philosophical theory of causality should explain the successes and failures of current calculi for causal reasoning.

Many common reasoning tasks are causal. We often need to predict the effects of an observed event, for instance, or to diagnose the cause of observed symptoms, or to predict the effects of several potential actions in order to decide which to instigate. Formal methods have long been used to aid and to automate such tasks. Currently in vogue are causal Markov methods such as Bayesian nets and structural equation models (See e.g., Pearl, 2000). These are formalisms that take as their starting assumption the Causal Markov Condition (which says that every variable under consideration is probabilistically independent of its non-effects in the model, conditional on its direct causes).

Of course any formalism must be used with a certain amount of caution. Each formalism has an associated range of safe application—those applications in which the formalism will give reliable results. Reliability depends on several factors: a formalism must be rich enough to treat the salient complexities of a potential domain of application, and it helps if the assumptions on which the formalism is based are true, or at least hold to some acceptable degree of approximation. This last requirement will of course depend on the interpretation of the terms in the assumptions. For example, the truth of the Causal Markov Condition depends on how one understands its terms, e.g., 'probabilistically independent' and 'causes' (Williamson, 2005). In this case considerations from the philosophy of probability and the philosophy of causality will be relevant. A philosophical theory of causality should thus tell us how to interpret the central assumptions of contemporary formalisms for causal reasoning, to thereby determine when the assumptions might be expected to hold, and to give an idea of each formalism's range of safe application.

Probabilistic accounts of causality often falter at this step. Such accounts tend to over-egg the Causal Markov Condition, taking the condition (or a special case of the condition, the Principle of the Common Cause) as a basic feature of causality, sometimes even a condition whose failure cannot be countenanced (Papineau, 1992; Pearl, 2000, §2.9.1; Spirtes, Glymour, & Scheines, 1993, §3.5.1; Spohn, 2002; Suppes, 1970). Consequently, probabilistic accounts come unstuck in situations where the condition fails. Counterexamples to the condition are widely documented and occur where probabilistic dependencies arise that are not solely attributable to causal relationships (See e.g., Williamson, 2005, Chapter 4).

Epistemology

It is not enough for a philosophical theory of causality to take the form 'causality is X'. As well as giving an account of of what causality is, it should also say how we can come to know causal relationships and why, when we apply the appropriate methods, we learn about X-causality and not Y-causality, some other notion of cause. (It seems in principle possible to come up with two coherent metaphysical notions of cause, X and Y, which perform equally well according to other desiderata. If it is unclear as to how one can come to learn X-causal relationships then the X-theory is clearly lacking, and Y-theory should be preferred if it does more to motivate a viable epistemology.)

This desideratum is perhaps controversial. Some metaphysicians might advocate the view that it *is* enough to give a coherent theory of the nature of causality. Indeed, one might argue that it is preferable to remain neutral regarding methods for causal learning (and calculi for causal reasoning) because these are matters for science and liable to change as science progresses: a commitment to particular methods will render the philosophical theory redundant as the methods are superseded. But such a stance does a disservice to metaphysics, for the following reasons. As suggested above, it can lead to under-determination: there may be umpteen consistent notions of cause; the only way to tell which is best may be to see which fits best with causal methodology and epistemology. Moreover, it divorces metaphysics from practice, and if metaphysics has no practical ramifications then for many there will be no reason to pursue it. On reflection, then, there seems to be little to recommend this divorce: it provides one less way of refuting metaphysical theories and one less way of improving them. If progress in science leads to improvements in metaphysics, so much the better for metaphysics.

Counterfactual accounts of causality can fall foul of this desideratum. David Lewis, for instance, developed a philosophical theory of causality which construes causal relationships in terms of counterfactual conditionals, and in turn understands counterfactual conditionals in terms of similarity of possible worlds (Lewis, 1973). Lewis remained neutral about causal epistemology. Accordingly, procedures for learning causal relationships must be gleaned from current scientific method. But it is hard to see how current techniques for learning causal relationships can at base be techniques for learning about similarity of possible worlds. More plausibly, current techniques attempt to find physical mechanisms or functional or probabilistic dependencies. If so, theories of causality which appeal to mechanisms or dependencies fare better than Lewis' account with respect to this desideratum.

Variety

We use causal language liberally and our causal claims are multifarious. Here are some examples of the variety of causal claims. Causal relationships can be positive or negative: smoking causes heart disease; exercise prevents it. Causal claims can be single-case or generic: Bart's bad behaviour this morning got him into trouble; in general, bad behaviour leads to trouble. Causal claims can be individual-level or population-level: the inequality of her compatriots (population-level but single-case) is a cause of Naomi's deterioration in health (individual-level, single-case) (Glymour, 2003); in Britain, inequality (population-level, generic) causes deterioration in health (population-level, generic). Causal relationships can complete or be pre-empted: firing the gun killed the man; his death by poisoning was thus pre-empted. Both presences and absences can be causes and effects: striking a match causes a flame; a lack of oxygen prevents a flame. Events, variables and properties can be causal relata, as can causal relationships themselves: the big bang caused the expansion of the universe; interest rate is a cause of inflation; the redness of the flower attracted the insect; smoking causing cancer causes governments to restrict tobacco advertising (Williamson & Gabbay, 2005). Causal relata can be concrete or abstract: the brick hitting his head caused Desmond's headache; the lack of a solution to the $P = NP$ problem causes headaches for computer scientists. As these examples show, causal claims are varied and ubiquitous—they crop up in the natural, social & biomedical sciences, as well as daily life.

