

To appear in *Synthese*

## Methodology, Ontology, and Interventionism<sup>1</sup>

### 1. Introduction

This paper has several interrelated goals. First, it defends an interventionist account of causation by construing this account as a set of *methodological* proposals, rather than as a set of theses about the *ontology* or *metaphysics* of causation. Second (and relatedly), it attempts to use the topic of causation to raise some more general issues about the relation between, on the one hand, methodology, and, on the other hand, ontology and metaphysics, as these are understood in contemporary philosophical discussion, particularly among so-called analytic metaphysicians. I will use the topic of causation to argue for the importance and value of methodology, understood as an enterprise that can be pursued largely independently of one variety of ontology/metaphysics (what I call below ontology<sub>2</sub>/metaphysics<sub>2</sub>). But at the same time I will also suggest a way of bringing methodology and ontology/metaphysics closer together, by re-construing or reconfiguring some traditional issues about the ontology of causation as methodological proposals, following the approach recently advocated in Hitchcock, 2012. I suspect that this irenic suggestion about reconfiguration is unlikely to fully satisfy the metaphysically inclined, since it involves abandonment of some core commitments embraced by contemporary metaphysicians, but in my view, it has the advantage of replacing questions that seem unresolvable with more tractable alternatives.

The remainder of this essay is organized as follows: Section 2 describes what I mean by methodology and distinguishes two conceptions of ontology. Section 3 introduces the idea that interventionism should be understood as a set of methodological proposals, rather than as a claim about the ontology of causation. This is then followed in Section 4 with a brief overview of interventionism. Sections 5- 8 describe various ways in which in which interventionist ideas can be used, methodologically, to clarify the content of causal claims and their relation to evidence. Section 9 compares the elucidation of causal claims provided by the interventionist program of associating such claims with hypothetical experiments with more standard metaphysically motivated demands that causal claims be elucidated by providing “truth conditions”. Section 10 concludes by considering the prospects of re-construing issues about the ontology/metaphysics of causation as issues in methodology.

### 2. “Methodology” and two conceptions of “Ontology”

As a point of departure, we need some common understanding of key terms such as “methodology” and “ontology”. I take methodology to have to do with how *we ought* (both the “we”—that is, we humans-- and the “ought” are important here) to go about

---

<sup>1</sup> A version of this paper was given as a talk at a conference on “Causation: New Prospects” at the College de France in December, 2013. I am grateful to the participants in the conference and especially Chris Hitchcock, Huw Price, and Claudine Tiercelin for helpful comments.

investigating, learning, and reasoning about various aspects of nature, about what sorts of theories we should construct, and about how we should reason about various important concepts in the scientific enterprise (such as “cause”). I will also assume that we should think about methodology within a means/ ends framework: we have certain ends or goals (cognitive or otherwise) and we then evaluate various methodological proposals in terms of how well they conduce or contribute to such ends. Methodological proposals are thus thought of as hypothetical imperatives, with their normative justification turning on how effectively they serve as means to our goals. Moreover, since effective achievement of our goals is what matters, methodology must take into account human epistemic limitations— for example, limitations on what we can observe, manipulate, and calculate.

Thinking about methodology in this way, we may note that methodological proposals vary a great deal in their level of abstractness and generality. On the one hand we have very general recommendations like “Construct falsifiable theories”. On the other hand, proposals about method may also be far more specific – in connection with the methodology of causal inference, for example, they may involve techniques for extracting causal information from time series data or claims to the effect that only randomized experiments are reliable ways of finding out about causal relationships<sup>2</sup>.

What about “ontology”? On one understanding of this notion found (when the word is used at all) in some areas of science, “ontology” simply refers to what are taken to be the most basic entities or properties or structures in some area inquiry or to the most useful or perspicuous way of classifying or conceptualizing these. (Here “basic” need not be thought of as carrying any metaphysical commitments; it instead it has to do with which assumptions or starting points or classifications are best from the point of view of

