

5

Causation, Intervention, and Agency: Woodward on Menzies and Price

Huw Price

1 Introduction

In ‘Causation as a Secondary Quality’ (Menzies and Price 1993; hereafter ‘CSQ’) Peter Menzies and I defended the view that, as we put it, ‘the ordinary notions of cause and effect have a direct and essential connection with our ability to intervene in the world as agents’ (CSQ: 187). We called this the agency theory of causation, and attributed it to Collingwood (1940), Gasking (1955), and von Wright (1975) before us (and tentatively also to Ramsey 1929). We argued that four common objections to this view are parallel to, and no more forceful than, four objections that could be raised to standard treatments of colour as a secondary quality (to all of which there are familiar responses). Hence our title: we were proposing that the agency view should be regarded as taking causation, too, to lie on the ‘secondary’ side of the primary/secondary divide; and that once this point is in the open, the usual objections to the agency view lose their force, because the familiar responses to the corresponding objections in the case of colour are easily transformed into the replies the agency theory needs in the case of causation.

As it turned out—a matter much more of correlation than causation, unhappily for us!—the decade following the publication of our paper was extremely fruitful for the investigation of links between causation and manipulation, thanks to the work of Judea Pearl, Peter Spirtes and his collaborators, and Jim Woodward, amongst others. In Woodward’s case, the decade culminated in the publication of his widely acclaimed book, *Making Things Happen* (2003). There, and in some more recent papers, Woodward devotes some space to distinguishing his view from earlier manipulability approaches, including particularly that of CSQ, and to criticizing the latter view, on several points.

In the present chapter, with the benefit of a further decade’s hindsight, I want to discuss Woodward’s criticisms. My response is mixed. On the one hand, I want to

argue that the CSQ view is not as different from Woodward's own as he believes, and that insofar as it is different, it has some advantages whose importance Woodward misses. On the other hand, I think that the CSQ view also lacks some elements whose importance Woodward rightly stresses, and I shall discuss the question as to whether it can be improved, to add those features.

I shall thus be offering an updated, 'Pricean' version of the Menzies and Price view. Some of the updates are recent, reflecting what I feel I have learnt from Woodward's work. Others are older, bringing in considerations from my own work on these topics in the years since CSQ was published. And some go right back to the period in which that paper was written, reflecting some respects in which my own take on the issues under discussion was not precisely aligned with the stance of the joint paper.

My discussions with Peter Menzies in the 1980s were immensely illuminating, from my point of view, and a huge influence on the direction of my own work. But much of their value stemmed from the fact that our philosophical dispositions were always a little way apart: Peter tended to be more of a realist, and more of a metaphysician, than I was (or am). The formulation of the view in CSQ was to some extent a compromise, to bridge this gap, and I want to take the opportunity below to offer a revised formulation, in my own voice, in a couple of places. As I shall explain, I think this is relevant to our disagreement, or at least *my* disagreement, with Woodward, at some points. (In order to distinguish my voice from that of CSQ, I shall refer to its authors in the third person from now on, and usually simply as 'MP', for brevity.)

2 Causation as a Secondary Quality

In their own words, the four arguments against the agency view of causation that MP consider are as follows:

1. *Agency accounts confuse the epistemology of causation with its metaphysics.* It is widely conceded that experimentation is an invaluable source of evidence for causal claims; the objection is that it is a confusion to suppose that the notion of agency should thereby enter into the analysis of causal claims.
2. *Agency accounts are vitiating by circularity.* It is argued that the bringing about is itself a causal notion, and that this introduces a vicious circularity into an agency account.
3. *An agency account cannot make sense of causal relations between events which are outside the control of any agent.* For example, it is argued that such an account cannot make sense of the claim that the earth's revolution around the sun causes us to experience the seasons.
4. *Agency accounts make causation an unacceptably anthropocentric phenomenon.* Agency accounts are said to imply what is obviously false, namely that there

would be no causal relations if there were no human agents (or different causal relations if there were different human agents) (CSQ: 188).

In the remainder of this section I shall summarize the replies that MP offer to these four objections, adding some comments about how I feel that these replies can be strengthened, in a couple of cases, if the project is given a more Pricean spin. I shall also take the opportunity to respond to some of Woodward's criticisms, where these relate closely to what I want to say about the MP replies to the original objections. In the case of the third and fourth objections, however, I shall defer most of my response to Woodward until later in the chapter, to allow a more lengthy discussion.

2.1 *Epistemology Confused with Metaphysics*

MP ask their readers to consider the familiar 'dispositional theory of colour, according to which an object is red, say, just in case it would look red to a normal observer under standard conditions' (CSQ: 192). They note that '[t]his theory makes colour a secondary quality in the sense that the concept of colour is taken to be an extrinsic or relational one, where the constitutive relation is to a certain kind of human response: in the case of the colour red, the "looks red" response'; and go on to say that although it is of course true that this theory has epistemological implications, it doesn't confuse epistemology for metaphysics. The metaphysics of colour properly involves reference to human responses, on this view, but there is no confusion: that's what it is to be a secondary quality, at least on this kind of elucidation of the primary–secondary distinction. And they suggest that the same is true of the agency view: '[T]he central point is that the concept of causation is to be explained by relation to our experience as agents in the same way that the concept of colour as a secondary quality is to be explained by relation to our experience as observers' (CSQ: 193).

As I shall note below, MP's use of the phrase 'experience as agents' at this point turned out to be misleading, in that it obscured for some readers (including Woodward) a point MP had earlier stressed, concerning the extent to which their view differed from standard empiricism. As they had put it earlier: 'Empiricists need to keep in mind that human subjects have access to the world in two ways: as observers, certainly, but also as *agents*, capable of intervening in the processes of the world at will' (CSQ: 191).

Putting that aside for a moment, I want to note that there is another way in which the response to this objection might go, more in keeping with my own predilections (then as well as now, so far as I can recall). It is explicitly to *disavow* that the project of the agency theory should be seen as *metaphysics* in the first place. Rather, it should be seen as what I have sometimes called philosophical anthropology: the task of explaining why creatures in our situation come to speak and think in certain ways—in this case, in ways that involve causal concepts. I think that this is one of a range of philosophically interesting cases in which the useful questions turn out to be questions about human thought and language, not questions about other aspects

of the world (such as the nature of causation). I think the same about the standard secondary qualities, of course—this shift in no way undermines the analogy that Menzies and I drew in our paper, in my view.

I cannot defend this general philosophical orientation here, and have done so at length elsewhere (see, e.g., the papers collected in Price 2011, and Price et al. 2013). My point is simply that if one has signed up for the view that the project of the agency theory is not metaphysics in the first place, there is no room for the objection that it confuses epistemology for metaphysics.

2.2 *Vicious Circularity?*

The distinction between metaphysics and philosophical anthropology is also relevant to what I would now wish to say about MP's response to the second of the four objections they consider. Here, the concern turns on the fact that the agentive notion of 'bringing about' is itself a causal notion. Doesn't this introduce a vicious circularity into the proposed account of causation? MP replied that dispositional analyses of colour avoid this difficulty because their appeal to notions such as 'looking red' can be cashed out in ostensive terms. To put it very crudely, we can say something like this: 'To be red is to be such as to illicit *this* response in normally sighted humans in normal conditions' (while showing our normally sighted audience some red objects in normal conditions).