A philosophical theory of causality should be able to cope with this variety of causal claims. If the theory takes one type of causal relationship as primitive, it should say how other types of claim can be reduced to the primitive relationships. It should also (c.f. Epistemology) explain how we can come to know these other types of claim and how we can reason with them (c.f. Calculi). Otherwise, it can at best be said to be a theory of just one type of causal relationship, rather than a theory of causality in general.

Mechanistic accounts of causality do not do particularly well here. Perhaps the leading mechanistic theory is that advocated by Salmon 1998 and Dowe 2000. Under this view mechanisms are construed as physical processes which possess or transmit a conserved quantity (such as charge). This account is designed to cope with causality in physics, but seems remote from causality in the social sciences or daily life. When we learn that interest rate is a cause of inflation it is hard to see how we can be finding out about linear momentum or some other conserved physical quantity; the fundamental quantities of interest are rather interest rate and inflation themselves.

Parsimony

Finally, a theory of causality should be ontologically parsimonious—it should posit as few new kinds of entity as possible.

Any theory of causality which views causal relationships as *primitive* and unanalysable would seem to suffer at the the hands of parsimony. Note, though, that evaluations of parsimony are relative to ontology. If one's ontology contains only causal relationships then this theory of causality is ontologically parsimonious. On the other hand if one's ontology contains other things, then the theory requires the addition of causal relationships to the ontology and is not parsimonious.

The parsimony desideratum is relatively uncontroversial, though hard to justify. It seems hopeless to argue for parsimony on the grounds that the world is simple—if it were so simple and parsimonious, science would not be so hard and entity-rich. A more promising strategy is to accept parsimony as a methodological or psychological requirement. In practice we need parsimony because we find simple theories attractive and find complex theories hard to understand and use. Desiderata are things desired; there is no reason to deny aesthetic and pragmatic desires. Having said that, the parsimony desideratum may be considered subsidiary to the other desiderata if aesthetic and pragmatic desires are judged less important than the reasons for accepting the other desiderata: *other things being equal*, a theory of causality should be ontologically parsimonious.

If we admit aesthetic desires, won't our list of desiderata veer into the realm of the subjective? Perhaps we should accommodate someone who, for instance, desires only theories that are written in rhyming couplets. But this is a step too far. The idea behind the desiderata is to find some yardstick to compare theories—if this yardstick is to be used to convince others as to the merit of a theory or the problems that beset another, then there needs to be at least intersubjective agreement as to the desiderata involved. Thus compelling reasons need to be produced to introduce new desiderata.

The desiderata considered above have not touched on the topic of explanation. One might think that a philosophical theory of causality ought to elucidate the connection between causality and explanation, and one might want to introduce a corresponding desideratum. I think that would be a mistake, for the following

reasons. First, explanation is itself a distinct and rather intractable philosophical topic: it would be too much to expect a philosophical theory of causality to contain a philosophical theory of explanation. Second, there are an extremely wide range of views on this question. While at one extreme some think that all explanation is causal explanation, at the other end of the spectrum there are those who disavow any principled connection between the two—on the grounds, for example, that causality is an objective notion but an explanation is subjective (an explanation is only successful to the extent that it furthers a particular agent's understanding, and understanding is inherently subjective, the domain of psychology). To insist that a theory of causality forge some connection with explanation would prejudge against the latter camp.

Similar remarks can be made about other potential desiderata. For instance, some might expect a philosophical theory of causality to forge a connection with decision theory; others may expect it to explain the connection between the causal asymmetry and the temporal asymmetry; others may expect it to account for the quirks and fallacies of human causal reasoning. Again I think it would be a mistake to enlarge our list of desiderata to make room for these issues. I do not wish to suggest that these desires or expectations should be entirely discounted—just that they hinge on controversial presuppositions that may be legitimately disputed. Arguably one should not dismiss a theory of causality on the grounds that it divorces causality and decision theory, or because it rejects any non-accidental link between causal direction and temporal direction, or because it neglects to offer an account of human practice. On the other hand, I have suggested that a theory which fares badly with respect to objectivity, calculi, epistemology, variety and parsimony should be rejected. While these latter are desires according to which a theory should be judged, the former may better be construed as hopes. (That said, not a lot hangs on the five desiderata outlined in this section being exhaustive: one may conceive of them as core desiderata and others as ancillary or subsidiary.)

Having introduced some means of evaluating philosophical theories of causality we shall now turn to an evaluation of Cartwright's dispositional theory.

Evaluating dispositional causality

In this section we shall run through the desiderata to see how Cartwright's dispositional theory of causality fares.

