Introduction

Understanding Counterfactuals and Causation

Christoph Hoerl, Teresa McCormack, and Sarah R. Beck

It seems obvious that there is a close connection between our understanding of certain causal claims and our understanding of claims such as the following: 'If a piece of metal had not burst its tyre, Concorde would not have crashed', 'If less violence was shown on television, the amount of violent crime would be lower', or 'If I were to prune this plant in the next few weeks, it would flower next year'. These latter claims exemplify a type often referred to as a *counterfactual conditional*, or *counterfactual*, for short.[1]

To date, the most prominent way in which the idea of a connection between causal and counterfactual claims has figured in philosophy has involved the idea that the meaning of the former can be analysed, at least in part, in terms of the latter. David Lewis (1973a) has put forward what is probably still the best-known example of a theory following this type of approach—i.e. what is often called a *counterfactual theory of causation*. Viewed more generally, counterfactual theories of causation form a category that also encompasses a number of other approaches that have emerged or come to more prominence since, most notably interventionist theories of causation such as the one put forward by James Woodward (2003; see below for discussion).

It is arguable, however, that at least some of the reasons as to why the general idea of connections between causal and counterfactual claims strikes us as plausible have to do with intuitions that are, at least in principle, quite separable from the issues at stake in counterfactual theories of causation. Counterfactual theories of causation (at least as typically conceived) trade on the idea of connections between the two

[1] The three quoted statements are also all conditionals in the subjunctive mood. As we will see below, there is some controversy over how exactly to construe the relationship between the notion of a subjunctive conditional and that of a counterfactual. For instance, on a narrow understanding of the notion of a counterfactual, as advocated by some theorists, the third type of statement we have quoted (a subjunctive conditional about the future) does not display all the features that should be taken to be characteristic of a genuine counterfactual, because it is not clear that its meaning differs from that of the indicative conditional 'If I prune this plant, it will flower'. Other theorists advocate a much broader notion of a counterfactual, on which even statements not in the subjunctive mood, such as 'If I prune this plant, it will flower' can count as counterfactuals.

types of claim on the level of truth conditions. Yet, we arguably also have intuitions about connections between causal and counterfactual claims in quite a different sense—namely empirically informed intuitions about connections between two types of *thinking* we actually engage in: thinking about causal relations and thinking using counterfactuals. For instance, as Woodward points out in his contribution to this volume, it simply seems to be a datum that people find it helpful, in considering complex causal scenarios, to engage in certain sorts of counterfactual thinking. This is not just an interesting fact about our mental lives, but is also reflected in practices that are part of British and American common law, or that inform engineering decisions at NASA. Similarly, it also seems to be a datum that people will spontaneously generate counterfactual thoughts in response to certain kinds of causal outcomes, especially if they were unexpected and distressing. Again, recognition of this fact goes beyond the anecdotal, and informs, for instance, aspects of post-traumatic stress counselling.

The central idea behind the current volume is that the psychological literature on counterfactual thought and its relation to causal thought provides a large, but as yet largely untapped, potential for exploring philosophical questions regarding the nature of causal reasoning in a way that may ultimately also impact on some of the issues at stake in theories of the type exemplified by counterfactual theories of causation. Conversely, philosophical reflection specifically on the nature of causal reasoning and its relation to reasoning with counterfactuals may help shed new light on some of the theoretical issues at stake in psychological studies that aim to probe, e.g. into the development of these reasoning abilities or the psychological capacities that underpin them. Thus, the chapters in this volume take as their starting point the types of intuitions about connections between causal and counterfactual thinking just mentioned, try to sharpen them up and enrich them through empirical means, and offer theoretical accounts as to how these intuitions are best explained.

This introduction cannot address the full range of perspectives from which these issues are explored in the various chapters in this volume. Rather, our aim in what follows is to draw out a small number of key lines of thought or debates that cut across several chapters, and across the divide between philosophy and psychology.

Counterfactual Process Views of Causal Thinking

How might counterfactual thought and causal thought be related? Perhaps the most ambitious general type of line one might take in this area is to try to argue for what Teresa McCormack, Caren Frosch, and Patrick Burns, in their chapter for this volume (this volume, p. 54), call a *counterfactual process view* of causal reasoning. According to such a view, engaging in counterfactual thought is an essential part of the processing involved in making causal judgements, at least in a central range of cases that are critical to a subject's understanding of what it is for one thing to cause another.