MP thus construed the core of the circularity objection to be a point about concept acquisition, comprising two claims: (i) that according to the proposed analysis, grasp of the concept of causation requires prior grasp of the notion of agency, for the latter is 'conceptually prior' to the former; (ii) that such prior grasp is impossible, however, because agency is itself a causal notion. MP's response is to accept (i) but reject (ii), arguing by analogy with the case of 'red' that the required concept of agency can be acquired by ostension definition:

[F]rom an early age, we all have direct experience of acting as agents. That is, we have direct experience not merely of the Humean succession of events in the external world, but of... doing one thing and thence achieving another. ... [T]hese cases provide direct non-linguistic acquaintance with the concept of bringing about an event; acquaintance which does not depend on prior acquisition of any causal notion. An agency theory thus escapes the threat of circularity. (CSQ: 194–5)

While I think this reply to the circularity objection stands up in its own terms—I shall respond below to some challenges Woodward raises to it—I would like to qualify it in two respects. The first is to note that there is at least one way to understand the project of a metaphysical analysis of causation to which it would not be an answer. Suppose we take ourselves to be puzzled by the nature of causality, expressing our puzzlement in questions like these: How does a world have to *be* to contain causation? Are the materials available in a bare 'Hume world' enough, for example, or do we need something else? And if so, then what, precisely? If a proponent of the agency

theory responds by saying something like this—‘If you’ve got agency in your world, you’ve got the basic raw materials for causation. Causation can be *constructed from*, or *reduced to*, the kind of materials that agency provides.’—then it does seem reasonable to object that the agency theorist has, in effect, helped herself to a special case of what we were looking for in the first place, in virtue of the fact that agency is a causal notion. (It would be as if bricks had to be made from a special kind of brick.)

I conjecture that Woodward interprets MP as being in the business of answering this sort of question. If he were right, then I think he would also be right that MP do not adequately respond to the circularity objection (unless perhaps by making agency a metaphysical primitive, which would be equally bad). But as far as I can see there is little if anything in CSQ to support this interpretation of their project, and there are several passages that count against it. For example, MP say:

[T]hese cases [of practical experience of agency] provide direct non-linguistic acquaintance with the *concept* of bringing about an event; acquaintance which does not depend on *prior acquisition of any causal notion*. An agency theory thus escapes the threat of circularity.

(CSQ: 195, emphasis added)

And earlier (CSQ: 194), they mention but decline a possible response to the circularity objection that involves conceding that ‘the theories [of causation and colour] in question are not meant to be reductive analyses which reduce the concepts of causation and colour to their atomic constituents’. The implication, presumably, is that they take it that the theories are ‘meant to be reductive analyses which reduce the *concepts* of causation and colour to their atomic constituents’. Insofar as MP’s project is a reductive one, in other words, it seems much closer to a kind of conceptual analysis than to the kind of metaphysical enquiry that would be vulnerable to the circularity objection. (In this respect, then, it is closer to Woodward’s own project than he realizes, for he explicitly disavows the kind of reductive ambitions that would fall victim to the charge of circularity.)

The second qualification I want to make about MP’s response to the circularity objection also turns on the issue of what we take the task of the agency theory to be. As I have explained, I think it is fair to say that MP take the task to be to answer questions such as ‘What is it for *X* to be a cause of *Y*?’, where the sought-for answer amounts to something like a conceptual analysis (this is how they differ from the more ‘material’ investigation just mentioned). Even in this conceptual form, the circularity objection has apparent bite. The agency theory is committed to giving an answer that mentions agency, after all, but how can we ‘get into the causal circle’, as it were, if the concept of causation reduces in this way to a concept that itself needs to be understood in causal terms?

The appeal to ostension provides an answer at this point, but in my view it is an answer that we only (seem to) need because we have asked the wrong question in the first place. If we make it clear at the beginning that we are not concerned with the project of providing a reductive *analysis* of the concept of causation, but rather with

the anthropological project of explaining its genealogy and use, then it is hard to see that there is even a *prima facie* concern about circularity. Our task as anthropologists is to explain a feature of what humans *say*—in this case, their use of causal concepts—in terms of what they *do*. It is no problem at all if *we theorists* characterize these doings in causal terms, so long as our subjects themselves don't need to do so, in order to get the linguistic behaviour in question off the ground.

This amounts to responding to the circularity objection by rejecting (i), rather than (ii), in the terminology I introduced above (in the second paragraph of this subsection). Once again, the colour case provides an example of the kind of thing that's needed, at least as a first step, so long as we move away from the dispositional analysis in that case, too. We simply imagine proto-humans habituating to grunting 'red' when they experience what we sophisticated anthropologists would describe as 'the seeing-red response'. Something akin to ostensive definition may well play a role here, too—proto-human Alice points at a tomato and grunts 'red' in the direction of proto-human Bob—but there is a crucial difference: this is an ostensive definition of 'red' itself, not of 'seeing red'. As I said, the latter notion can be confined entirely to the anthropologist's theoretical vocabulary, once the task is seen as anthropological explanation, rather than reductive analysis. The crudeness of this model doesn't blunt its message: provided we are focusing on *use*, on what speakers need implicitly to know how to *do*, it is no problem at all if our theoretical characterization of the practical capacities concerned itself employs some sophisticated descendant of the very concepts whose origins are in question.

This point is relevant to MP's options for responding to Woodward. To explain why, I turn to what seems to me a confusion in Woodward's reading of CSQ, albeit a confusion for which Menzies and Price deserve some of the blame.

2.2.1 WOODWARD ON MENZIES AND PRICE'S 'EMPIRICISM'

One of the remarks that Woodward makes 'by way of distinguishing [the Menzies and Price] position from [his] own' is as follows:

Menzies and Price's view of the origin of our concept of causality is a thoroughly empiricist one: we derive or learn the concept entirely from a characteristic kind of experience. As they see it, what is wrong with Hume's account is simply that he fixes on the wrong candidate for the relevant experience: it is our experience of acting as agents rather than the experience of regular succession that is crucial. But... the idea that our concept of causation is derived purely from experience (whether of agency or anything else) is simply mistaken. As with other concepts, the acquisition of the concept of causality involves a complicated interaction between prespecified neural mechanisms and 'learning'. Moreover, only some forms of learning are based on experience in the sense of that notion that Menzies and Price have in mind. (2003: 126)

Woodward then adds the following note:

A great deal of the learning that underlies the acquisition of causal concepts involves the acquisition of practical skills and habits that are not in any obvious sense 'based on' or derived from conscious experiences. There is now considerable evidence supporting the independence

of the systems involved in the acquisition of such ‘procedural’ memories from the ‘episodic’ memories of particular experiences on which classical empiricism is based . . . For this reason, among others, what is learned should not be equated with what is derived from conscious experience. (2003: 386)

He concludes:

There is no reason why a theory that takes the connection between causation and agency seriously should also be committed to the empiricist picture of concept acquisition advocated by Menzies and Price. This point is of considerable importance because it is this feature of concept acquisition that helps to ground the reductive features of their project. (2003: 126)

Two comments about this passage. First, as I have already said, I think that MP’s account can be improved by explicitly rejecting its *reductive* aspects—by accepting that the relation between agency and our causal concepts is best exhibited in a different philosophical vocabulary. As I’ll explain in a moment, I think that this makes it very easy indeed for MP to respond to Woodward.

