Woodward, James, *Making Things Happen: A Theory of Causal Explanation*, New York: Oxford University Press, 2003, pp. viii + 410, US$65.00 (cloth).

The concept of causation plays a central role in many philosophical theories, and yet no account of causation has gained widespread acceptance among those who have investigated its foundations. Theories based on laws, counterfactuals, physical processes, and probabilistic dependence and independence relations (the list is by no means exhaustive) have all received detailed treatment in recent years—and, while no account has been entirely successful, it is generally agreed that the concept has been greatly clarified by the attempts. In this magnificent book, Woodward aims to give a unified account of causation and causal explanation in terms of the notion of a manipulation (or intervention, terms which can be read interchangeably). Not only does he produce in my view the most illuminating and comprehensive account of causation on offer, his theory also opens a great many avenues for future work in the area, and has ramifications for many other areas of philosophy. *Making Things Happen* ought to be of interest not only to philosophers of causation and philosophers of science, but to any philosopher whose concerns involve assumptions about the nature of causation, laws, or explanation.

The pre-theoretical notion of a manipulation is of a causal influence produced by an agent. Correspondingly, there are two traditional lines of objection against theories of causation formulated in manipulationist terms. Firstly, the concept appears anthropocentric (or at least agent-centric), threatening to introduce an unacceptable subjectivism (or at least agent-dependence) into what is supposed to be the paradigmatic objective relation. Secondly, since manipulation itself appears to be a causal concept, there is a worry of circularity. Woodward takes both lines of objection to count against earlier agency and manipulationist theories, and works hard to dissociate his theory from these. The difference is that Woodward does not offer a purported reduction of causation, but rather the explication of causal claims in terms of a notion of intervention that is itself defined as a particular kind of causal relationship. The circle is virtuous, since the theory shows how a great number of diverse causal concepts can be defined in terms of this particular causal concept. This strategy also avoids anthropocentrism, though by a kind of fiat—since the theory is non-reductive, a fortiori it does not reduce to anything anthropocentric. In taking this result to count against anthropocentrism Woodward simply rests on our pre-theoretical confidence in the objectivity of causation. In the final part of this review I will argue that his account itself gives reason to reassess this confidence; but to begin, I will give an overview of the theory to show some of its virtues, consequences and open questions.

The non-reductive approach to causation advocated by Woodward resembles and is influenced by the formal causal modelling frameworks pioneered by Judea Pearl and by the trio of Clark Glymour, Peter Spirtes, and Richard Scheines. (Woodward gives the best introductory treatment of this work I have seen). But where those theories take as a primitive a notion of causal mechanism, and define interventions in terms of these, Woodward takes the notion of an intervention as a primitive, and defines causal mechanisms (and other causal concepts) in these terms. This is a significant achievement, in two respects. Firstly, it sets the formal frameworks on stronger philosophical footing—for one thing, Woodward provides an account of the meaning of causal claims as embodied in these frameworks, the lack of which has been the focus of recurring

criticism; for another, starting with interventions rather than mechanisms fits far more easily with the epistemological and methodological dimensions of causal explanation. Secondly, this inversion has the advantage that the path to a potential reductionist account of causation, in terms of agency, is made clear. Indeed, one of the great benefits of the book is that it brings together two traditions that have hitherto proceeded largely independently of one another—on the one hand a tradition originating in econometrics and experimental design, and continued in contemporary work on causal inference in computer science, which takes causal claims to encode claims about the results of hypothetical experiments; and on the other hand a philosophical tradition that attempts to analyse causation in terms of agency.

Starting with interventions leaves the question of their relationship to causal generalizations and laws. In Woodward's account it is the notion of invariance under interventions that plays the role laws of nature do in other theories—to distinguish between causal and merely accidental generalizations. Invariance under interventions holds when a particular generalization correctly captures the counterfactual relationships between two variables under a particular range of interventions. As Woodward notes, 'whether or not a generalization is invariant is surprisingly independent of whether it satisfies many of the traditional criteria for lawfulness, such as exceptionlessness, breadth of scope, and degree of theoretical integration' [17]. Independent, that is, because *weaker*—it may capture a generalization holding in quite particular circumstances, for quite particular interventions. This might appear *too* weak—a generalization only capturing the relationship between one or two possible interventions for some particular situation hardly merits the title—but the benefit is that we have a continuum, from minimal sorts of invariance all the way through to the ideally exceptionless invariance (invariance under all possible interventions) of the laws of physics. Indeed, Woodward considers laws 'as just one kind of invariant generalisation' [17]. Obviously, some explanations are more informative than others, and Woodward [18] proposes that explanatory depth, in the same way, can be analysed in terms of the degree of invariance that the explanations support. This is a lovely and intuitive way to characterize the difference between laws of nature and the laws of the special sciences, through to the sorts of everyday causal knowledge embedded in folk psychology. And it has the desirable consequence that we can see scientific knowledge as an elaboration and refinement of everyday causal thinking rather than taking the latter to involve implicit knowledge of the former [20], a point on which Woodward's account is clearly superior to rival models of explanation.

