A Functional Account of Causation

Or

A Defense of the Legitimacy of Causal Thinking by Reference to the Only Standard that Matters—Usefulness (as Opposed to Metaphysics or Agreement with Intuitive Judgment)

[×]

Abstract

This essay advocates a "functional" approach to causation and causal reasoning: these are to be understood in terms of the goals and purposes of causal thinking. This approach is distinguished from accounts based on metaphysical considerations or on reconstruction of "intuitions".

1. Introduction

In thinking about tonight's talk, I found myself considering two alternatives. One was a focused discussion of a single problem; the other something more general, ranging in a much less detailed way over a number of issues. Thinking that it would be difficult to find a single topic that would engage most people in the audience, that the hour is late, and that we are all tired, I've opted for generalities. What follows is a kind of overview/tasting menu of some issues having to do causal reasoning, long on pronouncements, expressions of attitude, and endorsements of general approaches, and short on details. For the latter, I can only suggest following up on some of the papers that I will mention.

2. Three Projects

The past several decades have seen an explosion of work on causation and causal reasoning, not just in philosophy, but also in many other disciplines including statistics, computer science, and psychology. Even within philosophy, many different projects having to do with causation have been pursued. A lot of this work falls into one of three categories. I will label these the *metaphysical project*, the *description of ordinary usage project*, and the *how does causation fit with physics project*, with no claim that these categories are either exhaustive or mutually exclusive.

First, the *metaphysical project* – exemplars include Armstrong, 1983, Bird 2005, and Tooley, 1977. Those pursuing this project think it important to provide a metaphysics for causal claims or to specify what causation "is", metaphysically speaking, or what the metaphysical "truth-makers" or "grounds" for causal claims are. Those

---

[×] This article is based very largely on the text of my talk at the 2012 Philosophy of Science Meetings, with small modifications. It was written for oral presentation and intended to be provocative. Hence many complications and qualifications that are usual in academic articles have been omitted. I've opted for preservation of the flavor of the original talk, rather than making any effort at precision.

committed to this enterprise sometimes invoke heavy-duty metaphysics involving relations of necessitation between universals, dispositional essences, and the like, although others prefer sparser, more "Humean" candidates for truth-makers- e.g., "laws" as understood within the Mill-Ramsey- Lewis framework).

Second, the *descriptive project*. Aspects of work by Lewis (e.g.1973), Hall, (2004) and Schaffer, 2000 and others appear to fall into this category, although these writers concurrently pursue metaphysical-sounding projects as well. Those pursuing the descriptive project attach considerable importance to constructing accounts whose aim is to describe or reproduce (what they take to be) the causal judgments of "ordinary folk" regarding various scenarios—Billy and Suzy throwing rocks at bottles and so on. Somewhat puzzlingly in view of this descriptive orientation, the focus is almost entirely on what people *say* about causal relationships in various situations. It does not extend to broader issues about how, as a descriptive matter, people learn about causal relationships, connect causal information to other sorts of concepts and cognitive activities, reflect such information in action and so on. However, one can certainly (as I would advocate) conceive of the descriptive project in this broader way, so that it becomes continuous with investigations into the empirical psychology of causal cognition, both in adult humans and others. Moreover, one might carry out this sort of descriptive study not just in connection with "ordinary folk" but also in connection with the causal cognition of scientists in various disciplines. (More on this below.)

The third project – the *fit with physics* project -- focuses on issues having to do with the relationship between causal claims, including the sorts of claims made in ordinary life and in the so-called special sciences, and what is imagined by some philosophers to be "fundamental physics". (cf. Field, 2003, Loewer, 2009.) These philosophers ask such questions as: what if anything in "fundamental physics" grounds or provides a basis for the causal claims made in everyday life and the special sciences?

Finally, many others (especially represented, perhaps, among philosophers of physics) view at least the first two of these enterprises—the metaphysical and descriptive projects -- with considerable skepticism, if not outright hostility. They dislike the metaphysics associated with the first project, and wonder why anyone should care about "what we would say" about whether the impact of Billy's rock caused the bottle to break and why this issue has any relevance to "real science", which of course they tend to think of as consisting largely if not entirely of physics. Those in this camp are often attracted to neo- Russellian "eliminativist" positions regarding causation and causal reasoning[1]. Eliminativists would respond to the question posed in "the fit with physics project": "what in fundamental physics grounds the causal claims of ordinary life and the special sciences? ", with the answer: nothing at all. Causation is nowhere (or perhaps only rarely and incidentally) to be found in fundamental physics and since fundamental physics contains all that is "real", causation is a confused or illegitimate or at least "ungrounded" concept. (Thus somewhat curiously, advocates of this position agree with the metaphysicians that it is really bad to be ungrounded, which perhaps suggests their views are less unmetaphysical than they may suppose.)

[1] Discussed but perhaps not fully endorsed, by several of the contributors to Price and Corry, 2007.

3. The Functional Project

I think of my own "interventionist" account of causation as not fitting very comfortably into any of the projects just described, although it has connections with some of them. Thus I would like to begin by sketching a distinct, alternative way of thinking about causation, causal reasoning, distinctions among casual concepts, how causal claims are connected to evidence and so on that I find attractive. (I will often use the phrase "causal cognition" as an umbrella term for this[2].) This alternative involves thinking of causal cognition in *functional* terms or, to put the idea a little bit differently, thinking of causal cognition from the perspective of what might be called *epistemic engineering*. More specifically, by a functional approach to causation, I have in mind an approach that takes as its point of departure the idea that causal information and reasoning is sometimes useful or functional in the sense of serving various goals and purposes that we have. It then proceeds by trying to understand and evaluate various forms of causal cognition in terms of how well they conduce to the achievement of these purposes. Causal cognition is thus seen as a kind of epistemic technology – as a tool -- and like other technologies judged in terms of how well it serves our goals and purposes.

Thinking about causation in this way has several other consequences. First, it suggests the importance of trying to *connect* causal notions to other concepts (and to reasoning patterns and procedures for investigating nature) that we employ. Understanding these connections turns out to play an important role in elucidating the point or usefulness of causal thinking and in evaluating various strategies of causal cognition. For example, from a functional perspective, it makes sense to try to understand how causal claims connect up with evidence and this in turn requires understanding the connections between causation and notions like statistical dependence/ independence (as in the Causal Markov condition) and with ideas from experimental design.