One fruitful way of approaching the different contributions to this volume is to think of them as providing materials, conceptual as well as empirical, for challenging

counterfactual process views of causal thinking, or for responding to such challenges. Some apparent challenges to a counterfactual process view of causal thinking emerge as soon as we look at some of the empirical work on causal and counterfactual thought reported in some of the empirical papers in this volume. Here is a small sample.

- When given a vignette detailing a sequence of events with a negative outcome, and then asked to generate suitable counterfactual statements, the counterfactuals that adults generate focus on antecedents that are different from the ones which they would normally judge to be the *causes* of the outcome. Rather, those counterfactuals seem to be focused on antecedents that would have been sufficient to *prevent* the outcome from occurring (Mandel, this volume).
- Given certain temporal cues, children, like adults, reliably interpret a particular physical arrangement as exemplifying a common-cause structure rather than a causal-chain structure, or vice versa. However, when asked counterfactual questions about potential interventions in the system, 5- to 7-year-olds, unlike adults, do not reliably provide answers that are consistent with their choice of causal structure (McCormack, Frosch, and Burns, this volume).
- When adults are asked to rate the probability of a conditional such as 'If car ownership increases, traffic congestion will get worse', which has a natural causal interpretation, there is little evidence that their answers draw on beliefs based on 'undoing' the antecedent. Beliefs based on undoing the antecedent only appear to come into play when adults are asked, e.g. to judge the *causal strength* of the relation between car ownership and traffic congestion (Feeney and Handley, this volume).

What, if any, implications such findings have for the prospects of a counterfactual process view of causal thought depends on a number of questions, such as the following: Is a counterfactual process view of causal thought committed to the idea that people can in fact explicitly articulate the relevant counterfactuals that underlie their causal judgements, or can we make sense of the idea of merely implicit counterfactual reasoning? To what extent is the truth of a counterfactual process view of causal thought compatible with the idea that people's explicit counterfactual judgements diverge, in certain respects, from their causal judgements? To what extent, if any, does a counterfactual process view of causal thought hinge on a notion of counterfactual reasoning according to which such reasoning necessarily involves some form of mental 'undoing'?

The chapters in this volume offer a variety of different views on these questions, some of which we will touch upon below. However, the above list of empirical observations that prompted these questions, as well as the many further findings reported elsewhere in this volume, also invite a more general comment. Anybody who is primarily familiar with the discussion about counterfactual theories of causation in the philosophical literature and then starts to engage with psychological research on

counterfactual reasoning and its connection with causal reasoning is likely to be struck by the sheer diversity of phenomena that are being studied as part of the latter. In particular, it seems clear from this diversity that there might be a real danger of setting things up in the wrong way from the start if we ask what *the* relationship is between causal and counterfactual understanding, as one might be tempted to if influenced by the discussion about counterfactual theories of causation in philosophy. Rather, there might be a multitude of ways in which different kinds or aspects of counterfactual understanding may interact with aspects of our understanding of causal relationships. We will discuss one important way in which this general consideration might be thought to be relevant to some of the chapters in this volume at the end of this introduction.

Philosophical Challenges

Perhaps the simplest version of a counterfactual process view of causal thought that one might think of would be a straightforward psychological counterpart of something like Lewis' (1973a) version of a counterfactual theory of causation. As already mentioned, counterfactual theories of causation are typically intended to give the truth conditions of causal judgements. The underlying motivation here is the thought that we can capture what it means to say that A causes B by stating that, for the judgement 'A causes B' to be true, a certain kind of counterfactual relationship has to obtain between A and B. It is important to note that, even if this thought is along the right lines, counterfactual theories of causation need not be seen to be descriptive of the psychological processes that people go through when making causal judgements. Yet, it is also easy to see how one might try and make a connection between these two issues: If the truth conditions of causal statements are to be given, at least in part, in terms of counterfactuals, it seems plausible to assume that people's reasoning about causal relationships should be sensitive to the obtaining of the relevant counterfactuals. And one very straightforward way in which one might then account for this sensitivity is by assuming that people actually engage in reasoning with counterfactuals when making causal judgements, i.e. by adopting a counterfactual process view of causal thought.

At its most basic, the kind of counterfactual process view we are envisaging here would have it that we arrive at causal judgements of the kind 'A causes B' by evaluating a counterfactual such as 'If A did not occur, B would not occur'. This kind of view is not actually advocated in any of the chapters in this volume, at least at this level of generality and without further qualification. Nevertheless, it serves as a useful model for considering some of the general types of challenges that counterfactual process views of causal thought face.