Second, I think that Woodward’s charge is in any case based on a misreading of what MP have in mind. The distinction Woodward himself refers to here—that of ‘the acquisition of . . . “procedural” memories from the “episodic” memories of particular experiences on which classical empiricism is based’ (2003: 386)—is the same distinction that MP have in mind in the following passage, for example:

In our view the best way to characterize these parallels between causation and colour is to say that both can be viewed as secondary qualities under a generalized understanding of this notion. The usual characterization of a secondary quality, as a quality which tends to elicit a characteristic sensory experience in human subjects under specified conditions, is too restrictive . . . [I]t applies only to those properties which have a sensory import. As such, it perpetuates a constant philosophical pre-occupation with passive observation to the neglect of active intervention in the world: it should be kept in mind that we interact with the world not only as observers but also as agents. We advocate the adoption of a more general notion of a secondary property, which expressly disavows [this restriction]. (CSQ: 201–2)

This point is obscured, unfortunately, because MP also characterize their view using such phrases as ‘direct experience of acting as agents’ (CSQ: 194). The term ‘experience’ seems to associate them with the very empiricist paradigm—that of passive observation—from which they here strive to distance themselves. But the use of the term ‘experience’ they have in mind is a perfectly ordinary one. It is the one we use when we advertise for job applicants with ‘teaching experience’. What we want, obviously, is candidates who have taught, not candidates who have simply observed teaching. Similarly, what MP have in mind when they speak of acquisition of causal concepts by subjects who have ‘direct experience of acting as agents’ is this practical, work-experience notion, not the conscious, episodic, perceptual notion. Elsewhere, for example, they characterize their proposal as ‘admitting action on a par with perception as a means of access to the world’ (CSQ: 191–2). In this respect, then, MP were always

on Woodward's page—always concerned to point out that what classical empiricism misses is that we *do*, as well as *observe*, in our interactions with the world.¹

Once this is clear, Woodward's charge of excessive empiricism can be set aside, I think. The resources that MP imagine to be available to our ancestors, as the basis for acquisition of causal concepts, are essentially the same as those that Woodward himself imagines to be available. From this point, there are two ways for MP to proceed. One, sticking closely to their original model, would employ this modified empiricist picture as the basis of an ostensive account of the acquisition of the notion of agency, which would then be available, without threat of circularity, in the services of their reductive analysis of the notion of causation.² The other, abandoning the reductive aspect of the view, would simply aim to provide a direct account of the acquisition of causal concepts, turning on the idea that these are concepts we acquire in virtue of our practical activity as agents, in just the way that Woodward himself proposes.

In the latter case, the response to the circularity objection is particularly direct. The suggestion was that if MP's agency view were correct, our ancestors could never have found their way into the circle of using causal concepts, because they would have needed the concept of agency first, and that itself is a causal concept. As we have seen, MP try to meet this challenge by appeal to the possibility of ostensive definition of the concept of agency. Free from the constraints of the reductive model, however, there is no need to insist that our ancestors have the *concept* of agency, in any form, ostensively acquired or otherwise. They need to *be* agents, but they don't need to *think of themselves as agents*. And at this point Woodward's alternative story about the acquisition of causal concepts—which nowhere mentions reductive definition, but does, as we shall see, accord a central place to human agency—seems to be exactly what MP need, for this modified version of their project.

2.3 Unmanipulable Causes

This objection turns on cases in which, as MP put it, 'it is physically impossible, given the capacities of a normal agent, to manipulate the cause and effect at will' (CSQ: 195). In response, MP first consider a counterfactual proposal:

[I]t might be argued that it is in fact true that if, *per impossibile*,... an agent were able to manipulate continental plates, he would thereby be able to bring about earthquakes. Perhaps, one might try to make these counterfactuals plausible by invoking a conception of an ideal observer or agent, a conception which abstracts away from the usual limitations of human perception and manipulation. (CSQ: 196)

¹ This needn't mean that there is *no* role for experience in the passive sense, and indeed Woodward himself elsewhere (2007: 29) seems to allow that there might be: 'I suggest that... human subjects do have a characteristic phenomenology which is associated with voluntary action; they typically have a sense of agency or ownership of their behaviour that is not present when they act involuntarily.' (For more of this passage see §5.1 below.)

² The difference from standard empiricist ostension is that the subjects under instruction will be required to *do* things, as well as to *observe* things.

They reject this idea for reasons to do with the possibility of finkish and masking dispositions, and instead propose this alternative:

[W]hen an agent can bring about one event as a means to bringing about another, this is true in virtue of certain basic intrinsic features of the situation involved, these features being essentially non-causal though not necessarily physical in character. Accordingly, when we are presented with another situation involving a pair of events which resembles the given situation with respect to its intrinsic features, we infer that the pair of events are causally related even though they may not be manipulable. (CSQ: 197)

They note that ‘this inference relies on [a] principle of analogical reasoning’ that also seems operable in the case of colours.

MP’s response to this objection is a particular focus of Woodward’s comments, and I shall return to this topic below, to discuss Woodward’s criticisms, possible responses to them, and alternative arguments that MP might give in response to the original objection. (My reason for deferring discussion of this point is to have some of Woodward’s remarks about his proposed alternative on the table first.)

2.4 *Anthropocentricity*

Here the objection is that agency accounts are said to imply what seems obviously false, that is, that there would be no causal relations if there were no human agents, or different causal relations if there were different human agents. MP’s response, once again, is to point out that there are familiar answers to the analogous charge against standard treatments of colour as a secondary quality. Such accounts do not imply that sunsets were colourless before sighted creatures like us came along, or would have been a different colour if our colour vision had been different. We simply apply our actual standards, in considering the circumstances in question. There is anthropocentricity, certainly, but where it ought to be, in our colour *concepts*, not in the objects and their properties. Similarly for causation, MP claim.

Woodward claims that his view is less anthropocentric than that of MP, and that this is an advantage. I now turn to those comments—once again, my strategy will be to argue that the two views are closer than Woodward appreciates, and that to the extent that they differ, that leads to problems for Woodward’s view.

3 Woodward’s ‘Manipulationist’ Theory

Under the subheading ‘Nonanthropomorphism’, in a list of what he takes to be characteristics of his own view, Woodward says the following about the relation of his approach to the agency theory of Menzies and Price:

Notions such as “human action”...do not occur as primitives [in my account]...In this respect [my view] is quite different from traditional agency theories (such as those of von Wright and Menzies and Price...). In these theories, the characterization of a manipulation (or intervention) makes essential reference to human agency or free choice, and the hope is

82 CAUSATION, INTERVENTION, AND AGENCY

that this can be somehow grasped or understood independently of the notion of causality. By contrast [in my theory] there is nothing logically special about human action or agency: human interventions are regarded as events in the natural world like any other and they qualify or fail to qualify as interventions because of their causal characteristics and not in virtue of being (or failing to be) activities carried out by human beings. (2003: 103–4)

But Woodward does allow that human agency plays a role in the development of our notions of causality and intervention, as he notes in a later passage in which he reintroduces the spectre of an excessively anthropocentric alternative:

[O]n the view I am advocating, our notion of causality developed in response to the fact that there are situations in which we could manipulate *X*, and by so doing manipulate *Y*. This fact led us (3.3.1) to form the notion of a relationship between *X* and *Y* that would support such manipulations and to contrast this with the notion of a mere correlation that would not support such manipulations. However, it is built into the notion of a relationship that will support manipulations in this way that (3.3.2) such a relationship would continue to hold even if we do not or cannot manipulate *X*, or if our beliefs and attitudes were different, or even if we did not exist at all. If it is asked why (3.3.2) is built into our notion of causation, my response is that any other view of the matter would involve a bizarre and magical way of thinking, according to which our ability to manipulate *X* or our practical interest in manipulating *X* or our beliefs about the results of manipulating *X* somehow make it the case that a means–end connection comes into existence between *X* and *Y* where this connection would not exist if we did not have the ability or interest or beliefs in question. Taken literally, such a view, if intelligible at all, would require human beings to have god-like powers that they plainly do not possess. (2003: 120, emphasis in bold added)

3.1 *Dismissing the Spectre of Anthropomorphism*

Woodward does not actually say at this point that he takes a commitment to his (3.3.2)—the view that causal relationships ‘would continue to hold even if we do not or cannot manipulate [the events in question], or if our beliefs and attitudes were different, or even if we did not exist at all’—to distinguish his theory from that of MP, but the formulation is strongly suggestive of the familiar anthropocentricity objection to the agency view. As noted above, however, and as MP themselves point out, the corresponding objection gets little or no grip in the case of familiar secondary qualities. We simply apply our *actual* standards to say that sunsets would have been red even if humans had developed different colour vision, or had never evolved at all. And this is entirely compatible with recognizing that had we evolved differently, we might have employed an entirely different set of colour concepts.