A second important consequence is that this way of thinking about causation leads naturally to a focus on *methodology*, broadly conceived—on *normative* assessment (and not just description) of various patterns of causal reasoning, of the usefulness of different causal concepts, and of procedures for relating causal claims to evidence. This normative or methodological dimension is another respect in which the functional project I recommend largely differs from the other three projects described above.

Finally, I suggest that to the extent that our project is functional in character, it is not obvious that it requires carrying out a "reduction" of causation or causal thinking to categories that are non-causal. As I will try to illustrate, the project of showing how various forms of causal thinking relate to our goals and purposes and to other sets of concepts can, to a considerable extent, be carried out in the absence of such a reduction.

To add to these rather abstract remarks with more specific detail, here is a partial list of issues and questions that naturally suggest themselves if we think about causal cognition (and causation itself) in functional terms.

---

[2] Use of this umbrella term is meant to signal that I see causation and how we think about it as intimately related

1) Why (or on what basis) do we (or should we) distinguish between cause and correlation? What is the point or points of such a distinction (that is, what goals does it serve)? (Section 4)

2) What distinctions might one usefully make *among* causal claims? That is, given various relationships that fall into the general category of "causal", what further discriminations among these might be fruitful? For example, as I have suggested elsewhere, biologists and others sometimes distinguish, within the general category of so-called type causal terms, claims that vary as to their *stability*, their *degree of specificity*, and the extent to which they satisfy a requirement of *proportionality* (cf. Woodward, 2010.) Within a functional framework, one can ask about the function or goals that served by these distinctions and whether these distinctions have defensible normative rationales. I comment briefly on this issue in connection with stability below. As another illustration, people often seem to distinguish among (a) causal claims describing relations of dependence involving so-called "production" (including but perhaps not limited to transference of energy and momentum, as when a rock strikes and shatters a bottle) and (b) causal claims involving dependence but not "production", as in the "double prevention" relations discussed below. Given goals plausibly associated with causal cognition, we may ask: what, if anything, is the basis or rationale for a distinction between production and other sorts of causal relationships? (Section 9).

3) What are the scope and limits of causal thinking? Under what circumstances and conditions is thinking causally useful or fruitful? Under what conditions, if any, is it not likely to be illuminating? Are there certain empirical conditions that a system must satisfy before it is profitable to try to analyze its behavior in causal terms?

What are the scope and limits of causal thinking? Under what circumstances and conditions is thinking causally useful or fruitful? Under what conditions, if any, is it not likely to be illuminating? Are there certain empirical conditions that a system must satisfy before it is profitable to try to analyze its behavior in causal terms?

4) Given some particular conception of the function of causal claims, what sorts of procedures for testing causal claims are reliable or warranted and what sorts of evidence is required to establish such claims? For example, given an interventionist account an obvious standard for causal inference from non-experimental evidence is this: are the evidence and background assumptions such that they support a conclusion about what the results of the hypothetical experiment associated with the causal claim would be, although we cannot actually perform the experiment? Although I lack the space for discussion here, some widely used inferential procedures, such as instrumental variables, are plausibly viewed as having a rationale grounded in an interventionist conception of causation in the sense that we can view them as furnishing information about the outcomes of hypothetical experiments without actually doing those experiments[3]. Other inferential procedures (such as assessing whether *X* causes *Y*

[3] For details see, e.g., Angrist and Pischke, 2009.

by regressing *Y* on *X*, controlling for every other variable the researcher can think of, demonstrably lack such a rationale.

5) How can a normative theory of causation based on a functional conception guide descriptive investigations into the forms of causal reasoning both ordinary folk and scientists employ? (Section 8)

6) Given goals associated with causal thinking, which variables should one "control for" or "hold fixed" in assessing causal claims? How can we understand such practices of control in terms of the goals of causal thinking? In particular, what should one hold fixed when relations of non-causal dependency like supervenience relationships are present? (Section 10)

4. Interventionism and the Goals of Causal Thinking

Assuming we want to pursue a functionalist project**,** what might be good candidates for goals or purposes associated with causal thinking? Interventionists think that the identification of relationships that are exploitable for purposes of manipulation and control is a central goal of causal thinking, but in principle a "functional" approach might evaluate causal thinking in terms of other goals as well—possibilities include the compact and unified representation of relationships useful for prediction, the codification of our commitments to various inductive strategies, as in Spohn 2012, or perhaps certain information-theoretic goals (implicit, arguably, in recent developments in machine learning approaches to causal inference such as Janzig et al., 2012.) I'll add that the general idea of thinking about key notions in philosophy of science, including, in addition to "causation", such notions as "explanation", "evidence", and "reduction" in functional terms seems to me to be a interesting and potentially worthwhile project in its own right, even if one does not find interventionism appealing as an account of causation.

I will give some more extended examples shortly of how the functional project might go in connection with interventionism, but a basic and familiar illustration is provided by the contrast between "mere correlation" and causation. What does this contrast consist in and why do (or should) we care about it? The (or at least an) answer provided by the interventionist framework is that only some and not all correlational relationships are potentially exploitable for purposes of manipulation and control; we regard those relationships that are so exploitable as causal and those not so exploitable as merely correlational.

As a concrete illustration, suppose you observe a correlation between attendance at a private school and scholastic achievement. You might wonder whether (first possibility) this is because attendance at a private school *causes* enhanced scholastic performance, in which case it might be an "effective strategy", in the sense of Cartwright (1979), for you send your child to a private school if you wish to improve his or her school performance. Alternatively (second possibility) the correlation might be entirely the result of some third factor which is a common cause of both attendance and performance. This would be the case if, for example, the same parental attitudes toward education that lead to selection of a private school also by themselves cause better scholastic performance among children. In this latter case it would be fruitless for

parents to send their child to a private school in order to enhance scholastic performance. The interventionist approach to causation takes the difference between these two possibilities (causation versus mere correlation) to be closely connected, in the way I have just described, to the question of whether or not it is true that intervening to alter whether or not one's child attends private school is a way of altering scholastic performance.