One class of challenges one might think of here is discussed in detail in Dorothy Edgington's chapter. Her strategy is to look at some philosophical problems that Lewis' counterfactual theory of causation faces, which also threaten to infect a psychological counterpart of it (of the kind sketched above). One of the key issues she raises is that the

way in which a counterfactual is to be interpreted can often itself depend on what we take the causal facts to be. An example she uses is that of a person tossing a coin and another person saying, 'If I had bet on heads, I would have won'. If the coin has in fact landed heads, the counterfactual is unproblematically true on the assumption that the second person's betting or not betting on heads had no causal impact on the outcome. However, it is not obviously true if her bet might have caused the outcome to be different—imagine that the person who tossed the coin is a swindler who can somehow influence which way it lands.

This poses a threat for a theory like Lewis', which tries to provide the truth conditions of causal claims in terms of counterfactuals. The threat is that the theory will be viciously circular, if the truth of the relevant counterfactuals, in turn, depends on that of certain causal claims. Arguably, this threat of circularity does not just affect Lewis' version of a counterfactual theory of causation, but also has an impact on the prospects of a psychological counterpart to Lewis' theory of the type we have been envisaging. More specifically, if Edgington is right, the problem she identifies with Lewis' account undermines the idea that we can give a *reductive* account of the meaning of causal claims in terms of counterfactual ones. And, as such, it also provides an argument against any counterfactual process account of causal thought with similarly reductive ambitions, i.e. any account that tries to base our understanding of causal claims on prior and independent counterfactual reasoning abilities.

Note, however, that there are versions of counterfactual approaches to causation that are not obviously affected by Edgington's argument. These are approaches that admit that there may be no possibility of giving a reductive account of causation in counterfactual terms, but which nevertheless maintain that an illuminating account of the meaning of causal statements can be given that makes essential reference to the holding of certain counterfactuals. Woodward's (2003) variant of an interventionist approach to causation, for instance, tries to account for the meaning of a statement of the type 'A causes B' in terms of the idea of an invariant relationship between A and B that holds under a range of interventions. This is a counterfactual account, in so far as it interprets 'A causes B' in terms of certain counterfactuals about the consequences of A being intervened on. However, the notion of an intervention is itself a causal notion (cf. Woodward, 2003: ch. 3). First of all, to say that A is being intervened on simply is to say that A is being caused to be a certain way (or caused to occur or not to occur). Moreover, for something to count as an intervention in a given causal system, in the sense relevant to interventionism, it must also meet a set of criteria regarding its own causal *in*dependence from other elements of the system at issue. For instance, we might observe an invariant relationship between A and B, even in the absence of A causing B, if A has a cause that also itself causes B, independently of causing A. In that case, bringing A about by means of this cause doesn't qualify as an instance of intervening on A in the sense at stake in interventionist approaches to causation. (The issue here is basically the one that is also behind the problem of confounding variables in empirical experiments.)

Even though the notion of an intervention is thus itself a causal notion, ruling out a reductive account of the meaning of causal claims in terms of interventionist counterfactuals, interventionists such as Woodward argue that this does not make interventionism viciously circular (see also Woodward, this volume, p. 34). Note, in particular, that we can specify the causal criteria an event must meet in order to count as an intervention that may settle whether 'A causes B' is true without touching on the particular causal relation, if any, that obtains between A and B itself. Thus, the most obviously damaging kind of circularity is avoided.

Suppose, then, that interventionist versions of a counterfactual theory of causation, and, by extension, their psychological counterparts, can avoid the kind of threat of vicious circularity that Edgington identifies in Lewis' theory. The points Edgington makes may still bear on the idea of a counterfactual process account of causal thought in a more subtle way. For they might be seen to put some pressure on the defender of such a theory to make more precise exactly how we should think of the role that counterfactual reasoning has in causal thought. Johannes Roessler, for instance, in his contribution to this volume, contrasts two quite different ways in which one might link up causal reasoning abilities with counterfactual reasoning abilities. According to one version, some counterfactual reasoning ability is required for causal thought, because it is required to grasp some of the essential *commitments* of causal claims. According to another version, causal reasoning is also required to marshal the canonical *evidence* for such claims.