Nevertheless, there is a range of open questions concerning how the idealized definitions of causal concepts provided by Woodward can be mapped onto real world practices, the resolution of which is especially important given Woodward's insistence early in the book that a theory of causation needs to involve an epistemological aspect that makes causal knowledge accessible to ordinary agents. At first glance, definitions such as his **DC** (Direct Cause [55]) do not satisfy this desideratum—in order to make a true direct-causal claim, for example, we would need to have the ability to perform (or otherwise counterfactually ascertain the result of) an intervention on a system with all other variables also held fixed by intervention. Needless to say, this is not something we ordinarily do, or are even in all cases capable of doing. Similar questions arise for other definitions Woodward provides, leaving a rich area of investigation for cognitive scientists and like-minded philosophers to explore. The benefit of the formal apparatus employed is that it

makes these questions precise. Indeed, the precision Woodward's approach gives to questions of conceptual clarification is remarkable. A quite minimal apparatus is required to both elegantly describe and then diagnose our judgments concerning classic cases such as cancellation of total causal impact along multiple causal routes, failures of causal transitivity (dog bite [57–9], falling boulder [79–81]), and various purported counterexamples to counterfactual analyses of causation (chestnut smasher, [67ff], thirsty traveller, [77ff], trumping [81–2]). Many controversial cases in the literature have continued to be debated even when all sides agree about the relevant counterfactual dependencies, and the apparatus Woodward uses both explains the various intuitions involved, and how they can be reconciled within a manipulationist framework.

Moreover, the framework involves no metaphysical claims whatsoever, simply employing a distinction between individuals, types, and variables used to represent those individuals and types. The causal relata on Woodward's view [111–14] are simply any particulars that can be manipulated—whether these be facts, events, tropes, or any other metaphysical candidate you wish to plug in (manipulation implies that these particulars be capable of taking different values; thus Woodward suggests that *variables* are the best way to characterize the causal relata). The lack of metaphysical claims masks, however, the degree to which the framework might help metaphysical debates in other areas. Take mental causation. Central to contemporary debates in this area has been the exclusion problem, where the possibility of alternative explanations for behaviour in terms of physics and in terms of belief-desire psychology are supposed to generate metaphysical worries about the efficacy of mental states qua mental. According to Woodward's account of causation, these explanations simply don't compete—each is framed in terms of a different variable set and is a bona fide causal explanation just in case the relevant counterfactuals are true. There is no *causal* sense in which physical explanations exclude or otherwise diminish mental explanations, though there might be further interesting questions concerning the relations between the two. In fact, Woodward [147] says that his account does not imply that all causal explanations are backed by exceptionless laws; and that if this is in fact the case, it will be an interesting *empirical* discovery that has no bearing on the truth of higher-level causal claims. So if we accept Woodward's account of causation, causation will not be a metaphysical concept driving reductionist arguments in philosophy of mind—there being no sense in which physical explanations are more causal than any other form of explanation. Reductionist arguments will have to find some other way of privileging the physical.

This line of argument is available because on Woodward's account, causal claims are relative to the specification of variables, both in the sense of which variables are included in the set [55–6] and of how fine or coarse grained the specifications are [378–9 n. 20]. This obviously leaves open a range of further questions, analogous to familiar issues in philosophy of science: How do we select a variable specification (model)? Can this be done independently of causal claims? Are some models better than others, and if so, on what grounds? Can models be compared independently of causal claims? Every model will make claims that can be objectively tested by performing the relevant hypothetical experiments they embody (or in cases where manipulation is practically impossible, by otherwise evaluating the relevant counterfactuals)—the questions here, however, arise in comparing models each of which is empirically adequate. While these are all pressing questions, it is nice to have

them disentangled from questions about causation per se, which on Woodward's account can play no role in their resolution.

In closing, I will make some brief comments on Woodward's claim that his theory avoids the agent-dependency of earlier manipulationist theories. Woodward [85–91] makes a distinction between what he takes to be the agent-independent patterns of counterfactual dependence that constitute the 'objective core' of causal claims, and the agent-dependent pragmatic features of those claims which he takes to consist in the selection of those counterfactuals that represent 'serious possibilities'. Nevertheless, we do make causal claims in the absence of the practical or physical possibility of performing the associated manipulations, a point which has been another common line of criticism of agency theories of causation. The solution Woodward adopts to this problem [131–2] is to weaken the strength of possibility required for interventions, so that it is only required that they be logically or conceptually possible, and that we have some means of evaluating the relevant counterfactuals (Woodward is critical [118ff] of projectivist views of causation, so it is interesting to note that this itself amounts to a form of projectivism). The cleanest form of logically possible intervention is simply to have the state of the world at the time of the change miraculously become such that (only) the change has occurred. (Such a change trivially meets Woodward's requirements for an intervention; formally, we can suppose the required intervention variable be *God's choosing to make it so*.) And the simplest means of evaluating the relevant counterfactuals is to use laws of nature— after all, these are the fundamental invariances. But these constraints are too weak for the purposes of recovering causal claims, since they will license counterfactuals in both the past and future directions in time. (This sort of claim is often put in terms of the time-symmetrical nature of the laws of physics, but here requires simply that the laws serve equally well for purposes of retrodiction and prediction).