These sorts of considerations motivate what might be described as a minimalist interventionist account of causation, which I label (**M**) for future reference. This account is "minimalist" in the sense that it distinguishes between cause and correlation but is otherwise rather weak and uninformative—it says merely that there is some manipulation of the candidate cause variable *X* in some background circumstances under which the effect variable *Y* will change but says nothing about, e.g., exactly how manipulations of *X* will change *Y*:

(**M**) *X* causes *Y* iff (i) there is a possible intervention that changes the value of *X* which is such that (ii) under this intervention the value of *Y* changes.

Here an "intervention" is an idealized unconfounded manipulation of *X* that changes *Y*, if at all, only through *X*. More detailed discussion and distinctions among different kinds of interventions can be found for example, in Spirtes, Glymour and Scheines, (2000), Woodward (2003) and in section 10 below.

I think of (**M**) as a functional account in the sense that it connects the cause/mere correlation contrast to a goal that it is uncontroversial that we have— the goal of being able to intervene in the world and manipulate things. **M** is motivated by the idea that the distinction between cause and correlation makes sense because of the way it contributes to this goal; "cause" versus "correlation" is a useful distinction to have, given the goal. One obvious problem faced by through –going eliminativist treatments of causation is that it seems crazy to deny that the contrast between mere correlations and relationships that can be used for manipulation is a real contrast—consider, for example, the difference between the claim that ingestion of some drug causes recovery from an illness and the claim that ingestion and recovery are merely correlated. Even if one wants to consign the notion of cause to the rubbish heap, one needs some way of making sense of the distinction between mere correlation and relationships exploitable for manipulation and control.

Indeed, in many areas of science (perhaps especially but by no means exclusively in the social and behavioral sciences), there are extended controversies about whether relationships are causal (where "causal is understood in the sense of **M**) or "merely correlational", and elaborate modeling and statistical techniques are devised to decide such questions. For example, in connection with the issue raised above, Coleman and Hoffer, 1987 and Chubb and Moe, 1990 are two book length discussions of whether private school attendance is a means for boosting scholastic performance, both employing statistical and econometric arguments of some complexity. It is possible, I suppose, that all of this work is directed at a goal that is confused (or a pseudo-problem), but this is (at the very least) a conclusion that requires detailed argument of a sort that I have not seen from skeptics about the whole notion of causation.

5. How the Functional Project Differs from the Other Projects Described Above

Let me now describe in more detail how this functional framework for thinking about causation differs from the other possible approaches mentioned above. First, I take the functional framework to involve no particular metaphysical commitments beyond a very modest realism. This modest realism consists in the (I would have thought uncontroversial) assumption that the difference between those relations that are merely correlational and those that are causal has its source "out there" in the world (as philosophers like to say) and is not , say, somehow entirely the result of our "projecting" our beliefs and expectations onto the world with the result that some relationships look causal even though none "really" are[4]. Of course, it is a fact about us and our interests that we value information about relationships relevant to manipulation and control, but it is the world (and not just our interests) that determines which such relationships hold and in what circumstances. A similar point holds for the other features of causation (like stability) that I will relate to the functionalist framework below—they also reflect features of the world that constrain how we reason and how we can successfully act.

To expand a bit on the way in which the functional project differs from the metaphysical one: whatever the merits may be of the various metaphysical accounts of causation on offer, they seem to me to tell us very little about the function or goals of causal thinking or what sorts of causally related concepts and reasoning strategies serves those goals well or badly. Suppose an oracle tells you that causal relationships are relations of necessitation between universals or that such relationships are "grounded' in laws of nature, understood in terms of the Mill-Ramsey-Lewis account of laws. What guidance does this give one in answering most of questions 1- 6 above? Conversely, it seems to me one can provide such a functional account while being non-committal about the metaphysics of causation beyond the minimal realism just described.

6. Interventionism and Truth-Makers

In this connection, a number of writers have claimed that it is a major defect in interventionism that (at least as currently formulated) it provide no account (or no metaphysically acceptable account) of the "truth-makers" for causal claims or the "interventionist counterfactuals" with which they are associated. Although I will not try to fully address the issues surrounding this criticism here, the following remarks may help to clarify how I see matters. I fully agree (who would deny this?) that if it is the case that some relationship *R* to the effect that interventions on *X* are associated with changes in *Y* holds (e.g., private school attendance boosts scholastic performance) then of course we should expect that there will be some deeper explanation, perhaps to be found in some other, more fundamental science, for why *R* holds in the stable way that it

[4] Some may favor a very expansive conception of metaphysics according to which even this minimal realism (indeed any claim about what exists) amounts to a metaphysical commitment. But this unhelpfully makes every empirical claim a matter of "metaphysics."

does. In *this* sense of "grounds" or "truth-makers", there will be underlying grounds or truth-makers which explain the stable holding of *R*[5]. However, as far as I can see, these "truth makers" will be ordinary empirical facts and ordinary causal/nomological relationships and the sorts of explanations they provide for claims like *R* will be ordinary scientific explanations. Metaphysicians may go on to try to interpret these empirical facts in terms of special sorts of entities, relationships, and categories they favor (so that the "underlying" laws appealed to are understood in terms of necessitation relations or a Lewisian trade-off between simplicity and strength and so on) but I don't see why this sort of interpretation is *required,* at least for functional purposes. If one fails to provide such a metaphysical story one is not denying that causal claims have truth-makers in the "ordinary" sense described above, rather one is just declining to do metaphysics.

7. Normative Components of the Functional Project

What is the relationship between the functional approach to causation and the descriptive project of characterizing, as an empirical matter, the causal judgments people endorse and the reasoning that surrounds these, both in everyday life and in the various sciences? Here the connection to the functional project is closer, in ways I will briefly discuss, but the projects are still distinct. One very important difference, already alluded to, is that the functional project, as I conceive it, has a *normative* (as well as a descriptive) component. The functional project is normative in the sense that we are interested in whether and to what extent, various causal concepts or ways of reasoning about causation are relatively well-adapted or not, functional or not, with respect to our goals, rather than just in describing what people (whether ordinary folk or scientists) in fact do (although the latter enterprise is regarded as important too). In principle, we might discover that some common ways of thinking about causation or some patterns of causal reasoning are not very functional after all— that is, are not effective means for achieving our goals. This might happen in any one of a number of different ways: the reasoning might, for example, rest on mistaken empirical presuppositions, or it might turn out that, contrary to what many people think, some candidate causal concept, characterized in a certain way, has no or very few real world applications. Or perhaps the concept or the way in which it is applied blocks or undermines various goals we are trying to achieve rather than furthering them. Or it might turn out that the concept conflates features that dissociate logically or empirically and hence are features that should be distinguished. Or perhaps people assume that some kind of evidence or testing procedure is a good one for determining whether a causal relationship, conceived in a certain way (e.g. as a claim about the result of a hypothetical experiment), is present when in fact the procedure is demonstrably not a good one for that purpose. These are respects in which the normative