If a reductive counterfactual theory of causation such as Lewis' was correct, this might perhaps also help make plausible the latter version of a counterfactual process account of causal thought. That is to say, if the meaning of causal statements could be reductively analysed in terms of counterfactuals, establishing whether the relevant counterfactuals obtain would arguably constitute the canonical way of finding out about the truth of causal claims.[2] Once we give up the idea of a reductive relationship between causality and counterfactuals, by contrast, this version of a counterfactual process account of causal thought also becomes harder to sustain. The prospects of the alternative version, according to which causal thought involves an ability for counterfactual reasoning because counterfactual reasoning is required to grasp some of the commitments of causal claims, are discussed in detail in Roessler's chapter. In particular, he discusses the extent to which it might be compatible with what he calls 'naïve realism concerning mechanical transactions', which involves the idea that perception can provide us with non-inferential knowledge of mechanical transactions (as opposed, e.g. to mere patterns of movement).

[2] Admittedly, there is scope for further debate on this matter. See, e.g. Woodward, this volume, p. 36, on related matters.

The Developmental Challenge

We have looked at a challenge to counterfactual process views of causal thought that is informed by philosophical considerations, in particular about the meaning of causal and counterfactual claims. But there are also a number of empirical challenges that counterfactual process views of causal thought face. Some of the contributions to this volume by developmental psychologists set out a very basic such challenge: In both verbal and non-verbal tasks, children seem to demonstrate an understanding of causal relations long before they appear to be fully competent with counterfactual reasoning (at least of certain kinds). Thus, it appears that a counterfactual reasoning ability cannot be an essential ingredient in the ability to make causal judgements, if we think of the latter as what is demonstrated in the relevant verbal and non-verbal tasks at issue.

In a very influential 1996 paper, Harris, German, and Mills claimed to have demonstrated that 'young children, including 3-year-olds, can consider counterfactual scenarios in trying to figure out both what has caused a particular outcome and how it might have been prevented' (Harris et al. 1996: 249). Harris et al. explicitly framed their paper in terms of a defence of what we have called a counterfactual process view of causal thought, and took themselves to have found evidence supporting such a view in the way in which children answered counterfactual questions regarding a number of different causal scenarios presented to them in stories. By contrast, the papers in the current volume by McCormack et al., Beck et al., and Perner and Rafetseder all come to a different conclusion. What emerges from them is a picture of counterfactual thought as a very sophisticated cognitive achievement, some elements of which do not in fact develop fully until the age of 10 or 12 years. There are no claims for similarly late developments in causal understanding in the developmental literature.

How is this discrepancy in views to be explained? Those developmentalists who stress the cognitive complexity of counterfactual thought can acknowledge that the children in Harris et al.'s experiments gave correct answers to questions put to them in the form of a subjunctive conditional. However, they are likely to maintain that the children did so on the basis of resources that fall short of genuine counterfactual reasoning, narrowly understood. Thus, for instance, a suggestion that can be found in the chapters by both Perner and Rafetseder and Beck et al., respectively, is that younger children, when asked a question using the subjunctive conditional form 'What would have happened if x had not happened?', actually merely entertain the indicative conditional 'If x doesn't happen, y happens', and answer on that basis. In many cases, at least in the typically rather simple worlds of developmental experiments, y will also in fact be the right answer to the counterfactual question, so the performance of children in counterfactual tasks may mask the fact that they do not genuinely engage in reasoning with or about counterfactuals.

One key underlying thought here is that counterfactual thought is psychologically demanding in as far as it requires, for instance, holding in mind both what could have happened and what actually happened (an idea also explored, within the context of

adult cognition, in Ruth Byrne's chapter; though see Woodward for a critical perspective). This has to be distinguished, the thought goes, from a more primitive ability to imagine what is in fact a non-actual state of affairs, but in a way that falls short of genuine counterfactual thinking. Thus, when asked the counterfactual 'What would have happened if x had not happened?' children might simply draw on their general knowledge to construct in their imagination a situation in which x does not happen, and then answer accordingly. However, they might not link this to their knowledge of what actually happened, as is required for genuine counterfactual reasoning, at least on the view at issue here.[3]

A theoretical perspective on the development of counterfactual and causal thought that differs somewhat from the line of thought just sketched is provided in David Sobel's chapter for this volume. Recall that the line of thought presented above had it that the lack of an ability to engage in genuine counterfactual reasoning might be masked in situations in which children can draw on general background knowledge in answering a question that is put to them in the form of a counterfactual. The idea here is that of a domain-general ability which young children lack (i.e. the general ability to engage in genuine counterfactual reasoning), but the lack of which can be masked in certain circumstances. Sobel, by contrast, can be seen to be pressing the point that, conversely, an existing general ability to engage in counterfactual reasoning might sometimes be masked by the fact that children do not have sufficient knowledge within a domain that they could bring to bear in evaluating counterfactuals about that domain (see also Woodward, this volume). In one of the studies reported by Sobel, for instance, 3- and 4-year-olds were asked counterfactual questions after listening to two stories that were arguably structurally identical. When told a story in which an event fulfils a character's desire, leaving him happy, the children could reliably judge how the character would feel had the desire been left unfulfilled. Yet, when presented with a story in which a character doesn't know that a certain event will happen, and is surprised when it does, the same children could not reliably judge how the character would feel if he had known about the event. As Sobel argues, the most natural interpretation of this finding is that it is to be explained in terms of differences in children's domain-specific knowledge: by the age of 3 or 4, they have already grasped certain facts about the functional role of desires, but still lack a proper understanding of the functional role of knowledge.