First Objectivity. For Cartwright, capacities are dispositions of causes to produce their effects. Dispositions satisfy what Cartwright calls the Two-Sidedness property: 'there is a distinction between the occurrence of a disposition and its being manifested' (Cartwright, 2003, p. 7). Cartwright argues that capacities are dispositions, and 'If the capacity is triggered properly and *is not interfered with*, then the canonical manifestation will result' (Cartwright, 2003, p. 10):

> The central feature that locates capacities in the family of dispositions, I claim, is ... malleability. All other members of the family of dispositions, habits and character traits seem to have at least one of the three central features of malleability I mentioned: They need triggering, they can be enhanced or weakened or they may produce different manifestations, or no manifestations at all, if they are interfered with (Cartwright, 2003, p. 8).

For Cartwright capacities are real—they are objective in the ontological sense:

> capacities are much like essences. If you are committed to the assumption that all of the internal properties of electrons are essential, this makes science a lot easier for you. You can measure the charge or mass on one, and you know it on all the others (Carwright, 1989, p. 148).

Therefore they are also objective in the logical sense:

> What makes everything in the disposition family belong there is two-sidedness. Within this family ''capacity'' seems especially like a power word. Nevertheless, I think that is the wrong way around to look at it. What marks out all capacities as capacities is not primarily that they enable systems to do things, but rather that they can be stopped—they can be interfered with. But that does not rule them out of science. Whether something is an interference in a given situation is a matter of fact; and it is a fact we can know about (Carwright, 2003, p. 12).

This is clearly an attractive feature of Cartwright's account: not only does it yield a notion of cause that is logically objective, it also provides an explanation for this objectivity, attributing it to ontological objectivity.

Cartwright goes half way towards meeting the Calculi desideratum by shedding a great deal of light on potential failures of causal Markov methods (Carwright, 1999, Chapter 5; 2001). The trouble is that Cartwright argues that failures of the Causal Markov Condition are ubiquitous; this makes it a surprise that causal Markov methods work at all. Perhaps they don't work, but then their prevalence in a wide range of fields—including data mining, econometrics and epidemiology—becomes a mystery. What is lacking then, is an explanation of the apparent success of causal Markov methods. Moreover, in the absence of causal Markov methods, we need some guidance as to how we should reason with causal knowledge. Cartwright advocates a bootstrap approach with respect to structural equation models: if their fundamental assumptions hold then one can discover causal relationships with them and one can reason with them (Carwright, 1989, Chapter 1). But the Causal Markov Condition is one of those fundamental assumptions (it is usually couched in terms of independence of error terms in structural equations). So if the Causal Markov Condition is likely to fail, we are lacking a calculus for causal reasoning.

This latter difficulty also impacts on the epistemology desideratum. If the assumptions behind structural equation models rarely hold, then we can not use them to discover causal relationships after all. But this is just the type of method that Cartwright advocates for learning causal relationships.[3]

On the plus side Cartwright does try to link the epistemology of causality to her metaphysics of capacities. The epistemology of causality to which Cartwright appeals is fairly standard, involving model building and the use of experiments. But Cartwright argues that capacities are required to understand these methods. The argument is not altogether conclusive however: even if we grant Cartwright that models ought to involve talk of capacities, it is not clear why one must commit to capacities

[3] Cartwright suggests that we glean causal laws from a *nomological machine* which 'has fixed components with known capacities arranged in a stable configuration' (Carwright, 1999, p. 122). A nomological machine is a type of model, so to discover causal relationships we must build a model that incorporates causal language, such as a structural equation model (Carwright, 1999, p. 55).

being *real*. Models are representational devices; they do not need to correspond to reality to permit successful inferences (indeed they often obviously do not). In her defence of real capacities Cartwright argues that any such rejection of the reality of capacities is unstable (Carwright, 1989, §4.4). The reason being that there is no epistemological demarcation between those basic things we accept as real and the concepts like causality which are viewed as representational; we learn about the former in the same way that we learn about the latter, so we had better either accept the reality of all these things or reject the reality of all these things. But this misses the point—it puts the cart before the horse. The fact is that in practice we find the reality of some things, such as tables and chairs, unproblematic. Other things, like causal connections, we have genuine trouble with, as the passage from Hume in §1 testifies. The question arises as to whether the problematic things might not be out there. One can hold the view that the unproblematic things are real and the problematic things are representational by appealing to a demarcation that is metaphysical rather than epistemological. Although we may learn about the real things and the representational things in the same way, we find the reality of some things unproblematic simply because they *are* real and therefore don't generate problems when they are classed as real; we find the reality of other things problematic simply because they are representational and therefore don't fit well into the mould of real things. In sum, while Cartwright attempts to tie epistemology to capacities, her epistemology fails to motivate the real-capacity theory over the representational-capacity theory. (We shall see in §4 that the epistemic theory views causality as representational rather than real.)