[5] I stress what needs to be explained is not just why *R* is true, but is why *R stable* or *invariant* (to the extent that it is) over interventions and other sorts of changes. Typically this is accomplished by showing not just that there are underlying laws and initial conditions from which *R* follows but also by showing *R* would hold continue to hold under some range of different initial conditions and/or in the presence of different underlying causal relationships. Example: Explaining the stable behavior of a gambling device by means of the method of arbitrary functions.

component of the functional project connects with methodology: with the investigation of what are good and not so good ways of learning about, reasoning with, and conceptualizing causal relationships.

As an illustration of the role of normative considerations, consider so-called Granger causality and the tests for Granger causality which are widely used in modeling and testing relationships between time series involving economic data. Oversimplifying considerably but I hope not fatally, Granger causation is the relationship that exists between two variables, *X* and *Y* when information about the value of *X* makes the values of *Y* more predictable, relative to some alternative in which information about *X* is absent (Granger, 1969). Simple examples show that *X* can Granger cause *Y* even though the relationship between *X* and *Y* does not satisfy condition (**M**)—that is, even though interventions on *X* will not change *Y*—a point that Granger himself recognizes Nonetheless it is not at all uncommon to read papers in macroeconomics and elsewhere in which the author moves seamlessly from evidence that *X* Granger causes *Y* to the conclusion that *X* is an interventionist cause of *Y* in the sense of (**M**), failing to recognize the difference between or conflating these two concepts, or at least failing to recognize that evidence that supports the a claim of Granger causation may not also be evidence for a relationship that conforms to **M** . The functional approach to causation does not rest content with simply describing these practices among econometricians but instead provides resources for asking critical questions about them.

A second illustration of this normative dimension is provided by the historical development of thinking about causation during the early modern period. On one reading, this moved from ways of thinking in which (from our present perspective) causal relationships are conflated with logical or conceptual relationships—a conflation which one finds, at least in some respects, in philosophers like Hobbes and Descartes -- to a view, which reaches its culmination in Hume, in which logical/conceptual relations and empirical causal relations are sharply separated (cf. Clatterbaugh, 1999.) I see this as a progressive, functional change, involving an *improvement* in our thinking about causation—an improvement in the sense that various conflations are removed, and the resulting notion of causation provides a better fit with methods of testing causal claims and with the goals of empirical science. Again, a functionalist approach would not confine itself to merely describing the possibly confused ways that people thought about causation in the past or may think about it at present. Instead it can suggest better ways of thinking.

As suggested above, issues of the sort that I have been describing fall into the general category of *methodology*, conceived of as a normative enterprise, in which we evaluate methods of testing and reasoning, and concepts we employ in terms of whether they help us to realize scientific goals. This methodological dimension is to a considerable extent absent from the alternative approaches to causal cognition described above. Methodology is a relatively neglected area in contemporary philosophy of science[6], although it (and in particular the methodology of causal inference) flourishes in other disciplines like statistics, artificial intelligence, machine learning and even cognitive psychology. One of the attractions of the functional way of thinking about

[6] In turn, methodology is a central concern of "general philosophy of science" and the waning of interest in the latter goes hand in hand with the neglect of the former.

causation is that it recognizes an important place for methodological considerations in thinking about causation, where methodology is conceived of as distinct both from metaphysics and mere descriptions of ordinary and scientific usage.

8. Preconditions for the Fruitful Application of Casual Analysis?

Next let me draw attention to another implication of the functionalist approach. Once one begins thinking about the extent to which various elements of causal cognition are well –adapated or not to our goals, one is struck by the possibility that some of these features may be functional in some contexts or with respect to certain kinds of problems—for example when certain empirical presuppositions are satisfied -- and yet not be functional with respect to other contexts or problems. After all, a tool or technology can be effective in connection with some problems but not with respect to others. This goes against the way that many philosophers are inclined to think about causation (and, for that matter, other concepts of philosophical interest). Many philosophers are inclined to think that causal concepts or thinking about nature in terms of causal relationships, if legitimate at all, ought to be legitimate everywhere: that if causation is a well-behaved concept, it ought to be universally applicable or describe some aspect of the fundamental furniture of the universe. (Its "cement", as Hume and Mackie have said.) This issue of universal applicability has particular resonance when we consider the role played by causal concepts in physics. Consider the question of whether there are "fundamental" physical theories that are not fruitfully interpreted in causal terms . One might wonder, for example, whether it is helpful to interpret the field equations of General Relativity as making a claim about a causal relationship between what is represented by the stress-energy tensor and the space-time metric or whether instead some other way of conceiving of that relationship is more appropriate.

The general issue of the role of causation in physics is a complex one that I will not try to address here. I do, however, want to advocate two ideas. The first (i) is that from the point of view of a functional approach to causation, it is entirely possible that there may be some contexts or domains of inquiry in which causal thinking and representation, or at least the kind of causal thinking associated with interventionism, is not useful or functional. One way in which this might happen is that the physics of the situation precludes, for deep, non-contingent reasons, the satisfaction of the conditions that must be met by interventions—perhaps the physics makes inapplicable any non-trivial notion of a local change in X which affects a second variable Y, only through this change in X[7]. Which situations are situations which are inhospitable to causal reasoning is something that must be decided on a case by case basis, but in principle I see no reason why "cause" might not turn out to be like "entropy" in the sense that it is a notion that it usefully applicable to physical situations with a certain structure when analyzed at a certain level of description, and not usefully applicable elsewhere. The second idea I want to advocate (ii) is that even if it is true that causal thinking has a little or no functional application in some domains, such as parts of fundamental physics, this is

[7] For example, if the stress-energy tensor cannot be specified independently of the space-time metric, the notion of an intervention on the former with respect to the latter becomes problematic.

consistent with causal thinking being highly useful in connection with many other problems and domains, in which the requirements for its fruitful application are met. In other words, if we decide that that causal notions have limited application in connection with General Relativity, it does not follow that they are illegitimate or confused when deployed in molecular biology or economics. "Doesn't work everywhere" does *not* imply "works nowhere". Instead, if causal notions have limits on their applicability of the sort I have been imagining, the research project, for functionalists, is to more precisely delimit the conditions under which causal thinking allows us to achieve our goals and when it does not.