[3] There is an influential idea in the literature on children's developing understanding of the notion of a representation that can be seen to provide a historic model for this position. It is a well-established finding that children can engage in pretend play (e.g. acting as if a banana was a telephone) before they can pass false belief tests (i.e. correctly predict the actions of a person who lacks key pieces of information). Perner (1991) explains this developmental dissociation in terms of the idea that pretence only involves the ability to switch between two representations (representing the banana as a banana, and representing it as a telephone), whereas an understanding of false belief requires modelling the other's belief, but as a belief that actually aims at the world one's own beliefs represent to be different. In other words, false belief understanding does not just require entertaining two representations, but relating them to one another.

The advocate of domain-general changes in children's counterfactual reasoning abilities can, of course, admit that the ability to engage in counterfactual reasoning can be constrained by a lack of background knowledge within a domain in the way envisaged by Sobel. What is ultimately at stake in his or her position is the question as to whether, in addition, we can also make sense of, and give empirical substance to, the idea of changes in children's very understanding of possibility. Perner and Rafetseder pursue this issue, for instance, by trying to find ways of disentangling empirically counterfactual and other types of conditional reasoning, and they do find that younger children struggle with cases in which using the latter will not yield the correct answer.

One interesting possibility, however, which emerges from both the chapter by McCormack et al. and that by Beck et al. is that the strongest empirical evidence relating to developments in children's understanding of possibility might in fact emerge from work in which the children are asked to produce counterfactuals that are *different from* the counterfactuals that philosophers putting forward a counterfactual theory of causation have traditionally focused on. (Compare, for instance, McCormack et al.'s discussion of studies on children's comprehension of counterfactuals in situations featuring cue competition, or Beck et al.'s discussion of studies on what they call 'open counterfactuals' and counterfactual emotions.) Clearly, once we look at counterfactuals that are different from the ones that, according to a counterfactual theory of causation, encapsulate the causal relations obtaining in the relevant situation, the particular methodological worry we described in connection with Sobel's contribution to this volume no longer applies. By the same token, however, it might be argued that any developmental differences in children's understanding of possibility that might be found in such studies are of less obvious relevance to the question as to whether some form of counterfactual processing view of causal thought can be sustained. It is to a version of this issue that we turn next.

Two Notions of 'Counterfactual'

The kind of developmental claim that we considered in the previous section—to the effect that genuine counterfactual thought emerges later in development than causal thought—typically hinges on a specific understanding of what a counterfactual is, which we might call a 'narrow' understanding of the notion of a counterfactual. It is important to note here that the issue as to whether young children can engage in genuine counterfactual reasoning, as e.g. Beck et al. or Perner and Rafetseder see it, isn't one about children's linguistic competence. They can allow that counterfactual reasoning abilities might be manifested in tasks that don't require children to produce or evaluate explicit statements of the form 'If x hadn't happened, y would have happened' or similar, but instead look, e.g. at the development of feelings of regret, or at children's understanding of 'almost happened' statements. Rather, central to the 'narrow' understanding of the notion of a counterfactual those authors invoke is the idea of a sharp distinction between counterfactual and indicative conditionals.

A similar narrow understanding of the notion of a counterfactual can be found in Lewis' book *Counterfactuals* (Lewis, 1973b). In Lewis, the idea that there is a sharp distinction between counterfactual and indicative conditionals comes out, for instance, when he explains why 'Subjunctive Conditionals' would not have served as an equally good title for his book. Lewis admits that counterfactuals, as he understands them, are typically expressed in the subjunctive mood, but then he goes on to say that 'there are subjunctive conditionals pertaining to the future, like "If our ground troops entered Laos next year, there would be trouble" that appear to have the truth conditions of indicative conditionals, rather than of the counterfactual conditionals I shall be considering' (Lewis, 1973b: 4).[4]