We turn to the Variety desideratum. Cartwright takes single-case causal relationships to be fundamental, with causal laws reducible to these single cases. The relata of causal relationships are varied: capacities can attach to properties or to objects (Carwright, 1999, pp. 53–54). However, it is hard to see how capacities might attach to causal relationships themselves to make sense of statements like 'interest rate rises causing a slow-down in the housing market causes interest rate rises.' Worse, if capacities are real they won't attach to unreal things like absences to cope with claims like 'an absence of oxygen extinguishes a flame'. One way to accommodate absences is to treat them as values of variables. Thus one can construct a variable oxygen which takes possible values *present* and *absent* and a similar variable flame and claim that oxygen causes flame. But variables are representational devices, not things to which real capacities can attach. Other representational and non-physical relata are equally problematic: if rate of change of salary causes stress, or the axiom of choice causes the Banach-Tarski paradox, then capacities ought to attach to rates of change or axioms. In which case it appears that some capacities are real but non-physical—are they abstract entities? How do they interact with physical capacities? At best, Cartwright's theory leaves important questions unanswered. More pessimistically, its commitment to real capacities leaves these questions unanswerable.

Parsimony is much on Cartwright's mind: 'The point of this book is to argue that we must admit capacities, and my hope is that once we have them we can do away with laws. Capacities will do more for us at a smaller metaphysical price' (Carwright, 1989, p. 8). So the obvious metaphysical expense of the inclusion of capacities in one's ontology is to be offset by ridding laws from it. Clearly this only avoids metaphysical overdraft if (i) laws can be reduced to capacities, (ii) laws can not be reduced to anything else in the ontology, and (iii) capacities are no more of an ontological

commitment than laws. All of these claims are controversial, but (iii) is particularly implausible in the light of the above discussion. Capacities must enter into an ontology not just in great number—think of the number of capacities that a humble brick has—but also in great variety, attaching to objects, properties and perhaps abstract entities.[4] Thus they are ontologically expensive. On the other hand, Cartwright argues variously that laws are hard to find, false, or lacking in universality (Cartwright, 1983; 1989, p. 8; 1999, §1.2), none of which suggests ontological expense approaching that of capacities. Assuming the existence of capacities, then, goes well beyond assuming the existence of the things they attach to, and it is hard to see how laws, if as elusive as Cartwright suggests, could compensate for this lack of parsimony.[5]

While I have focussed above on its limitations, Cartwright's dispositional view must surely be one of the leading philosophical theories of causality: it makes headway on all five desiderata outlined in §2, which is more than most theories do. It successfully accounts for the objectivity of causality and makes a great deal of progress towards understanding contemporary calculi. I have argued that Cartwright's insistence on the reality of capacities leads to shortcomings with respect to the other three desiderata. The epistemic theory, to be introduced next, rejects the reality of causality. We shall examine in §5 the extent to which this pays off with respect to the desiderata.

Epistemic causality

The starting point of the epistemic theory of causality is David Hume's idea that causal connection or power is something mental:

> after a repetition of similar instances, the mind is carried by habit, upon the appearance of one event, to expect its usual attendant, and to believe that it will exist. This connexion, therefore, which we *feel* in the mind, this customary transition of the imagination from one object to its usual attendant, is the sentiment or impression from which we form the idea of power or necessary connexion (Hume, 1748, paragraph 59).

Immanuel Kant concurred:

> To the synthesis of cause and effect there belongs a dignity which cannot be empirically expressed, namely, that the effect not only succeeds upon the cause, but that it is posited *through* it and arises *out of* it (Kant, 1781, B124).

The following thesis can be motivated by the writings of Hume and Kant (Williamson, 2005, §9.2). First, the notion of production is central to causality: intuitively causes *produce* their effects. But, second, this notion of production is just intuition: production is purely *mental*, a 'transition of the imagination'. Accordingly, there is no physical correlate of production; the world is just set up in such a way that often, when the cause event happens, the effect event happens.

[4] If there are capacities then there must be far more of them than there are things that they attach to, since few if any real objects or properties are genuinely inert and those that aren't can be applied to a vast number of tasks.

[5] Note that the parsimony problem may be resolvable if one resorts to the view that all properties are dispositional—see e.g., Shoemaker (1998), Bird (2005), and Armstrong (2004, §10.4) for a contrary view.

To say that causal connection is mental does *not* imply that causality is subjective (in the logical sense). There may be a fact of the matter about what causes what, since there may be a fact of the matter about which transitions of the imagination are most appropriate. Neither does it imply that that there can be no physical explanation for the occurrence of an effect. One can appeal to facts about the set-up of a snooker table—e.g., the positions, velocities, momenta of the balls—to explain the potting of the black; facts about gravity, energy, materials, positioning, wind etc. explain the sound that occurs after dropping a brick. As Bertrand Russell argued, a full explanation of an effect requires a lot more than its cause and some connection between the two: if not the whole state of the universe then at least a significant chunk of it needs to be invoked (Russell, 1913, p. 8).

The epistemic theory of causality offers one way of fleshing out the idea that causal connection is just a transition of the imagination (Williamson, 2005, Chapter 9). The main object of interest for the epistemic view is a rational agent's causal-belief state. According to the epistemic theory an agent's causal beliefs are like a map—they help her reason about the world; they help her make strategic decisions, predictions, diagnoses, for example. Consider the causal graph of Fig. 1, where the vertices are the variables smoking, bronchitis, lung cancer and chest pains and the arrows indicate direct causal connection. An agent who has causal beliefs that are represented by this graph can predict the occurrence of chest pains when lung cancer or bronchitis is present, can diagnose bronchitis if chest pains are present and lung cancer absent and can decide not to smoke in order to avoid bronchitis, lung cancer and chest pains.