9. Functionalism and the Descriptive Project

I said above that because of its evaluative dimension, the functional project is distinct from the project of just describing the causal judgments people make in ordinary life and in science. Nonetheless, I see the functional project as connected to the descriptive one in the following way: if we find, as a descriptive matter, that ordinary people or scientists make certain distinctions in their causal judgments or engage in certain patterns of causal reasoning and not others, it is often a good strategy to consider the possibility that there is some goal or point to what they are doing and to try to identify what that goal or point is. In other words, we should take seriously the possibility that people's causal cognition is often fairly well adapted to the problems they face or the goals they are pursuing. Of course, as I have said, this will by no means always be the case—sometimes we will find that, given goals and values people hold, some patterns of causal cognition that people exhibit don't further those goals at all. But often the opposite will be the case and when this is so, the descriptive and the normative will fit together nicely. Indeed, in many cases there are reciprocal connections between the normative and descriptive projects, with the results of the latter lending support to normative ideas and normative ideas in turn guiding descriptive research[8].

One illustration of how functional theorizing about causation with normative commitments can fit with experimental work about human causal reasoning in an illuminating way is provided by the experiments described in Lombrozo (2010). These explore the causal judgments of ordinary subjects regarding double prevention scenarios— scenarios in which in which an effect *e* is such that it will be prevented from occurring by some other event *f* unless a third event *c* occurs which prevents *f* from occurring, thus leading to the occurrence of *e*.

Such scenarios are sometimes illustrated in the philosophical literature with toy examples: Billy shoots down an enemy fighter which, had it not been shot down, would have shot down the bomber Suzy is piloting, preventing her from dropping bombs on her target. The question is then asked whether Billy's actions "cause" the bombing. On the one hand, the bombing counterfactually depends on Billy's action; on the other hand, the relation between the two lacks features possessed by many paradigmatic causal relationships—there is no transference of energy or momentum and Billy's action can be arbitrarily far away in space and time from the bombing, without any continuous set of intervening links between the two. If one is inclined to distinguish between "mere

[8] For more on this theme, see Woodward, 2012 and Forthcoming a.

dependence" and "production" (holding the latter to require transference or a connecting process or mechanism, which are marks of "genuine" causation), double prevention seems to involve dependence but not production.

One might well wonder whether questions about the causal status of double prevention relations have anything to do with any area of science. In fact, however, as a number of authors have noted, so-called causation by double prevention is quite common (and important) in many biological contexts—for example, many cases of genetic regulation work by double prevention, including the classic lac operon model of lactose synthesis devised by Jacob and Monod (Cf. Woodward, 2002). So understanding the causal significance of double prevention relations is of some importance in science.

Putting this aside, in her empirical investigations, Lombrozo finds (as one would expect) that not only do ordinary subjects distinguish between some double prevention relations and other sorts of casual relationships such as those involving production, but that, in addition, they make distinctions *among* double prevention relations with respect to their causal status. In particular, rather than judging that all double prevention relations are causal or that none are, subjects think that some are more paradigmatically causal than others; for example, they judge that double prevention relations which are the results of biological adaptations are more paradigmatically causal than other sorts of double prevention relations, such as those present in the Billy/Suzy scenario[9].

By itself this might seem to be an isolated curiosity, but Lombrozo proposes a functional explanation for *why* this pattern of judgment exists— she argues that double prevention relations can differ in their degree of *stability* and this is what explains the distinctions among double prevention relations that subjects make.

Stability is a notion I have discussed elsewhere (Woodward, 2006). In this context, we may understand it to mean something like the following: supposing *X* and *Y* are related in such a way that *Y* changes in value under some interventions on *X* (i.e., that the relationship between *X* and *Y* satisfies **M**), to what extent will this dependency relationship continue to hold as various other factors in the environment or background change"? To the extent that a relationship continues to hold under such changes, it is more stable. Both arm-chair philosophical reflection (e.g. Lewis, 1986) and more serious psychological investigation support the claim that, as an empirical matter, subjects judge that more stable relations of intervention-supporting counterfactual dependence are more paradigmatically causal than less stable intervention –supporting counterfactual dependence relationships. Stability thus has to do with a distinction we make among causal relationships, rather than with the distinction between cause and correlation(cf. Issue 2, Section 3).

Lombrozo contends that double prevention relations that result from natural selection, like those involved in genetic regulation, are typically more stable than double prevention relations like those present in examples like the Billy/Suzy scenario and that this difference explains why her subjects make the distinctions they do among double prevention relations. In particular, because many prototypical causal relations involving production are rather stable, our concern with stability helps to explain why we think of such relationships as paradigmatically causal and to distinguish between them and

[9] For more details about this example, see Woodward 2012 and Forthcoming a.

relatively unstable dependence relations not involving production, as in the Billy/Suzy example.

This explanation is a functional story in the following sense: It is unmysterious why both ordinary people and scientists should value causal relationships that are more stable—among other things, such relationships afford more extensive and reliable opportunities for manipulation and for prediction—and they are more generalizable or projectable to new situations People's judgments about double prevention cases makes sense or are rational given this concern with stability. In this sense, it is normatively or methodologically appropriate for both ordinary people and scientists to make the judgments they do.

Notice that in making this argument I have not tried to appeal to "intuitions" (my own or other people's) about whether double prevention relations are "really" causal or not. Instead my focus has been on understanding why, in functional terms, people judge as they do with respect to double prevention relations and with the normative basis, if any, for such judgments. People's judgments function as inputs to a functional analysis rather than as a source of special intellectual insight into the "nature" of causation.