Not all philosophers working on conditionals share Lewis' views of a sharp distinction between counterfactual and indicative conditionals. Edgington, for instance, makes the point in her chapter that it appears that any acceptable indicative conditional can, as she puts it, 'go counterfactual', given the right context. Broadly speaking, the understanding of the notion of a counterfactual she employs here is that of a past-tense conditional in the subjunctive mood. (Note that this, according to the quotation above, should also count as a counterfactual by Lewis' lights.) Yet, if counterfactuals, in this sense, can be generated by transformation from indicative conditionals, it seems implausible that the original indicative conditional and the resulting counterfactual should require two completely different kinds of semantic analysis. This, in turn, has a direct impact on the prospects for a counterfactual theory of causation as there are obvious examples of indicative conditionals that don't 'track causation', as Edgington puts it, such as 'If she's not at home, she's out for a walk'. If these can be transformed into counterfactuals, as Edgington uses the term, the latter will clearly be unsuitable for figuring in an analysis of causation. Her conclusion is that 'counterfactuals are too wide a class to hope to capture causation in terms of them' (this volume, p. 239).

Woodward, too, thinks that the narrow understanding of the notion of a counterfactual that we have been considering does not single out a phenomenon with a 'fundamentally different type of semantics' (this volume, p. 26) from many other conditionals. However, he develops this idea in a somewhat different way from Edgington. Woodward in fact advocates a broad understanding of the notion of a counterfactual, according to which even a conditional such as 'If I drop this pencil, it will fall to the floor' should be counted as a counterfactual. What governs whether a conditional counts as a counterfactual or not, on this understanding, is whether evaluating it requires 'the insertion of a change into conditions as they are in the actual world, the alteration of some additional features, and the retention of others' (ibid.; see Woodward's chapter for further elaboration of this idea). As Woodward argues, this is the case for the conditional just mentioned.

[4] Lewis (1973b: 3) admits, though, that the title 'Counterfactuals', too, may be misleading, as it might be seen to have the implication, which he rejects, that he is dealing with a class of conditionals the antecedent of which must be false.

There is also a further point that Woodward makes, which is particularly pertinent to the project of the present volume. Even if a narrow understanding of the notion of a counterfactual did turn out to be useful for some purposes, he argues, it might not be the most helpful when it comes to examining potential ways in which causal and counterfactual thought might be connected. Rather, it is much more plausible that it is the broad understanding that we should be focusing on for the specific purpose of examining such connections. In particular, Woodward makes the point that a psychological account of the processes involved in causal thought is likely to assign special significance to causal thought in the context of planning and deliberation. In those contexts, however, conditionals such as 'If I drop this pencil, it will fall to the floor' seem just as central (if not more so) as conditionals such as 'If I had dropped this pencil, it would have fallen to the floor'. Thus, it is natural to think that, if a counterfactual process theory of causal thought is on the right track, the relevant notion of a counterfactual will be the broad notion that Woodward has in mind.[5]

We can look to the chapters by Aidan Feeney and Simon Handley, and by Ruth Byrne, for some empirical work that, whilst perhaps speaking against a counterfactual process view of causal thought on a 'narrow' reading of the notion of a counterfactual, seems consistent with Woodward's views. Indeed Feeney and Handley explicitly interpret one of their results as being in line with Woodward's approach. They used a 'probabilistic truth table task' to study adults' comprehension of what they call causal conditionals, i.e. conditionals most naturally construed as expressing a causal relation. In the task, participants were asked to rate the probability that a causal conditional such as 'If car ownership increases, traffic congestion will get worse' was true, and they were then also asked to rate the probability of each of a set of conjunctions: in each conjunction, the antecedent or a negation of the antecedent was combined with the consequent or a negation of the consequent. Thus, for instance, in addition to the conditional just mentioned, participants would also be asked about the probability of each of the following: 'Car ownership will increase; traffic congestion will get worse', 'Car ownership will increase; traffic congestion will not get worse', 'Car ownership will not increase; traffic congestion will get worse', and 'Car ownership will not increase; traffic congestion will not get worse'. Feeney and Handley found that, in such a task, participants' responses to the original conditional were strongly correlated only with their responses to conjunctions featuring the antecedent, but not with responses to conjunctions featuring the negation of the antecedent. This suggests that they make sense of the relationship expressed primarily by simply imagining a situation in which the antecedent is true, rather than also considering an imagined situation in which the antecedent is not true. Moreover, Feeney and Handley also found the same result when the causal conditional was in the subjunctive mood, suggesting that that there is no sharp distinction in the way the two types of conditionals are understood.