According to the epistemic interpretation of this causal map, the arrows are just heuristic devices, they are just representational. An arrow does not correspond to some physical feature of cause or effect nor to some physical link between the two. Moreover, a full explanation of an occurrence of chest pains would be extremely complex, involving a significant chunk of the history of the universe culminating in the chest pains, rather than just the causes and effects and their connections.[6] That the arrows are just representational does not imply that their placing on the graph is subjective: there may be a fact of the matter as to the correct representation, not because the arrows correspond to reality, but because different representations yield different predictions, diagnoses and so on, and there may be a fact of the matter as to which representation yields optimal inferences.

Consider an analogy. Construct an undirected graph by taking names of towns as vertices and linking two towns if *one can travel between them in 2 h*. Fig. 2 is such a graph, involving Stroud, Birmingham, London and Canterbury. An agent who has travel beliefs that are represented by this graph can make successful inferences. She can predict, for instance, that she can travel between Stroud and Canterbury within 4 h. But the edges of the graph are just representation, they do not correspond to

[6] Of course in some cases the arrows in the causal graph might appear to roughly correspond to the key components of a full physical explanation, e.g., to the salient physical processes or mechanisms. (If so, a causal explanation may look structurally much like the physical explanation.) That does not mean that salient physical features will invariably match causal arrows, as is required for an analysis of causality in terms of these features. Thus the epistemic view does not preclude there being elements of reality that sit between cause and effect as arrows sit between labels of cause and effect; but it does not take such features to be constitutive of causality.

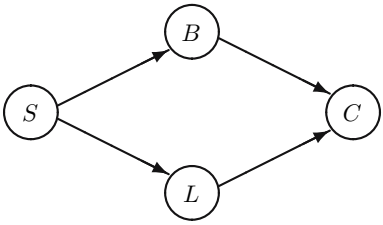

Fig. 1 A causal graph. Smoking, bronchitis, lung cancer and chest pains

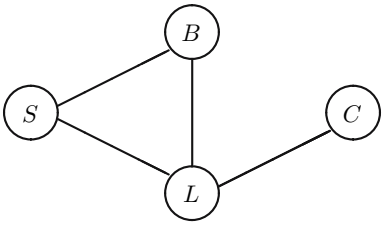

Fig. 2 A travel graph. Stroud, Birmingham, London and Canterbury

some physical link between towns nor to some feature of one or other town.[7] A full explanation of a particular journey from Stroud to Canterbury taking less than 4 h would be extremely complex, involving more than just the towns and links between the two. Neither is the graph subjective: there is a fact of the matter as to the best representation—it is the graph that admits the most reliable inferences.

While the optimal graph may be objectively determined, in practice an agent's belief graph will be relativised to her background knowledge. For example, an agent's travel graph will depend on her knowledge of different modes of transport. If she knows about rail travel as well as road travel, her travel graph is likely to differ from that produced on the basis of knowledge just of road travel. The question thus arises as to what causal beliefs an agent should adopt given her background knowledge. How does background knowledge constrain choice of causal graph?

If a causal graph is valuable to the extent that it yields reliable inferences, and we would like to isolate the best graph given an agent's background knowledge, then we need to identify how causal beliefs give rise to reliable inferences. Arguably causal beliefs allow one to make diagnoses, predictions and strategic decisions because (i) normally (but not always) cause and effect are probabilistically dependent, and (ii) normally (but not always) intervening to change a cause changes its effects. The qualification 'normally (but not always)' does not preclude the utility of causal beliefs for inference—they will not always result in a correct inference, but if inferences are normally correct, that is quite an achievement. (But this qualification does put paid to probabilistic or count-

[7] Do the edges just correspond to generalisations of particular journeys? No. Suppose a travel company decides to introduce a 90-min helicopter shuttle between two far-flung places. But no one uses this shuttle, so it never flies the route and is scrapped after a day. There are no journeys to generalise from, but this knowledge is enough to induce a travel belief that will arguably lead to correct inferences on the day in which the shuttle operates, so there is an edge in the travel graph.

erfactual analyses of causality, which require some invariant feature which can be used to replace talk of causal connections.)[8]

These two features of causality can be combined and made precise by the following claim: normally *strategic dependencies* accompany causal connections. Here cause and effect are said to be strategically dependent if they are probabilistically dependent when intervening to set the value of the cause and controlling for the effect's other direct causes, i.e., they are probabilistically dependent if one holds fixed the other direct causes of the effect and some set of the cause's non-effects that includes its direct causes.