10. The Causal Exclusion Problem

As a final example of thinking about causation within a functional framework, consider the so-called causal exclusion problem, the subject of a great deal of discussion in philosophy of mind and also, I think, lurking in the background of many recent discussions of so-called inter-level causation in philosophy of biology. Recall Kim's iconic diagram (Figure 1 below from Kim 1998), which I will use for heuristic purposes, despite its being misleading in many ways.

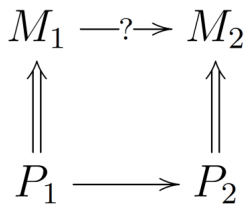

Figure 1

Here M_1 and M_2 are mental properties (or, more generally, whatever is represented by variables describing the "mental") and P_1 and P_2 their respective physical supervenience bases. Supervenience relationships are represented by a double-tailed arrows and ordinary causal relationships by single-tailed arrows. (Aside: there are many reasons for disliking the notion of supervenience but I ignore them in what follows, because they will not affect the points I want to make.) Assume for the sake of argument that the supervenience relations represented in this diagram are appropriate to some form of non-reductive physicalism: M_1 is not identical with (and is not caused by) P_1 and M_2 is not identical with (or caused by) P_2.

Suppose also that P_1 causes P_2, as represented by the single-tailed arrow from P_1 to P_2. The exclusion argument claims that these assumptions, in conjunction with several others which I will not go into, "exclude" the possibility that M_1 causes M_2. If so, mental

causation never happens. Moreover, if correct, the argument seems to generalize (although proponents of the argument sometimes deny this): It seems to follow that there are no causal relations involving "upper level" variables anywhere – in biology, economics, or, for that matter, thermodynamics. Any causal action must go on at a "lower" "physical" level. And if there is no (or little) causation in fundamental physics, perhaps it follows that there is no (or little) causation anywhere.

It will be instructive to reformulate this "argument" within an interventionist framework, which will allow us to focus on some key parts of the reasoning. Recall the notion of an intervention, already mentioned. Here we need a bit more detail: as characterized in Woodward, 2003 and elsewhere, an intervention on a variable *X* with respect to a second variable *Y* causes a change in the value of *X* which is such that any change in the value of *Y* occurs only through this change in the value of *X*, and not in some other way. In particular, the intervention *I* on *X* should not be such that it changes the value of *Y* via some route (in the causal graph characterizing the system of which *X* and *Y* are a part) that does not go through *X*. *I* also should not be such that it is correlated with any variable that affects *Y* via a route that does not go through *X*. Thus possibilities like the following are ruled out if *I* is to count as an intervention on *X*:

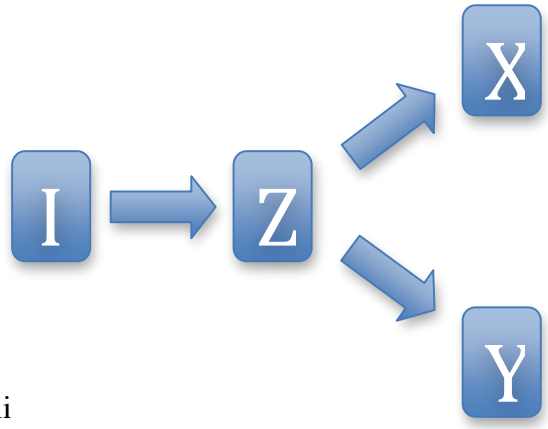

Figure ii

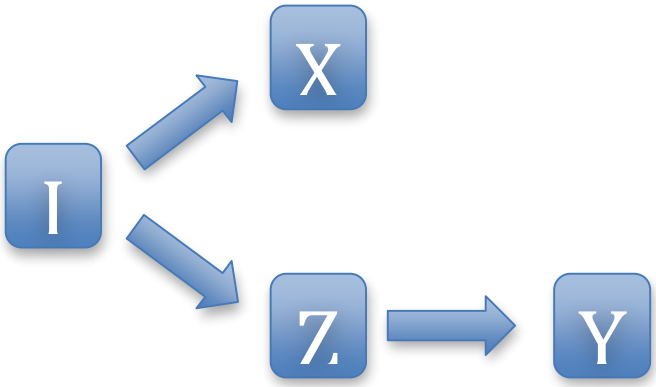

Figure (iii)

Summarizing, we might say that the notion of an intervention just described requires that (among other things) one control for *off-path variables*—that is, variables like *Z* in (ii) and (iii) that are on a causal path from *I* to *Y* that does not go through *X*.

The rationale for these requirements is (I hope) common-sensical (and "functional"). If *Y* changes under a manipulation of *X*, but this manipulation affects *Y* via a route that does not go through *X*, we have a badly controlled or ill-designed experiment for the purposes of determining whether *X* causes *Y*, an experiment that is confounded by *Z*. In such an experiment we have not ruled out the possibility that *X* does not cause *Y* and the reason why *Y* changes under the intervention on *X* is that this change in *Y* is caused by *Z*.

Now consider how these ideas might be applied to a diagram like Kim's. You may find it tempting to argue as follows: An intervention on M_1 for the purpose of determining whether M_1 causes M_2 requires that one control for variables that are not on the path from *I* to M_1 to M_2 and then ascertain whether, under some such intervention on M_1, M_2 changes. But P_1 and P_2 are such off-path variables and hence (so the argument goes) they need to be controlled for. But holding fixed P_1 and P_2, while intervening to change M_1 is of course impossible because of the nature of the supervenience relation—the supervenience of M_1 on P_1 excludes the possibility that M_1 might change while P_1 does not change. Hence M_1 does not cause M_2; it is causally inert with respect to M_2. All the real casual action represented in this diagram is in the $P_1 \rightarrow P_2$ relation, just as the exclusion argument claims. In effect, this is a version of the exclusion argument, restated within an interventionist framework.

Versions of this argument have been used recently by several writers (e.g., Baumgartner 2010) to contend that contrary to what I have always supposed, interventionism implies that mental (and presumably biological and economic etc.) variables are all causally inert; interventionism, on this view, requires that causal relationships, if they exist at all, hold only at some much "lower" level.

The first thing to say about this line of argument is that it misinterprets the technical requirements on an intervention as presented in Woodward, 2003. Those

requirements require control for variables that are off-path in the *causal graph* representing the causal structure of the system involving *X* and *Y*. A causal graph only represents casual relations and (by means of the absence of arrows) the absence of such relations; a diagram, like Kim's in which non-causal supervenience relations are present, is just not a causal graph in the relevant sense. In particular, P_1 and P_2 in Kim's diagram are *not* off-path variables in a causal graph representing the relationship among M_1, M_2, P_1, and P_2.