Once again, MP argue that the same is true of causation, though they argue that there is a difference of degree: it is harder to imagine the required variation in the case of causation than colour. Indeed, MP suggest that there might be no variation possible in this case, except variation that would result in having *no* concept of causation:

In the previous section we saw that by appealing to a principle of analogical reasoning an agency approach may extend its scope well beyond the domain of those things in a particular

world that the agents of that world can actually influence. (This was the gist of our reply to the non-manipulability objection.) In consequence, it is far from clear that any modification of mere degree in our powers as agents will issue in any modification in the causal relations we are thus inclined to ascribe. On the contrary, it seems that agents with different capacities will nevertheless envisage the same range of possible causal relations, provided that they employ the principle of analogical reasoning we noted earlier as licensing the extrapolations of their manipulative capacities.

This suggests that in the case of agency, the only relevant possible world for the purposes of the anthropocentricity objection is the limiting case: the world in which, like Dummett's intelligent trees, cognitive beings have no powers as agents. (CSQ: 200–1)

In more recent work (e.g. Price 1996, 2007), I have defended a different view. I have argued that there is at least one very significant variation that we can imagine, involving agents whose perceived direction of time is the opposite of ours. (They are imagined to live in a region of the universe in which the thermodynamic 'arrow' points in the other temporal direction.) I maintain that just as such agents would disagree with us about the direction of time, they would also disagree about the direction of causation.

In Price 2007 I compare this to familiar perspectival categories, such as near and far, left and right, and foreigner and local. The people on the other side of the border mean the same by 'foreigner' as we do, in one obvious sense. (We assume for the sake of the example that they speak English too.) But whereas we apply it to them, they apply it to us—annoying of them, perhaps, but it is hard to maintain that they are actually *wrong!* I point out that this does not involve denying the reality of foreigners. Foreigners are 'not figments of our collective imagination, or social constructions, or useful fictions'; on the contrary, they 'are as real as we are' (Price 2007: 250). Nevertheless, we learnt something when,

minds broadened by travel, we realized that foreigners themselves use the very same concept, but apply it to us! ... [T]he reality of foreigners notwithstanding, there's a sense in which foreignness is a less objective matter than we used to think. (2007: 250–1)

Let me now relate this comparison—of *causation* to *foreignness*—to Woodward's characterization of his own view. As I noted above, Woodward says the following:

[O]n the view I am advocating, our notion of causality developed in response to the fact that there are situations in which we could manipulate *X*, and by so doing manipulate *Y*. This fact led us (3.3.1) to form the notion of a relationship between *X* and *Y* that would support such manipulations and to contrast this with the notion of a mere correlation that would not support such manipulations. (2003: 120)

Similarly, we might say, our notion of foreignness developed in response to our realization that there are people who are not of our tribe. This led us to form a notion of a characteristic—'foreignness', as we came to call it—possessed by all and only the people of whom this was true.

84 CAUSATION, INTERVENTION, AND AGENCY

In this case, we come to see that there is a contingency involved in the application of the term: had we been different in identifiable ways—had we been foreigners, in fact!—the same term would have applied to different objects. I claim that the same is true of causation, construed as Woodward here described. Had *we* been otherwise, the same procedure would have led us to pick out *different* relations between *X* and *Y*—or the same relation in the opposite direction, at least.

3.1.1 TWO LESSONS WE LEARN FROM ‘FOREIGNERS’

It is worth distinguishing two lessons that emerge from the case of the notion foreignness, both of which I take to be applicable to causation, too. The first is what we might call the context-sensitivity or perspectivity of the concept, the fact that for speakers in different circumstances (in the case of foreignness, belonging to different tribes), the concept picks out something different. *Our* use of the concept picks out *them*, and vice versa, but there’s an obvious sense in which it is the same concept in both cases. The second is more subtle—we might call it the interest-relativity of the concept. In the case of foreignness, it turns on the contingent fact that we are tribal in the first place. Creatures who were not tribal would not be in a position to employ the notion of foreignness at all (because, as we might put it, the rule for using the concept requires that one be the member of a tribe).

This distinction corresponds to a distinction between two ways in which speakers may differ. The first kind of difference—call it an *intramodal* difference—is that between speakers who both have the kind of context or perspective required for the use of a perspectival concept, but occupy *different* contexts of that kind. This is the difference between us and them with respect to foreignness, in a tribal society. The second kind of difference—*extramodal* difference—is that between speakers who occupy a context of the relevant kind and those who occupy no such context. This is the difference between us and our distant non-tribal descendants, for whom the notion of foreignness is an unusable relic of another age.

In my view, we can make sense of both kinds of difference with respect to causation, too. We differ *intramodally* compared to creatures who are also agents, but have the opposite temporal perspective to our own. We differ *extramodally* with respect to creatures who are not agents at all, and therefore lack ‘what it takes’ to employ the concept of causation in the first place. To appreciate the sense in which causation is an anthropocentric notion, we need to recognize the possibility (in principle!) of both kinds of variation, with respect to our own situation, and its implication for the use of the concept, in each case.³

³ It is worth emphasizing that this kind of anthropocentricity is something visible from the anthropological viewpoint (focusing on concepts), not from the metaphysical viewpoint (focusing on the causal relations themselves). When we adopt the latter standpoint, we typically rigidify on the basis of our actual perspective, as in the colour case.

The upshot, I think, is to undermine or at least significantly qualify the view that interventions are simply a mind-independent category, to which our manipulative practices give us access (by instantiating interventions themselves, at least to some extent). The possibility of intramodal variation reveals that nature offers (at least) two alternative ways of carving out such a category, and that which one we latch onto depends on contingencies about us. The possibility of extramodal variation reveals that in a deeper sense, too, the kind in question reflects a way of modelling the world that depends on the fact that we are agents.⁴

3.1.2 HOW MANY ALTERNATIVE CAUSAL VIEWPOINTS?

Agents with the opposite temporal orientation to our own would provide a stark illustration of intramodal variation with respect to causation, but do we need to go so far afield? The question is of pragmatic as well as theoretical interest, from my point of view, for my experience is that the ‘sci-fi’ nature of the time-reversal case tends to limit its impact—at least among recipients not already au fait with ‘the view from nowhen’ (Price 1996)! But I think we can bring the point down to earth, and indeed connect it with some issues raised by Woodward himself, if we think about the general features of an agent’s perspective that the time-reversal case exploits. I discussed these features in Price 2007, and proposed this general characterization of the nature of deliberation:

In any deliberative process, presumably, there must be a range of things that the deliberator in question takes to be matters for deliberation: in other words, the alternatives among which she takes herself to be deliberating. For formal convenience, let’s regard these alternatives as a class of propositions, denoted by *OPTIONS*. These are the propositions the agent takes herself to have the option of ‘deciding to make true’, in other words. It will be helpful to subdivide this class into *DIRECT OPTIONS*, comprising those matters over which an agent takes herself to have immediate control, and *INDIRECT OPTIONS*, comprising those ends she takes herself to be able to accomplish indirectly, by an appropriate choice from her *DIRECT OPTIONS*. And let the *FIXTURES* denote everything else—all matters of fact that are not held to be a matter of choice in the deliberation in question.