This strategic dependence property motivates an answer to our question about how background knowledge should constrain choice of causal graph (Williamson, 2005, §9.5). Arguably an agent's causal beliefs should reflect her background knowledge but should not go beyond her background knowledge: her causal belief graph should include arrows that are warranted by her evidence but no extra unwarranted arrows. An arrow from A to B is warranted by her background knowledge if she already knows that A causes B (if causality is objective in the logical sense then an agent can have causal knowledge as well as causal beliefs), or if she knows that B is strategically dependent on A and she doesn't know of anything that contra-indicates A causing B. Contra-indications include knowledge that there is some other connection between A and B that accounts for their dependence (e.g., a mereological, logical, mathematical or semantic connection),[9] and knowledge that there is no possible physical explanation for instances of B occurring that would centrally involve instances of A.[10] In sum then, an agent should adopt as a representation of her causal beliefs the smallest causal graph that fits her background knowledge.

Interestingly, if there are no contra-indications to arrows in the causal graph then the Causal Markov Condition must hold (Williamson, 2005, Theorem 9.1). Moreover, if also all arrows in the agent's causal graph correspond to strategic dependencies, then the graph will be a minimal graph that satisfies the Causal Markov Condition (Williamson, 2005, Corollary 9.2). A whole host of methods have been developed to construct such graphs (Pearl, 2000; Spirtes et al., 1993), and current evidence suggests that on average all but the directions of two arrows will be fully determined by background knowledge, i.e., the agent's causal belief graph is nearly

[8] To reiterate: under the epistemic view causal claims are not claims about the world, inasmuch as causal connections are representation not reality. But they do tell us something about the world, since they tell us about effective reasoning and there must be something about the world that makes causal reasoning effective. However, the explanation for the effectiveness of causal reasoning involves more than a simple physical relationship between cause and effect, of the sort that would constitute an analysis of causality. The world is more complicated than causal models make out.

[9] The mereological case includes knowledge that A and B are overlapping physical events and so not the kind of things that are likely to stand in a causal relation—a dependence is likely to be attributable to their overlap rather than a causal connection. The other cases are discussed in Williamson (2005, §4.2).

[10] Perhaps it is known that a B only ever occurs before an A; or that A and B are treated exactly symmetrically by physical theory, so that A could not be said to cause B without also saying that B causes A.

uniquely determined by background knowledge (Gillispie & Perlman, 2002; Williamson, 2005, §9.7).

Having discussed causal beliefs, we move on to causal knowledge and causal facts. What is it for one thing to be a cause of another? Thus far, causal belief graphs vary with background knowledge, but this question asks for an unrelativised causal graph. Perhaps the most straightforward way of generating such a graph is to take it to be the causal belief graph that ought be adopted by an agent who knows everything (everything non-causal, that is) (Spohn, 2002, pp. 166–167; Williamson, 2006, §§ 4.3, 7). It is this ideal or *ultimate* causal graph that determines the facts of causality. Of course it is hard to know what such a causal graph would look like, but presumably as background knowledge increases, an agent's causal belief graph would tend to this ultimate causal graph. If so, well-entrenched causal beliefs would be reliable indicators of arrows in the ultimate causal graph. Entrenched causal beliefs, then, may qualify as causal knowledge.

By focussing on rational causal beliefs, the epistemic theory of causality places causal relationships firmly in the mind of the rational beholder. While not in the mind of any particular agent—no agent knows everything—causal relationships are epistemic in the sense that they are delimited by knowledge: they are defined in terms of rational causal beliefs which are in turn a function of an agent's background knowledge. We have these causal beliefs because of their utility as a basis for inference, not because there is any physical connection between cause and effect that we perceive or infer.

Evaluating epistemic causality

Having sketched the epistemic theory, we shall now see how it fares according to the desiderata of §2.

For any mental theory of causality, the objectivity desideratum is likely to be a significant hurdle. While theories that ascribe reality to causal relationships can invoke the causal relationships themselves as truth-makers to causal claims, any theory for which a causal connection is just a transition of the imagination must give some other grounds for the truth or falsity of causal claims. The epistemic theory takes the ultimate causal graph as the determinant of causal facts: 'A causes B' is true if an omniscient rational agent should believe 'A causes B'. Causal claims are objective, then, to the extent that this ultimate causal graph is uniquely determined—if two omniscient rational agents can disagree as to whether A causes B and neither be wrong, then 'A causes B' is not objectively true or false. Now we saw in §4 that, according to current evidence and assuming no contra-indications, a causal belief graph will be uniquely determined except for the directions of two arrows on average. It has to be said that current evidence is somewhat limited and more must be collected before we can be close to certain about this conclusion for large graphs. But from what we now know, as long as the above result does not significantly change with the introduction of contra-indications, it seems likely that the ultimate causal graph will be to all intents and purposes unique (two arrows pale into insignificance in such a large causal graph). If so, then the epistemic theory yields an

objective notion of cause.[11] Tentatively, then, the epistemic theory overcomes this hurdle.