For this reason, the requirements on interventions in Woodward, 2003 (and also the very similar requirements in writers like Pearl, 2001 and Spirtes, Glymour and Scheines, 2000) do *not* require that one control for P_1 and P_2 in assessing whether M_1 causes M_2. Indeed, since these writers do not consider cases in which supervenience relations among variables are present, read literally, their requirements on interventions say nothing at all, one way or another, about which variables should be controlled for or how one should conceive of interventions in circumstances in which supervenience relations are present.

This consideration, however, does not address the following set of issues. Suppose that (for whatever reason) we do want to talk about causal relationships and interventions in circumstances in which supervenience (or other sorts of non-causal dependency relationships) are present. This raises the following question: Independently of what Woodward, 2003 may have said or implied, what variables *should* one should control for in such cases? Is it normatively correct to "control for" supervenience bases in assessing the causal efficacy of the variables on which they supervene? More generally (and expressed within the functionalist framework) the issue is how we might most reasonably (in light of our normative goals) extend the interventionist apparatus to cover cases in which supervenience relations are present.

Although I won't try to do this in detail here, in fact it is perfectly possible to extend the interventionist apparatus (and to characterize notions of intervention and of causal dependence) in a way that has the following features. First, (i) the characterization captures a notion of its being "possible" to intervene on M_1 that does not require changing M_1 while holding its supervenience base P_1 fixed. Instead, (ii) we characterize the notion of an intervention in such a way that interventions on M_1 automatically change P_1 in whatever way is required by the supervenience relationship between M_2 and M_1. Third, (iii) doing this has the result that M_1 comes out as causing M_2 on the standard interventionist treatment of causation (that is, M_2 changes under such an intervention on M_1 which is the standard test for whether M_1 causes M_2.). Thus within this extended framework, supervening variables come out as causally efficacious, contrary to what the exclusion argument claims. The details of how one can consistently do this are somewhat complex but can be found in Woodward, forthcoming b.

But although it is possible (without incoherence) to extend the interventionist apparatus in the way I have just described, one might wonder *why* one should to this—that is, what is the justification or rationale for extending the notion of intervention (and imposing requirements on what one should or should not control for) in the way I have described? Why isn't this just an ad hoc maneuver to save the causal efficacy of upper level causes? It is here that the functionalist perspective becomes particularly valuable.

To develop this perspective, let me return to the question of why we care about (what purpose/function is served by) controlling for off-path or confounding variables

like *Z* in ordinary causal graphs (without supervenience relations) like (ii) – (iii) above. In (ii)- (iii) *X* and *Y* are correlated when *I* is used to manipulate *X*, but this correlation arises for some other reason besides *X*'s causing *Y*—in particular, in this example, it arises because *I* is correlated with (in fact causes) some third variable *Z* which causes *Y* but is not on any causal path from *I* to *X* to *Y*. Because of this, when we come to other, new cases in which we also have an opportunity to manipulate *X* in the hope of controlling *Y*, it is entirely possible that when we perform this new manipulation *X* and *Y* will no longer be correlated in the way in which they are in the diagram (i). For example, this would happen if, in the new situation, the causal structure is as in (i) but the new intervention directly causes a change in *X* without the mediation of *Z*. In such a case, *Y* will not change under an intervention on *X*.

Thus if our goal is to determine whether the relationship between *X* and *Y* is such that we may reliably use it for manipulation, we need to distinguish between cases in which the correlation remains between *X* and *Y* when we intervene on *X* and cases in which it does not. This is what the requirement that we control for off-path variables in the characterization of an intervention achieves. In other words, the role of or rationale for such control for off-path variables is to enable us to distinguish cases in which a correlational relation can be reliably exploited for purposes of manipulation and those cases in which it cannot and to prevent us from being misled, by the presence of confounders, into thinking that a correlational relation can be so used when it cannot. In this sense there is an obvious functional justification for controlling for off –path or potentially confounding variables.

Now let us ask whether a similar rationale or functional justification extends to controlling for supervenience bases like P_1 in Kim's diagram if we wish to assess whether M_1 causes M_2. I take it that one has only to raise this question to see that the answer is "no". In the sort of case represented by Kim's diagram, the nature of the supervenience relation requires that M_1 cannot change independently of P_1, so there is no analogue to being misled into supposing that a merely correlational relation is exploitable for manipulation in the manner described above. To spell this out: Suppose one manipulates M_1 while failing to "control for" P_1, its supervenience base (so that P_1, also changes under this manipulation of M_1 in a way that respects the supervenience relation) and that one observes a change in M_2. Does one have to worry about the possibility because of this failure to control for P_1, one will be misled about whether there is a causal relationship between M_1 and M_2? That is, should one be concerned about the possibility that in some other context, in which a new manipulation I^* of M_1 occurs, P_1 might behave like *Z* in (i), changing or failing to change in such a way that there is no change in M_2? Given the nature of the supervenience relation, the answers to these questions must be "no". Whenever one manipulates M_1 to change its value in any way, the value of P_1 will automatically also change, so that one never finds oneself in a situation in which, because changes in M_2 are "really" caused by P_1 and M_1 is changed independently of P_1, the relationship between M_1 and M_2 disappears.

I take this to be just a statement of the common-sense idea that the rationale for controlling for "off path" variables in a structure consisting entirely of causal relationships, with no supervenience relations present, cannot be extended to or does not transfer to provide a rationale for controlling for supervenience bases in assessing the causal efficacy of supervening variables. In other words, P_1 is not a "confounder" for any

putative causal relationship between M_1 and M_2 in anything like the way in which Z is a confounder in testing whether X causes Y in (i). Indeed, treating P_1 as such a confounder is a highly non-functional and pointless thing to do in the sense that it defeats or undermines any attempt to find relationships among macroscopic variable that might be used for manipulation and control. Thus if we think in terms of the function of causal claims or in terms of interpreting such claims in a manner that allows them to be useful to us, demands that we "control for" supervenience bases in assessing the causal efficacy of upper level variables seem completely unmotivated—they derive from a failure to think through what such control is intended to accomplish[10].