[5] By the same token, any difficulties children may have with certain counterfactuals, more narrowly understood, do not have to stand in the way of such a counterfactual process view of causal thought.

Byrne in fact uses the term 'counterfactual' in a way that is closer to what we have called the 'narrow' understanding. In particular, she takes it that understanding a counterfactual conditional, in contrast to understanding an indicative conditional, involves 'thinking of two possibilities'. This understanding of the term 'counterfactual' differs from that advocated by Woodward, who explicitly rejects a similar idea put forward by Perner and Rafetseder (Woodward, this volume, p. 21f.). Yet, Byrne also claims that when people think about what she calls 'strong causes', they in fact only envisage a single possibility. An example would be thinking about the claim 'Heating water to 100 degrees causes it to boil'. Byrne proposes that people understand this claim by thinking about the possibility that water is heated to 100 degrees and boils; they do not think about the alternative possibility, which is also consistent with the claim, that the water is not heated to 100 degrees and does not boil. Despite the terminological disagreement with Woodward over the term 'counterfactual', Byrne's view is thus actually consistent with his idea that a basic form of causal thought may be centred on the idea that causes are *sufficient* for their effects, which requires a grasp of counterfactuals only in Woodward's broad sense (Woodward, this volume, p. 42).

Causal Judgement and Causal Selection

For Byrne, the idea that understanding causal claims does not always require thinking of two possibilities (and thus a grasp of counterfactuals in the narrow sense she adopts) is connected to the idea of a distinction between 'strong causes' and 'enabling causes'. Enabling causes, she claims, require individuals to think about the same two possibilities as counterfactuals (again, in the narrow sense) do. Thus, there is a specific sort of connection, on her account, between counterfactual reasoning and thought about enabling causes.

The distinction between strong and enabling causes, in Byrne's sense, relates to a topic sometimes referred to as 'causal selection'. Confronted with a scenario in which a certain type of event happens, individuals can make judgements not just as to which factors in the scenario are amongst the causes of the event, which are merely correlated with it because they are other effects of a common cause, and which of them are causally unrelated to it. In our causal judgements we also typically single out one or a small group of factors belonging to the first category as *the* cause or causes of the event in question.

Woodward, in his chapter, argues that the question as to what principles govern causal selection is quite separate from the question as to how we distinguish, for instance, between causation and mere correlation. It is specifically the latter question that Woodward's own version of a counterfactual process view of causal thought, involving the broad reading of the notion of a counterfactual, is focused on. In this, Woodward's main theoretical interests (at least in his chapter for this volume) may be seen to mirror in certain respects those of Lewis, who even went as far as denying that the topic of causal selection was of any significant philosophical interest. As Lewis puts the point,

> We sometimes single out one among all the causes of some event and call it 'the' cause, as if there were no others. Or we single out a few as the 'causes', calling the rest mere 'causal factors' or 'causal conditions'. Or we speak of the 'decisive' or 'real' or 'principal' cause. We may select the abnormal or extraordinary causes, or those under human control, or those we deem good or bad, or just those we want to talk about. I have nothing to say about these principles of invidious discrimination (Lewis, 1973a: 556f.).[6]

A rather different attitude towards the issue of causal selection can be found in Peter Menzies' chapter for this volume. In some respects, Menzies' chapter is the one that most closely adheres to the project of providing a counterfactual theory of causation along traditional Lewisian lines. However, he also points out that there is a class of counterexamples to Lewis' original theory, the common theme of which is that the theory over-generates causes. For instance, if the gardener fails to water a plant and it dies, Lewis' theory counts his failure as a cause of the plant's death. Yet it also counts the Queen's failure to come and water the plant instead as a cause in the same way. This counter-intuitive consequence is simply the result of allowing absences to figure as causes at the same time as treating causal selection as not reflecting any differences of genuine philosophical significance.

Menzies traces back the problem of over-generation of causes to a particular feature of Lewis' theory, namely a centring principle that Lewis imposes on his semantics for counterfactuals. As he points out, Lewis' definition of counterfactual dependence, which is to be the basis for the analysis of causation in counterfactual terms, requires the truth of two counterfactuals:

(i) If c were to obtain, e would obtain.
(ii) If c were not to obtain, e would not obtain.

Because of the centring principle Lewis imposes, however, (i) comes out as trivially true if c and e in fact obtain. Menzies, by contrast, argues that (i) should not be regarded as trivially true if c and e obtain—indeed, we can give examples where (i) seems clearly false, even though both c and e obtain (see also Edgington and Woodward's chapters on related points). Thus, once we give up the centring principle, counterfactual dependence becomes a much stronger condition. And, as Menzies goes on to argue, this allows us to avoid the problem of over-generation of causes that besets Lewis' theory.