FIXTURES will contain a subset, *KNOWNESSES*, comprising those facts the deliberator takes herself to *know* at the time of deliberation, and also a larger subset, *KNOWABLES*, comprising matters she regards as either known or knowable, at least in principle, before she makes her choice. Why must *KNOWNESSES* and *KNOWABLES* be subsets of *FIXTURES*? Because it seems incoherent to treat something both as an input available to the deliberative process, at least in principle, and as something that can be decided by that process. Control trumps a claim to knowledge: I can’t take myself to know that *P*, in circumstances in which I take myself to be able to decide whether *P*, in advance of that very decision. (2007: 275)

⁴ Again, terms such as ‘near’ and ‘far’ provide an excellent analogy. There, too, we have both intramodal variation, in virtue of the fact that speakers may occupy different locations, and at least a possibility of extramodal variation, in virtue of the fact that a speaker might in principle occupy no particular location.

As I go on to say, this gives us

a very simple template, characterising a deliberator's view of the world. In terms of this template, acting, or intervening, is a matter of fixing something not already fixed—of moving something from *OPTIONS* to *FIXTURES*, as it were. (2007: 276)

For present purposes, the importance of this characterization of the abstract structure of an agent is that it brings into view the real sources of the contingency of our causal perspective. Everything turns on what we can know before⁵ we act, and what we take to be under our control (under idealization, no doubt, in both cases). The possibility of time-reversed agents provides a dramatic and (at least in some sense) physically well-motivated way to vary these factors, and so produce alternative causal viewpoints, but it isn't the only way, and the abstract characterization provides a recipe for constructing more.

And at this point, in fact, a consideration noted by Woodward himself as a source of some 'subjectivity' (2003: 89) in ordinary causal judgements seems to fit neatly into this abstract model. Woodward (2003: 86–91) discusses the dependence of the acceptability of ordinary causal claims on what speakers take to be 'serious possibilities' in the circumstances under consideration. When a patient dies for lack of antibiotics, for example, we don't hold a stranger in a distant city causally responsible for the death, even though it may be true that had the stranger visited the patient, bearing antibiotics, he would have survived. Such an occurrence is not regarded as a 'serious possibility', and is hence discounted as a causal factor. Woodward acknowledges that the decision as to what to treat as a serious possibility depends in various ways on our own interests and beliefs, and concludes that this does introduce at least a small element of 'subjectivity' into his interventionist account—though, as he argues, it is an element that other approaches to causation will be hard-pressed to avoid. As he puts it later, the fact seems to be that there is 'a limited respect in which . . . which causal claims we accepted as true . . . are influenced by what we take to be a "serious possibility"' (2003: 118).

In terms of my model, the way to describe this kind of case is to say that by default, we treat the behaviour of distant strangers as part of the *FIXTURES*, rather than the *OPTIONS*, direct or indirect. But these choices are contextual, in the way that Woodward notes, and this shows up in our causal judgements. Differences between speakers—in Woodward's terms, cases in which one speaker treats something as a serious possibility and another does not—can thus represent familiar, homely examples of intramodal variation, in my notation.

In the homely as in the sci-fi cases, I take the lesson to be that when Woodward says that 'our notion of causality developed in response to the fact that there are situations in which *we* could manipulate' (2003: 120, emphasis added), the indexical term 'we' is ineliminable. Agents with different epistemic 'situations' to our own will

⁵ As I note, this 'before' should be understood in terms of the personal time of the deliberator.

make different judgements about what could be manipulated by manipulating what, and there's no objective sense in which we are right and they are wrong—to think otherwise is to accord our own viewpoint a god-like priority that, as Woodward says, it plainly does not possess. (Here, as in many other cases in the history of science and philosophy, it is the modest, 'subjectivist', Copernican view that does the better job of recognizing the contingencies and limitations of the human standpoint, and the objectivist view that confuses us with gods.)

3.2 *Wasteful and Gratuitous?*

Similar comparisons also provide a response to a further objection that Woodward raises to the MP view, immediately following the passage quoted above:

This conclusion [i.e. if I interpret Woodward correctly, the conclusion that the MP view 'would involve a bizarre and magical way of thinking'—HP] is reinforced by [a] naturalistic, evolutionary perspective... According to subjectivist accounts, causal relationships have their source in facts about us—facts about our expectations, attitudes, and so on—which we 'project' on to the world... [W]hat is the evolutionary story about the benefits we derive from this projective activity? After all, our projectivist tendencies systematically lead to beliefs that, by the subjectivist's own account, are mistaken or ungrounded—mistaken in the sense that they ascribe a false objectivity to causal claims or involve thinking of the distinction between causal and correlational claims as having an objective basis in nature rather than in facts about us. Why should we and other animals go to the trouble of distinguishing between causal and correlational relationships if all that is 'really out there' in the world are correlations? All that projecting seems wasteful and gratuitous. (2003: 120–1)

Once again, we need only think about the case of the secondary qualities, or of 'perspectival' asymmetries such as there–here, past–present, you–me, or foreigner–local. None of these properties or asymmetries are simply 'there' in the world, visible from an Archimedean point of view. They all reflect our viewpoint, or 'location', in one way or another. But there's no mystery about why we have evolved so as to draw such distinctions.

Take the case of the familiar indexicals, for example. To paraphrase Woodward, why should we go to the trouble of distinguishing between here and there, now and then, self and other, if all that is 'really out there' in the world are the bare non-indexical facts? All that projecting seems wasteful and gratuitous! But Perry (1979) and others have shown us why it isn't wasteful and gratuitous, in the indexical case. On the contrary, as Perry puts it, the indexical is essential, for creatures in our circumstances: creatures who need to coordinate their own actions and observations with third-person maps of their environment.

The general lesson is something like this. Many of our concepts are useful to us in virtue of contingent features of our own circumstances—for example, in the indexical case, the fact that we are 'located' in space, time, and communities of individuals. It is not surprising at all, from a naturalistic perspective, if some of our concepts reflect

these ‘located’ features in essential ways—that is, roughly, in such a way that we cannot understand the concept in question except with reference to the feature in question. (At least one way in which this might happen is for the ‘location’, in this generalized sense, to play a role in the use-rules governing the concept.) There is no affront to naturalism in this idea: on the contrary, it would surely be extraordinary if our conceptual structures did not reflect these contingencies. We are not gods, so why should we think in the kind of conceptual repertoire that gods might use?

I take the insight of the agency view to be that causation is one of these ‘located’ concepts. Its particular link is to the fact that we are agents, capable of intervening in our environment at will. This might seem to leave the view open to the charge that, as Woodward puts it, it ‘flies in the face of any plausible version of naturalism: it makes agency out to be a fundamental, irreducible feature of the world and not just one variety of causal transaction among others’ (2003: 123). But, as I noted above, this is a mistake (perhaps encouraged by MP’s tendency to characterize the project of an agency theory in a metaphysical key rather than an anthropological key). The agency view requires that we have a practical acquaintance with agency ‘from the inside’, as it were, so that we are able to acquire implicit use-rules that, *if made explicit*, would need to refer to it. But this is in no way incompatible with regarding agency as an element in the causal web, ‘one variety of causal transaction among others’, once we have the concepts and start to reflect on such matters.

All of this seems to be entirely in keeping with the way in which Woodward frames the motivation for his own project, at one point:

As a preliminary motivation, let me begin with a question that is not often asked in philosophical treatments of causation: What is the point of our having a notion of causation (as opposed, say, to a notion of correlation) at all? What role or function does this concept play in our lives? An important part of the appeal of a manipulability account of causation is that it provides a more straightforward and plausible answer to this question than its competitors. (2003: 28)

The difference, if there is one, is that I have on the table the possibility that the answer to this question will need to appeal to contingencies of our own nature, in such a way that any theory of concepts that sees their role in crudely representationalist terms will simply be blind to the need for some interesting theoretical work somewhere else (i.e. in the story about how the use of the concepts depends on the contingencies in question, in the sense manifest in possibility of intramodal and extramodal variation).