A common response I have heard to the argument just presented is that regardless of whether it would produce anything useful for purposes of prediction and control, metaphysical considerations or perhaps "our concept" of causation *require* that one hold fixed supervenience bases in the manner demanded by the exclusion argument. Too bad if one wants a notion of causation that is useful for the special sciences or for dealing with the macroscopic world—it turns out that this is ruled out by the metaphysics/conceptual considerations embodied in the exclusion argument and there is nothing we can do about this. Now first of all, I think it is obvious that our concept of causation imposes no such requirement for control for supervenience bases in assessing the causal efficacy of macro-level variables. This is reflected in the complete failure to recognize any such requirement in the scientific literature, or in discussions of experimental design or causal inference outside of philosophy. But suppose it could be shown that our present concept of or way of thinking about causation imposes such a demand (or that certain metaphysical commitments require it) it. If so, I say: so much so much the worse for our present concept of causation and for the metaphysics in question. If our metaphysics or present patterns of thinking about causation don't serve our purposes (or even undercut them, as the exclusion argument in effect claims), we should replace these with concepts and patterns of thinking that better serve our purposes. The question is: Who's in charge here anyway? I say it is us (constrained of course by ordinary empirical facts) not metaphysics.

11. Conclusion

I have tried in these remarks to provide some illustrations of various ways in which thinking about causation and causal reasoning in functional terms can be illuminating. There is much more to this project than what I have been able to describe tonight. For example, I believe there is a story to be told within a functionalist framework that allows one to make sense of the notion of so-called downward causation

[10] Also relevant in this connection is the following observation (For details see Woodward Forthcoming b): It is perfectly possible for M_1 to supervene on P_1, M_2 to supervene on P_2 and for P_1 to cause P_2 (in the interventionist sense of cause captured by **M**), and yet for M_1 to *fail* to cause M_2. In this sense whether there is a causal relationship between M_1 and M_2 (and whether it is appropriate to draw an arrow from M_1 to M_2) is a "further fact" that cannot just be read off from the information in the lower part or "remainder" of Kim's diagram. Drawing or not drawing such an arrow thus conveys additional information and is in this sense "functional", rather than superfluous.

and also of structures involving causal cycles. Both of these are ubiquitous in biology and other disciplines such as economics, although they are often rejected as unintelligible by philosophers.

A commentators have expressed puzzlement over the larger project I was pursuing in *Making things Happen*. They have asked: Is interventionism a metaphysical doctrine? A purely descriptive account of the ordinary or scientific usage of words like "cause"? My answer is that is none of these: instead it is the functional project just described.

References

Armstong, D. 1983. *What Is a Law of Nature?* Cambridge: Cambridge University Press.

Baumgartner, M. 2010. "Interventionism and Epiphenomenalism." *Canadian Journal of Philosophy* 40: 359-383.

Bird, A., 2005, "The Dispositionalist Conception of Laws", *Foundations of Science* 10: 353–370.

Cartwright, N. 1979. "Causal Laws and Effective Strategies" *Noûs* 13:419–437.

Chubb, J and Moe, T. 1990. *Politics Markets and America Schools* Washington, DC: The Brookings Institution.

Clatterbaugh, K. 1999. *The Causation Debate in Modern Philosophy, 1637-1739*. New York: Routledge.

Coleman, J. and Hoffer, T. 1987. *Public and Private High Schools*. New York: Basic Books.

Collins, J. Hall, N., and Paul, L. 2004. *Causation and Counterfactuals*. Cambridge : MIT Press.

Field, H. 2003. "Causation in a Physical World" in M. Loux and D. Zimmerman, eds., *Oxford Handbook of Metaphysics*. Oxford: Oxford University Press.

Granger, C. 1969. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". *Econometrica* 37: 424–438

Hall, N. (2004): "Two Concepts of Causation", in Collins, Hall, and Paul. 2004, pp. 225–76.

Janzing, D., Mooij, J., Zhang, K., Lemeire, J., Zscheischler, J., Daniusis, D., Steudel, B. and Scholkopf, B. 2012."Information-geometric Approach to Inferring Causal Directions" *Artificial Intelligence* 182-183: 1-31.

Kim, J. 1998. *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation*. Cambridge, MA: MIT Press.

Lewis, D. 1973. "Causation." *Journal of Philosophy* 70: 556–67.

Lewis, D. (1986) Postscripts to "Causation" in *Philosophical Papers*, Vol 2. New York: Oxford University Press, pp. 172-213.

Loewer, B. 2009. "Why is There Anything Except Physics?"*Synthese* 170: 217-33.

Lombrozo, T. 2010. "Causal-explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions". *Cognitive Psychology*, 61, 303-332

Paul, L and Hall, N. 2013. *Causation: A User's Guide*. Oxford: Oxford University Press.

Angrist, J. and Pischke, J-S. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

Price, H. and Corry, R. (2007) *Causation, Physics, and the Constitution of Reality*. Oxford: Oxford University Press.

Schaffer, J. 2000. "Causation by Disconnection." *Philosophy of Science* 67:285-300.

Spirtes, P., Glymour, C. and Scheines, R. 2000. *Causation, Prediction and Search*, Cambridge: MIT Press.

Spohn, W. 2007. *The Laws of Belief*. Oxford: Oxford University Press.

Tooley, M. 1977. "The Nature of Laws." *Canadian Journal of Philosophy* 7: 667–69

Woodward, J. 2002. "What is a Mechanism? A Counterfactual Account." *Philosophy of Science* 69: S366-77.

Woodward, J. 2006. "Sensitive and Insenstive Causation." *Philosophical Review* 115: 1-50.

Woodward, J. 2010. "Causation in biology: Stability, specificity, and the choice of levels of explanation" *Biology and Philosophy* 25: 287-318.

Woodward, J. 2012. "Causation, Interactions Between Philosophical Theories and Psychological Research" *Philosophy of Science* 79: 961-972.

Woodward, J. (Forthcoming a) "Causal Reasoning: Philosophy and Experiment", forthcoming *in Oxford Studies in Experimental Philosoph*y.

Woodward, J. Forthcoming b. "Interventionism and Causal Exclusion." Available at http://Philsci-archive.pitt.edu.