Moving on to the Calculi desideratum, we must assess the relationship between the epistemic theory and current formalisms for causal reasoning. According to the epistemic theory, the Causal Markov Condition holds if there are no contra-indications to arrows in an agent's causal belief graph. This happens if the agent has no knowledge of non-causal relationships that give rise to strategic dependencies. This is just to be expected: the Causal Markov Condition implies that all probabilistic dependencies are accounted for by causal connections; if it is known that this is not the case, then the condition is bound to fail (Williamson, 2005, §4.2). Thus the epistemic theory tells us when causal Markov methods will work, and when to expect them to fail. Moreover, the theory provides some guidance as to how to proceed when the methods do fail, since it says how an agent's causal graph is determined in such circumstances. As yet, however, formalisms for causal reasoning where the Causal Markov Condition fails are by no means as efficient as the causal Markov methods themselves. So while the theory does the required philosophical work, explaining the successes and failures of current calculi for causal reasoning, there is more formal work to be done.

Next to epistemology. The epistemic theory provides its own account of how a rational agent's causal belief graph is determined by her background knowledge: it is the smallest directed acyclic graph that, given her background knowledge, best accounts for strategic dependencies. This is an epistemology of sorts, and one that is clearly tailored towards the epistemic view: what we learn when we use this method is *epistemic*–causal relationships, rather than some other notion of cause.

Of course this is not a very practical epistemology since it does not provide a method for incrementally or efficiently building the causal belief graph. But standard methods can be used in special cases or as approximations. *If* there are no contra-indications then causal Markov methods can be used to build a causal graph. Controlled experiments will tell us about strategic dependencies, and one can incrementally build up a graph by adding arrows to represent those strategic dependencies that are not contra-indicated by background knowledge—the resulting graph can be viewed as an approximation to the rational causal belief graph. Standard methods should thus be interpreted as ways of getting to the epistemic notion of cause, rather than any other notion.

While in comparison with the dispositional theory the epistemic theory has its work cut out accounting for the objectivity of causality, the tables are turned when it comes to the variety of causal claims. This is because the epistemic theory imposes no constraints on the relata of causality: if causal connection is just a transition of the imagination then the things connected need not be uniform or real either—they can be whatever works; products of the imagination even, the imagination which knows

[11] We need a qualification here. Given the nodes of the ultimate causal graph, the arrows may be close to uniquely determined. But the nodes themselves may not be uniquely determined by full background knowledge: for uniqueness of nodes the background knowledge would have to determine precisely one optimum conceptualisation of the causal relata; this may be a tall order. Uniqueness of nodes is not crucial, though. What matters is that the graph is determined to an extent that accounts for intuitions of logical objectivity. If there is more than one optimum way of conceptualising the domain, but no two conceptualisations lead to incompatible causal graphs (i.e., causal graphs whose arrows intuitively conflict), then causal relationships are logically objective in the required sense.

no bounds. Since it takes no type of causal relatum to be primitive, the epistemic theory need not reduce relata such as absences or abstract entities to other types of relatum. It can take a standard line on prevention and pre-emption, and can take either single-case or generic causal claims to be more basic, or neither. The wealth of causal claims can be integrated into an epistemic account whether they occur in science, daily life, law, art or mathematics. So the epistemic theory has little problem with the variety desideratum. Moreover, since it treats causal relationships homogeneously, it appears to be at an advantage over those theories which are forced to invoke multiple stories to cope with the variety of causal relationships—one for presences, one for absences, one for physics, one for economics and so on—with little to offer in the way of explanation for the fact that we seem to have one concept of cause rather than four, or however many stories are posited.[12]

Parsimony, as noted in §2, is relative to ontology. Presumably an ontology will include some means to interpret rational beliefs of the sort discussed here. If so, then epistemic causality comes at no ontological cost, since the epistemic theory characterises causal relationships in terms of rational beliefs. Of course if this is to be a genuine reduction, the interpretation of rational beliefs must not itself appeal to the notion of causality—otherwise the notion of cause will have to be taken as an ontological extra after all.

A full analysis of rational belief is beyond the scope of this essay, but a few remarks about rational causal beliefs are in order. According to the epistemic view of causality, a causal belief is a type of belief—a directed belief, one that can be represented by an arrow in a causal graph. A causal belief is *not* a belief *about* causality, which is a belief about the contents of the ultimate causal graph. (Consider the analogous probabilistic case: under the Bayesian interpretation of probability, a probabilistic belief is a type of belief—a degree of belief, one that can be represented by a real number—*not* a belief *about* probability.) I suggested earlier that an agent's causal beliefs are rational if they are determined appropriately by her background knowledge. Moreover, these beliefs are classed as rational because they yield reliable predictions, diagnoses, strategic decisions etc. For there to be a fact of the matter about which causal beliefs are rational, these beliefs do not have to be held by any particular agent—indeed there do not have to be any agents at all—they just need to be construed as possibilities. A causal belief graph is just a set of possible causal beliefs. Given some possible domain and some possible background knowledge base, a possible causal belief is rational if it occurs in the causal belief graph determined by the domain and background knowledge. Apparently this story does not presuppose causality in its telling; it is not circular (Williamson, 2006, §7). If that is correct then parsimony is achieved.