4 The Problem of Unmanipulable Causes

Let us now return to objection three. Woodward says that MP ‘face the obvious problem about the extension of causal concepts to circumstances in which manipulation by human beings is not possible’ (2003: 123). He argues that MP’s response (as above) in terms of resemblances in intrinsic properties is unsatisfactory because,

as he puts it, he sees ‘no reason to believe . . . that this notion of resemblance can be characterized in noncausal terms’ (2003: 125).

The problem with this suggestion becomes apparent when we consider, for example, the nature of the ‘intrinsic’ but (allegedly) ‘noncausal’ features in virtue of which the movement of the continental plates ‘resemble’ the artificial models the seismologists are able to manipulate. It is well-known that small-scale models and simulations of naturally occurring phenomena that superficially resemble or mimic those phenomena may nonetheless fail to capture their causally relevant features because, for example, the models fail to ‘scale up’—because causal processes that are not represented in the model become quite important of the length scales that characterize the naturally occurring phenomena. Thus, when we ask what is [it] for a model or simulation that contains manipulable causes to ‘resemble’ phenomena involving unmanipulable causes, the relevant notion of resemblance seems to require that the same *causal* processes are operative in both. I see no reason to believe (and Menzies and Price provide no argument) that this notion of resemblance can be characterized in noncausal terms. But if the extension of their account to unmanipulable causes requires a notion of resemblance that is already causal in character and that, *ex hypothesi*, cannot be explained in terms of our experience of agency, then their reduction fails. (2003: 125)

However, I think that if there were a problem here for MP, it would equally be a problem for Woodward’s own view. As Woodward will agree, we extend our causal notions into many regions in which we can be sure that we will never intervene: the inside of the sun, distant galaxies, and the distant past, for example. We take it for granted both that there is causation in these regions, and that it is broadly similar to causation in more familiar regions—it doesn’t work backwards, for example. On what basis do we take these distant regions to be so similar to our own, in causal respects?

There are two possibilities at this point. One is that we rely on similarities in noncausal respects to ground the inference. But this would be to grant what Woodward here wants to deny to MP, namely, that there are relevant similarities, characterizable in noncausal terms. The second option is that there are inference principles of some kind—perhaps grounded in physical symmetries, and/or whatever else might be held to underpin the normal inductive procedures of science—that are taken to license the inference directly.

What are these inferences? Just the ones needed to support counterfactuals. As Woodward puts it elsewhere:

It seems uncontroversial that the claim that *C* causes *E* can be true even if *C* is not actually manipulated—any account that suggests otherwise is a non-starter. This observation suggests that manipulationist accounts should be formulated as *counterfactual* claims connecting causal claims to claims about what would happen *if* certain manipulations *were* performed. (2009: 236)

In this passage Woodward is using the term ‘manipulation’ in a way I take to be neutral between his own preferred version of the manipulationist account—the ‘interventionist’ approach, as he calls it—and MP’s agency view. But it is hard to

see what basis there could be for the claim that the required counterfactuals are harder to justify in one case than the other. Indeed, if MP had opted for their first suggestion, and responded to the problem on unmanipulable causes by appealing explicitly to counterfactuals, then it would be even more difficult to see how there could be space for them to be in trouble at this point, while Woodward is not.

4.1 *Extension to Remote Cases*

Moreover, I think that Woodward's view that his approach is more 'realist' than that of MP is likely to prove more of a hindrance than a help at this stage, in that it makes him more prone to sceptical worries about whether there is *really* causation inside the sun, or whether causation *really* runs past-to-future in neighbouring galaxies. To introduce this point, let me appeal once more to the analogy with indexicals. Wearing our old Newtonian hats, we have no trouble in making sense of the question as to whether it is *now* light or dark at some specified point on the surface of a distant planet, where no sentient creature could possibly exist. Does this commit us to the view of the so-called A-theorists, that the distinction between past, present, and future is an intrinsic feature of reality? Pretty obviously not. A B-theorist will say that our extension of the indexical notion *now* to remote places requires only the non-indexical notion of *simultaneity*: an event is happening *now* on a remote planet if and only if it is simultaneous with what is happening *now*, here. Simultaneity thus provides a tenseless notion of similarity, that grounds our extension of the tensed notion from one context (our own neighbourhood) to another (the remote planet).

The A-theorist might object at this point that the notion of simultaneity is *not* tenseless: on the contrary (she claims), two events are simultaneous iff it is, was, or will be the case that they are co-present (or something of that kind), so that the notion of simultaneity depends on that of presentness. Whatever might be said in favour of this view, however, it had better not stand in the way of whatever ordinary processes we take to determine whether it is now night or day on the distant planet—the A-theorist needs those inferences as much as anybody. (And, *prima facie*, her additional realism about tensed properties makes things more difficult, in that it introduces new sceptical possibilities. How do we know that the A-theorist's notion of simultaneity tracks the physicist's notion of simultaneity, after all?) Notice that relativity cuts equally on both sides of this debate. It undermines the idea that there is an objective notion of simultaneity to ground the extension of an indexical *now* to remote locations, thus requiring the B-theorist to acknowledge that what she took to make sense—the question whether it is now night or day at the remote location—actually does not make sense (unless relativized to a suitable reference frame). But it surely requires the same concession of the A-theorist, too, unless she is to be left in the position of arguing that there is a fact of the matter, but that it is inaccessible to us.

Can we imagine the same state of affairs in the case of causation, read in manipulationist terms? I think that we can. Imagine that some distant region in spacetime turns out to be linked to our own region only via two wormholes; and that

these wormholes turn out to have opposite temporal parity, in the sense that if one of a matched pair of clocks is sent through each, the clocks are running in opposite temporal senses when they reach the other side. (We do not assume that there is a fact of the matter about which is ‘right’.)

In these circumstances, I think it is difficult to maintain that we have a clear sense of what we could do to manipulate what, in the region on the far side of the wormholes. It would all depend on which wormhole we used! By my lights, this is an example of how the extension of our anthropocentric notion of causation to regions in which we cannot actually manipulate things is in principle always provisional, and subject to correction in the light of learning more about the relevant physics. (I want to say the same about Woodward’s example involving scaling—of course we can get things wrong!) And once again, I think Woodward faces a dilemma: either he relies on the same principles of extension, and is hence subject to the same exigencies; or he is left defending an implausible objectivity, vulnerable to scepticism in cases such as these.

4.2 Summary—the Objectivist’s Dilemma

The problem was to explain how an agency account of causation could explain the extension of the concept of causation from cases in which we can make manipulations to cases in which we cannot. Let us call this the problem of extending causal models from *local* cases to *remote* cases. (‘Local’ and ‘remote’ are thus terms of art, for present purposes.) The MP proposal was that the extension works by dropping down to a subcausal level of description, and extending our models exploiting similarities at that level (plus, presumably, some sort of supervenience principle).

The new proposal we have on the table is that the extension from local to remote cases takes place at the higher level, exploiting such things as physical symmetries (‘spatial translation doesn’t make a difference’, for example). My argument has been that Woodward’s own view requires some such extension principles at this point, and whatever he uses, MP can use too. A possible response on Woodward’s behalf is that he has in mind extension principles that would take us to regions where human agency cannot sensibly be considered to go (inside the sun, or into distant galaxies, for example). My reply is to point out that unless the extension does avail itself of constraints grounded in our (actual) agents’ perspective, it cannot resolve ambiguities that stem from the contingencies of that perspective. (It cannot provide any justification for taking causation to have the same temporal orientation in the distant galaxy, for example.)

So there is a dilemma for anyone more objectivist than MP about these cases.⁶ Without the constraint imposed by being able to extend *our* standards into

⁶ Strictly speaking this argument isn’t available to MP in the causal case, because they claim that there are no other possible causal perspectives (except the ‘no causation’ option associated with intelligent trees). But as I have said, I think that MP were wrong at this point.

counterfactual cases, we are left unable to resolve the ambiguity that stems from the contingency of the original notion—its relativity to *our* situation and interests! Once again, objectivism leads to scepticism, and any principle the objectivist invokes to deal with the problem will serve equally well for the more subjective view.