Setting aside some of the technicalities, Menzies' suggested strengthening of the definition of counterfactual dependence can be seen as trying to capture the intuitive idea that a cause is a disruption to the way things proceed normally. Thus, if c and e in fact both obtain, but circumstances in which c obtains are not normally circumstances in which e also obtains, we do not count c as a cause of e. This is why we count the

[6] Compare also John Stuart Mill: 'Nothing can better show the absence of any scientific ground for the distinction between the cause of a phenomenon and its conditions, than the capricious manner in which we select from among the conditions that which we choose to denominate the cause' (Mill, 1846: 198).

gardener's failure to water the plant as a cause of its death, but not the Queen's failure to do so.

It is at this point that Menzies makes a connection between his own proposal and some of the psychological literature dealing with what is commonly referred to as counterfactual availability. Psychologists have discovered a number of factors that determine which particular counterfactuals individuals are most likely to generate in response to a given causal scenario. It has long been assumed in much of this literature that counterfactual availability might hold the key to causal selection, i.e. to the question as to which factors people single out as the cause (or causes) of a given event. As David Mandel discusses in his chapter, however, the most straightforward way in which one may try to make the connection is not supported by empirical research. When people are asked to generate 'but for' counterfactuals about a certain causal scenario (i.e. counterfactuals corresponding to clause (ii) in the above definition of counterfactual dependence, taken in isolation), their answers typically do not correspond to the answers they would give if asked about the causes of the outcome of the scenario. Even if this is true, though, Menzies' chapter suggests an alternative way in which counterfactual availability might still be related to causal selection. Put crudely, the proposal would be that counterfactual availability governs the way in which we decide whether the definition of counterfactual dependence *as a whole* is met in a particular case or not. In elaborating a proposal along these lines in more detail, Menzies introduces the further technical notions of a deviant and a default counterfactual, before showing how the proposal might be brought to bear on the cases that prove problematic for Lewis' theory.

That there is, in fact, a connection between counterfactual availability and causal selection is also supported by a number of empirical studies reported in Christopher Hitchcock's contribution to this volume. In these studies, a range of factors that influence counterfactual availability are also shown to influence causal selection. What Hitchcock's chapter brings out in particular is that causal selection can be influenced not just by empirical norms, but also by social, legal, and even moral norms. This, though, raises an important general question that might be brought out by noting something of a difference in approach between Menzies and Woodward.

Menzies offers a unified account of the truth conditions of causal claims that effectively builds the normative criteria governing causal selection into those truth conditions. That is to say, on Menzies' account, the very meaning of a causal claim can turn on what we take to be the norms in operation in a particular situation. As a result, Menzies' theory has the feature—which he acknowledges to be controversial—of making the meaning of causal claims in some sense subjective.

Woodward, by contrast, explicitly mentions the fact that causal selection may turn out to be an irreducibly subjective matter as a reason for separating out what he sees as two quite different projects. One of them is to account for causal selection, the other to account for the principles according to which we distinguish between, say, causation and mere correlation. Thus, a key question that Woodward can be seen to be driving at

here is this: To what extent is it possible, by separating out different roles that counterfactual reasoning might play in our thinking about causation, to isolate something like an objective core to our thinking about causation? Or does adopting what we have called a counterfactual process account of causal thought ultimately commit one to a form of anti-realism about our ordinary notion of causation?[7] This is just one way in which considerations about potential psychological connections between counterfactual and causal thought can, ultimately, be seen to lead right back to some of the fundamental types of questions philosophers have been asking about causation.

References

Harris, P.L., German, T., & Mills, P. (1996) 'Children's Use of Counterfactual Thinking in Causal Reasoning', *Cognition* 61: 233–59.

Lewis, D. (1973a) 'Causation', *Journal of Philosophy* 70: 556–67.

—— (1973b) *Counterfactuals*. Oxford: Basil Blackwell.

Mill, J.S. (1846) *A System of Logic*. New York: Harper & Brothers.

Perner, J. (1991) *Understanding the Representational Mind*. Cambridge, MA: MIT Press.

Price, H. & Cory, R. (2007) (eds) *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Clarendon Press.

Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

[7] Related issues of realism vs anti-realism about causation are at the forefront of many of the chapters in Price and Cory (2007). Compare also the way in which the question of realism figures in Roessler's contribution to the present volume.