We see then that the epistemic causality copes well with the desiderata. Currently it looks as if the epistemic theory yields an objective notion of cause. It shows us when we should expect current calculi to work, and explains why they work. Its epistemology is tied closely to its concept of cause. It imposes no limits on the variety of causal claims that fall under its scope. Nor need it add any entities to an ontology. Arguably these are all the features we require of a philosophical theory of causality.

[12] Armstrong (2004, §5.2.2) for example, puts forward an error theory for talk of absences causing. It seems unattractive to have two stories: one for present causes and one for absent causes. If our use treats them the same, it is arguably better that our theory treat them the same too.

Conclusion

As Hume noted, the idea of causal production is not something for which we have direct evidence. Cartwright claims that we have indirect evidence for its existence and infers its reality. I have tried to show that there is an alternative strategy: taking causal connection seriously as a transition of the imagination. Indeed I have argued that a fallacious projection of causality onto reality is partly responsible for the trouble that the desiderata of §2 pose for the dispositional theory. In contrast the epistemic theory—which characterises causality in terms of a rational epistemic state instead of taking it as an ontological extra—shows considerable promise when judged according to the desiderata. These desiderata, I have argued, provide an apt yardstick by which to stand any philosophical theory of causality.

I have put questions about the interpretation of probability to one side in this paper; these are discussed in detail in Williamson (2005). The case of probability mirrors that of causality: objective Bayesianism is to the propensity interpretation of probability as the epistemic theory is to the dispositional theory of causality.

Acknowledgments I am very grateful to Laurence Goldstein, Federica Russo and two anonymous referees for comments on earlier drafts of this paper.

References

Armstrong, D. (2004). *Truth and truthmakers*. Cambridge University Press: Cambridge.
Bird, A. (2005). Laws and essences. *Ratio*, *18*, 437–461.
Cartwright, N. (1983). *How the laws of physics lie*. Clarendon Press: Oxford.
Cartwright, N. (1989). *Nature's capacities and their measurement*. Clarendon Press: Oxford.
Cartwright, N. (1999). *The dappled world: A study of the boundaries of science*. Cambridge University Press: Cambridge.
Cartwright, N. (2001). What is wrong with Bayes nets? *The Monist*, 84(2):242–264.
Cartwright, N. (2003). What makes a capacity a disposition? Causality: Metaphysics and methods discussion paper 10/03, London School of Economics Centre for Philosophy of Natural and Social Science.
Dowe, P. (2000). *Physical causation*. Cambridge University Press: Cambridge.
Gillispie, S. B., & Perlman, M. D. (2002). The size distribution for Markov equivalence classes of acyclic digraph models. *Artificial Intelligence*, *141*, 137–155.
Glymour, B. (2003). On the metaphysics of probabilistic causation: Lessons from social epidemiology. *Philosophy of Science*, *70*, 1413–1423.
Hume, D. (1748). Enquiry into the human understanding. In *Enquiries concerning human understanding and concerning the principles of morals*, third (1975) edition. Clarendon Press: Oxford.
Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge University Press: Cambridge.
Kant, I. (1781). *Critique of pure reason*. Macmillan (1929), second (1787) edition. Trans. Norman Kemp Smith.
Lewis, D. K. (1973). Causation. In *Philosophical papers* vol 2, (pp. 159– 213). Oxford University Press (1986), Oxford.
Menzies, P., & Price, H. (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science*, *44*, 187–203.
Mill, J.S. (1843). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*, seventh (1868) edition. Longmans Green Reader and Dyer: London.
Papineau, D. (1992). Can we reduce causal direction to probabilities? *Philosophy of Science Association*, *1992*(2), 238–252.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press: Cambridge.
Price, H. (2005). Causal perspectivalism. In: Price, H. & Corry, R., (Eds.), *Causation, physics and the constitution of reality: Russell's republic revisited*. Oxford University Press: Oxford.
Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society*, *13*, 1–26.
Salmon, W. C. (1998). *Causality and explanation*. Oxford University Press: Oxford.
Schaffer, J. (2005). Contrastive causation. *The Philosophical Review*, *114*(3).
Shoemaker, S. (1998). Causal and metaphysical necessity. *Pacific Philosophical Quarterly*, *79*, 59–77.
Spirtes, P., Glymour, C., & Scheines, R. (1993). *Causation, prediction, and search*, second (2000) edition. MIT Press: Cambridge, MA.
Spohn, W. (2002). Bayesian nets are all there is to causal dependence. In: Galavotti, M. C., Suppes, P., & Costantini, D., (Eds.), *Stochastic causality*. University of Chicago Press: Chicago, IL.
Suppes, P. (1970). *A probabilistic theory of causality*. North-Holland: Amsterdam.
Williamson, J. (2005). *Bayesian nets and causality: Philosophical and computational foundations*. Oxford University Press: Oxford.
Williamson, J. (2006). Causality. In D. Gabbay & F. Guenthner (Eds.), *Handbook of philosophical logic*, vol 14, Springer.
Williamson, J., & Gabbay, D. (2005). Recursive causality in Bayesian networks and self-fibring networks. In: D. Gillies (Ed.), *Laws and models in the sciences* (pp. 173–221). King's College Publications: London. With comments pp. 223–245.