Moreover, I stress that there is nothing unique about causation here. The same is true of any of the vast range of concepts that have some in-built relativity to our own situation and interests. In all cases, an unambiguous extension to remote cases depends on our being able to map the relevant aspect of our situation and interests into those remote circumstances—to the extent that we can't do that, we have no basis to resolve the ambiguities in one way rather than another, in the remote circumstances. (Where the extension of our particular perspective really doesn't make sense, in other words, objectivism is in trouble.)

5 Objectivity Again

So far, I have been arguing that Woodward's criticisms of the MP view are largely unsuccessful, especially if the latter view is tweaked and clarified in various respects. I have suggested that the MP view is actually closer to Woodward's own position than he realizes, and moreover that if Woodward tries to establish a difference by moving in the direction of 'greater objectivity', then danger lurks—danger that MP's more modest view avoids. But I now turn to one aspect of Woodward's discussion of the objectivity of causality that seems to me a clear advance on the MP view, and that I want to endorse (almost) without qualification. Once again, I think that the MP view can take it into account, in ways that turn out to be fully in the spirit of the original comparison between causation and the familiar secondary qualities. In that sense, then, it doesn't in the end represent a damaging objection, but it is certainly an important addition.

The point turns on a distinction between three varieties of agent that Woodward draws in the following terms:

1. An agent whose instrumental behaviour and learning is purely *egocentric*. That is, the agent grasps (or behaves as if it grasps) that there are regular, stable relationships between its manipulations and various downstream effects but stops at this point, not recognizing (or behaving as though it recognizes) that the same relationship can be present even when it does not act, but other agents act similarly or when a similar relationship occurs in nature without the involvement of any agents at all.
2. An agent with an *agent causal* viewpoint: the agent grasps that the very same relationship that it exploits in intervening also can be present when other agents act.
3. An agent with a *fully causal* viewpoint: The agent grasps that the same relationship that the agent exploits in intervening also can be present both

when other agents intervene and in nature even when no other agents are involved. This involves thinking of causation as a tertiary relationship (Woodward 2007: 32).

One of the interesting things about this three-way distinction, from my point of view, is that it, too, has obvious echoes in the case of the familiar secondary qualities. There, too, there seems to exist a similar range of options: one might think of what one's senses deliver as a private, purely egocentric experience; as an experience that other observers will share; or as a revelation of a property of the object, present in nature in the absence of observers. Much concern about the nature of the secondary qualities turns, in effect, on whether they reach the third level. (If a tree falls in an uninhabited forest, does it make a sound? Or, as Galileo puts it, do the sensory properties 'have their residence solely in the sensitive body?') But there is also fascinating work—here I am thinking particularly of Sellars's classic discussion in 'Empiricism and the Philosophy of Mind' (Sellars 1956)—of the step from stage one, on the one hand, to stages two and (perhaps also) three, on the other. Here, at least at first pass, the question is something like this: What is involved in coming to regard our colour experience as a means of access (and therefore fallible access) to something objective?⁷

I mention this mainly to call attention to the importance and interest of the comparison between the ways these issues play out in the two cases—causation, on the one hand, and colour and the other secondary qualities, on the other. It now seems to me (this is what I take from Woodward) that a full defence of the thesis of CSQ would require a study of these analogies.

It might be felt—perhaps Woodward himself would feel this way—that the analogy will fail, because the end point (what we get to at stage three) is clearly more objective in the causal case than in the colour case. In the colour case, stage three is always a little half-hearted, in the sense that we recognize that the contingencies of our visual systems are never entirely eliminated: once we've noticed those contingencies, then there's no getting away from the fact that had we been different, we would have reached a 'different' stage three. Whereas for causation (it might be felt), there's only one possibility: one set of relations on which any creature capable of making the journey will inevitably converge. But I've argued that this is a mistake. There are ineliminable contingencies in the causal case, too—most strikingly those of temporal perspective, though these are merely the most stark manifestation of something deeper, and elsewhere much more familiar.

⁷ One concern about this formulation of the question might be that it doesn't adequately distinguish the situation of the individual from that of the community as a whole. It is not all clear that an individual language learner needs to 'come to' the objective view, rather than simply taking it to be the default. (In the latter case, we could read Sellars's story of John and the tie shop as telling us how we learn about subjectivity, having started from a presumption of objectivity.) So less contentiously, then, we could say that the general concern is simply to understand the relationship between 'objective' and 'subjective' viewpoints, in the cases in question.

5.1 'Not All Actions are Interventions'

These issues are also relevant to another objection that Woodward raises against the MP view. He points out that there are cases that an overly naive agency theory will be liable to get wrong:

As an illustration, consider a case in which an experimenter's administration of a drug to a treatment group (by inducing patients to ingest it) has a placebo effect that enhances recovery, even though the drug itself has no effect on recovery. There is a correlation between ingestion of the drug and recovery that persists under the experimenter's free act of administering the drug even though ingestion of the drug does not cause recovery. (Woodward 2013: §4)

Woodward then goes on to say that to deal with this problem we need the notion of what has come to be called an intervention—the problem, in effect, is that not all 'free actions' actually count as interventions, and it is the latter notion that matters, if we are to 'get the causal facts right'.

Examples like those just described show that if we wish to follow Menzies and Price in defending the claim that if an association between *A* and *B* persists when *A* is given the right sort of 'independent causal history' or is 'manipulated' in the right way, then *A* causes *B*, we need to be much more precise by what we mean by the quoted phrases. There have been a number of attempts to do this in the recent literature on causation. The basic idea that all of these discussions attempt to capture is that of a 'surgical' change in *A* which is of such a character that if any change occurs in *B*, it occurs only as a result of its causal connection, if any, to *A* and not in any other way. In other words, the change in *B*, if any, that is produced by the manipulation of *A* should be produced only via a causal route that goes through *A*. Manipulations or changes in the value of a variable that have the right sort of surgical features have come to be called interventions in the recent literature... The characterization of the notion of an intervention is rightly seen by many writers as central to the development of a plausible version of a manipulability theory. (2013: §5)

I think that Woodward is entirely right here, but that the point in no way requires that we abandon the basic thought of CSQ, that causation is analogous to a secondary quality, with agency substituted for sensory perception. For again, as Sellars (1956) teaches us, a similar dialectic exists in the case of colour, too. There, too, as Sellars's example of John and the tie shop illustrates so vividly, naive colour judgements come to be treated as provisional, and subject to revision. What John learns, in that example, is something important about how to revise his colour ascriptions (e.g. to take into account unusual lighting conditions). The upshot is that while our initial colour judgements are taken as *prima facie* reliable, they come to be embedded within a socially mediated practice that allows them to be revised—indeed, that's what it is for them to come to be genuine *judgements*, in Sellars's view.

This Sellarsian picture of revisable positive-presumptive judgement, based on our usually reliable abilities to track colours, seems to me to be strikingly analogous to the

picture that Woodward himself proposes with respect to agency and intervention. Woodward bases his proposal on ‘the following hypothesis’:

Human beings (and perhaps some animals) have (a) a default tendency to behave or reason as though they take their own voluntary actions to have the characteristics of interventions and (b) associated with this a strong tendency to take changes that temporally follow those interventions (presumably with a relatively short delay) as caused by them. (2007: 29)

He goes on to say that this hypothesis suggests a way in which ‘it is . . . possible for such subjects to use their interventions . . . to reach fairly reliable causal conclusions’ (2007: 29), provided two conditions are met:

First, subjects must have some way of determining (some signal that tells them) when they have performed a voluntary action and this signal must be somewhat reliable, at least in ordinary circumstances. Second, voluntary actions (again in ordinary, ecologically realistic circumstances) must—not always, but often enough—have the characteristics of an intervention. (2007: 29)

Woodward suggests that ‘both claims are true’. Concerning the second, he notes that ‘the correlation between voluntariness and satisfaction of the conditions for an intervention is imperfect’ (2007: 29).