There are several different strategies that might be used to recover the temporal asymmetry of causation here, but I think that it suggests that agent-dependency is not so easily evaded. We can put the question to Woodward in the same form as he puts a very similar question to his rival accounts. Woodward asks, of those who propose that causation is a disunified cluster concept, why we shouldn't abandon our concept of causation in favour of some revised version, causation* [93], and with respect to Lewis's similarity metric, why we shouldn't exchange it for another metric and a corresponding notion of smausation [137]. So the question for Woodward is—why this pattern of counterfactual dependence and not another? Or put slightly differently, why one sort of counterfactual antecedent and not another? Why is it that one sort of counterfactual is the sort that we can use for the purposes of manipulation, and not other sorts? Again, given that any variety of counterfactual meeting the criteria of an intervention will give us a variety of manipulation, why is it only some subset of these that we are interested in? Why shouldn't we abandon counterfactual for counterfactual*, especially if counterfactual* will enable us to cause* past events? The answer, I think, is that we can't, in fact, bring about counterfactual* antecedents (at least in all cases we know of)—but this is in part a fact about the sorts of agents we are.

Early in the book [28], Woodward suggests that the demand for a reductionist account of causation 'virtually forces one' to an anthropocentric conception of causation. And the train of thought underlying much of the resistance to such a reduction seems to be that anthropocentrism is equivalent to subjectivism, and the insistence that whatever causation is, it can't be subjective. The mistake here is in

supposing agent-dependence to be equivalent to subjectivism—the fact that we can't bring about counterfactual* antecedents might be agent-dependent, but it is certainly not subjective. Here as elsewhere, *Making Things Happen* helps to focus the issues in a way that allows theoretical progress; it deserves to form an axis around which future debates in causation and explanation revolve.

Brad Weslake
University of Sydney

Dennett, Daniel, *Sweet Dreams: Philosophical Obstacles to a Science of Consciousness*, Cambridge MA: MIT Press, 2005, pp. xiii + 199, US$28 (cloth).

In *Sweet Dreams* Dennett presents himself as a Lockean under-labourer, 'removing some of the rubbish that lies in the way to knowledge'. The rubbish in question is identified in the book's subtitle: 'philosophical obstacles to a science of consciousness'. Dennett's central claim is that the intuitions and thought experiments that dominate philosophical discussions of consciousness are hampering the scientific study of consciousness. Removing these obstacles will reveal that consciousness poses no hard problems and raises no explanatory gaps; indeed, it will reveal that a mechanistic explanation of consciousness is not just possible but is 'fast becoming actual' [7].

Although written with Dennett's characteristic panache, *Sweet Dreams* is less than kind on its reader. Dennett's Jean Nicod lectures (chapters two through five) are sandwiched between various papers on consciousness, but rather than being presented as stand alone papers they are arranged as chapters. There is too much repetition between them for this arrangement to be successful. Numerous claims and indeed entire paragraphs reappear verbatim. Even in these environmentally conscious times I wouldn't have thought that rubbish-removal entailed quite so much recycling.

Also likely to try the patience of some readers is Dennett's characterization of his opponents. Zombiephiles rely on an intuition that is 'almost entirely arational, insensitive to argument or the lack thereof' [22], while scientists who suspect that consciousness might not succumb to current scientific methods have been 'tempted' or 'blackmailed' into holding these views [134]. In light of these comments one might have expected *Sweet Dreams* to be packed with arguments. Not so. There are arguments here and there, but it is not always easy to find them among the metaphors (consciousness as fame; consciousness as cerebral celebrity; consciousness as fantasy echo) and stories (the Tuned Deck, Mr. Clapgras, and the very entertaining Indian Rope Trick). In fact, Dennett is rather pessimistic about the ability of reason to resolve the qualia wars:

> the tempting idea that there is a Hard Problem is simply a mistake. I cannot prove this, and some who love the Hard Problem find my claim so incredible that they admit, with some hilarity, that they can't take it seriously. So I won't make the tactical error of trying to dislodge with rational argument a conviction that is beyond reason.
>
> [72]

It is not hard to have some sympathy with this pessimism: it *is* difficult finding argumentative traction on the issues that divide qualiaphiles from qualiaphobes. Is this, as Dennett's implies, because the qualiaphiles are allied with the forces of