In a badly designed clinical trial, an experimenter might be subconsciously influenced, in his decisions to give a drug to some patients and withhold it from others, by the health of the patients; his decisions are voluntary and yet correlated with an independent cause of recovery in a way that means that the conditions for an intervention are not satisfied. (2007: 29–30)

‘Nonetheless,’ Woodward says, ‘[i]t seems plausible that many voluntary actions do, as a matter of empirical fact, satisfy the conditions for an intervention’ (2007: 30).

If I come upon a wall switch in an unfamiliar house and find that there is a regular association between my flipping the position of the switch and whether a certain overhead light is on or off, then often enough my flippings will satisfy the conditions for an intervention on the position of the switch with respect to the state of the light . . . The existence of causal illusions in which we experience or ‘perceive’ salient changes that follow our voluntary actions as caused by them similarly suggests that such a heuristic is at work. (2007: 30)

Thus Woodward, like Sellars, offers us a story in which we reach our mature concept by learning to correct the deliverances of our naive judgements—judgements which ‘get things right’ not all the time, but often enough. If there is a difference between a Sellarsian version of the view of CSQ and Woodward’s view, I think it will lie in the conception of the order of explanation between the notion of intervention and that of agency—in the thought on Woodward’s part that his notion of intervention is somehow ‘more objective’ than anything that could be achieved by Sellarsian objectification, beginning with our (practical, not perceptual) experience of agency. However, I have argued that if there is such a difference, it counts against Woodward’s view: interventions are not a sufficiently natural category, and scepticism looms, if we

try to imagine that they are. So the more plausible approach will be the modified MP view, to the extent that there is a difference of this kind.⁸

6 Summary

I close by summarizing what I take to be right about MP's responses to the four objections they consider, and how I think these responses can be improved:

1. *Agency accounts confuse the epistemology of causation with its metaphysics.* The MP response stands, in my view, but the point can be strengthened by a version of the agency view that takes itself to be in the business of philosophical anthropology, not metaphysics.
2. *Agency accounts are vitiated by circularity.* Again, the MP response stands up in its own terms, in my view, and Woodward's accusation that it depends on excessive empiricism rests on a misreading. But again the response is greatly strengthened by an anthropological rather than a metaphysical conception of the project, for in this case there is no need to say that acquisition of the concept of causation depends in any sense on prior acquisition of a *concept* of agency, ostensibly defined or not. On the contrary, such a version of the view can help itself to Woodward's own account of the acquisition of causal concepts.
3. *An agency account cannot make sense of causal relations between events which are outside the control of any agent.* Here, too, MP seem able to allow Woodward to do the required work on their behalf. If Woodward seeks to establish a difference, based on the idea that his view is 'more objectivist', or extends causation into regions that an agency view can't reach, then he faces a major difficulty: scepticism looms. Hence my conclusion: either MP and Woodward are on the same side at this point, or his side is at a disadvantage, due to the threat of scepticism.
4. *Agency accounts make causation an unacceptably anthropocentric phenomenon.* Again, the original MP reply survives unscathed, in my view, in the sense that the analogy with colour does show that the anthropocentricity can be 'contained'—we are not committed to the view that ancient sunsets were colourless, or absurdities of that kind. Again, the anthropological stance makes this easier to say, because it focuses from the beginning on the concepts, which is where the anthropocentricity resides. However, I have argued that the concept of causation is *more* anthropocentric than either Woodward or MP themselves realize—there are more contingencies, more opportunities for variation, at least

⁸ Both sides agree that *within the practice*, a speaker must take there to be a fact of the matter—that is, to take her naive judgements to be subject to correction, subject to a norm that *makes it possible* for them to be right or wrong. The issue is whether we conceive of this practice on the model of Sellarsian objectification built on shared contingencies, or as something more metaphysically robust. I've argued that the latter view runs into trouble.

in principle. This may be surprising, but that's a feature, not a fault: the first-order anthropological investigation of our concept of causation reveals to us a contingency that isn't obvious 'from the inside'. Objecting that this makes causation unacceptably anthropocentric is like objecting that Copernicus makes our ordinary description of the heavens unacceptably anthropocentric.

Finally, I hope that these comments give some impression of the extent to which I feel that Woodward's work can be read as a magisterial vindication of the philosophical viewpoint whose colours Peter Menzies and I nailed to the mast in CSQ.⁹ I think that the arguments of that paper, updated as above, do have something to offer to Woodward, by way of a commentary on the task of locating his project on a bigger philosophical map. Even if accepted, however, this contribution does little to repay the debt that Menzies and I, and the agency tradition in general, owe to Woodward (and to Pearl, Spirtes, and others), for showing us how much can be done with the insight that causation is intimately linked to manipulation.¹⁰

References

- Collingwood, G. 1940. *An Essay in Metaphysics*. Oxford: Oxford University Press.
- Gasking, D. 1955. 'Causation and Recipes', *Mind*, 64: 479–87.
- Menzies, P., and Price, H. 1993. 'Causation as a Secondary Quality', *British Journal for the Philosophy of Science*, 44: 187–203.
- Perry, J. 1979. 'The Problem of the Essential Indexical', *Noûs*, 13: 3–21.
- Price, H. 1996. *Time's Arrow and Archimedes' Point*. New York: Oxford University Press.
- Price, H. 2007. 'Causal Perspectivalism', in H. Price and R. Corry (eds), *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press, 250–92.
- Price, H. 2011. *Naturalism Without Mirrors*. New York: Oxford University Press.
- Price, H., with Blackburn, S., Brandom, R., Horwich, P., and Williams, M. 2013. *Expressivism, Pragmatism and Representationalism*. Cambridge: Cambridge University Press.
- Ramsey, F. P. 1929 (1978). 'General Propositions and Causality', in D. H. Mellor (ed.), *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. London: Routledge and Kegan Paul, 133–51.
- Sellars, W. 1956. 'Empiricism and the Philosophy of Mind', in H. Feigl and M. Scriven (eds), *Minnesota Studies in the Philosophy of Science*, vol. I. Minneapolis: University of Minnesota Press, 253–329. Published separately in R. Brandom (ed.), *Empiricism and the Philosophy of Mind: With an Introduction by Richard Rorty and a Study Guide by Robert Brandom*. Cambridge, MA: Harvard University Press, 1997.
- von Wright, G. 1975. *Causality and Determinism*. New York: Columbia University Press.

⁹ It was nailing to the mast from my point of view, at any rate, though Peter may have felt that we were merely testing the waters!

¹⁰ Thanks to Helen Beebe, John Maier, Alejandro Pérez Carballo, and Jim Woodward for helpful comments and discussion.

98 CAUSATION, INTERVENTION, AND AGENCY

- Woodward, J. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.
- Woodward, J. 2007. 'Interventionist Theories of Causation in Psychological Perspective', in A. Gopnik and L. Schulz (eds), *Causal Learning: Psychology, Philosophy, and Computation*. New York: Oxford University Press, 19–36.
- Woodward, J. 2009. 'Agency and Interventionist Theories', in H. Beebe, C. Hitchcock, and P. Menzies (eds), *The Oxford Handbook of Causation*. Oxford: Oxford University Press, 234–63.
- Woodward, J. 2013. 'Causation and Manipulability', in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2013 edition), <https://plato.stanford.edu/archives/sum2013/entries/causation-mani/>.