---

<sup>2</sup> An anonymous referee asks how methodology, as I conceive it, relates to epistemology. Since I am not an epistemologist, I hesitate to make claims about this matter, but insofar as the methodology of causal inference has to do with how one *learns about* or comes to *know* about causal claims, I suppose one might, if one wishes, think of *this portion* of methodology as in part a branch or subspecies of epistemology. However, as I try to make clear, methodology as I conceive it, also has to do with issues that at least some readers may not think of as squarely within the province of epistemology –e.g., issues having to do with conceptual clarification, with discovery and search strategies, and with the evaluation of reasoning patterns. (If one wants to think of these issues as part of epistemology as well, that is fine with me.) I will add that , as I think of it, methodology is not particularly concerned with many issues on which contemporary analytic epistemologists focus such as the explication of such ordinary language epistemic notions as “X knows that p” or the construction of responses to skepticism. In addition, methodology is, often at least, domain specific in the sense that it operates with the assumption that the methods that are appropriate for learning about or reasoning with facts in one domain of inquiry may not carry over into investigations of other domains—for example, methods for testing non-causal statistical hypotheses are not identical with methods for causal inquiry and so on. So methodology often requires some degree of subject matter specific knowledge—it is not “theory of knowledge”, conceived of as general, domain-independent inquiry. I will also add that in my opinion the legitimacy of methodology as a subject of philosophical interest does not require that it be subsumed under or connected to the concerns of analytic epistemology.

scientific theory construction, allowing for the possibility that different answers to this question may be appropriate depending up one's goals and the level of analysis sought.) Let us call this *ontology*<sub>1</sub>. It is ontology in the sense of *ontology*<sub>1</sub> that geneticists have in mind when they speak of constructing a "gene ontology" or that cognitive neuroscientists have in mind when they speak of the importance of constructing a "cognitive ontology". For example, in the latter case, a cognitive ontology might provide a catalog or typology of basic cognitive processes and operations of a sort that can be used to answer questions about how the operations performed by different neural regions or circuits should be characterized at a cognitive level (does amygdala activity involve "fear processing" of some sort or more general reward processing of positive and negative stimuli?), whether different neural regions should be thought of as performing the same or different cognitive operations, whether various experimental tasks involve the same or different cognitive processes and so on. On this understanding, constructing an "ontology" (that is, an *ontology*<sub>1</sub>) is a matter of ordinary empirical or scientific investigation and the entities in such an ontology are not taken to have any sort of special "metaphysical" status (whatever that might be) – they are just familiar objects of inquiry of the sort studied in the various sciences. If by ontology we mean *ontology*<sub>1</sub>, it seems completely uncontroversial that methodology and ontology are and ought to be closely linked. The reason for this is the truism that our methods for investigating particular scientific domains should be attuned to the entities and structures those domains contain and that different sorts of investigative and reasoning methods may be fruitful for different sorts of entities and structures, depending on the features of the latter. For example, if the correct cognitive neuro-ontology is that the basic structures or units of analysis in the brain are distributed networks of various sorts, then different methods for identifying and reasoning about these will be appropriate than if one thinks that the basic units are highly localized neural areas. As another example, if the *ontology*<sub>1</sub> of some domain is that it contains structures in which values of key variables change over time in a way that is causally influenced by previous values of those variables and complex feedback relationships are present, generating data in the form of time series, such domains will likely require different methods of causal analysis than structures which are acyclic and can be assumed to have settled into some sort of equilibrium state which generates cross sectional data.

In contemporary philosophical discussion, however, "ontology" and "metaphysics" (in so far as this is connected to ontology) typically seem to be understood in a rather differently from the notion of *ontology*<sub>1</sub> just described and in a way that makes connections with "methodology" much less straightforward. I will call these enterprises *ontology*<sub>2</sub>/*metaphysics*<sub>2</sub>. Although I am unable to provide a general characterization of when a project counts as *ontological*<sub>2</sub>/*metaphysical*<sub>2</sub>, there are certain diagnostic markers: including a tendency to focus on certain distinctive questions and to go about attempting to answer them in certain distinctive ways as well as the use of a distinctive terminology and (in some cases) a tendency to postulate special sorts of entities that seem different from more ordinary objects of scientific investigation. For example, *ontologists*<sub>2</sub> often focus on questions like the following: what are the "truth-makers" or "grounds" for causal claims (laws of nature?, powers and dispositions?, relations of necessitation between universals?) According to many *ontologists*<sub>2</sub>, failure to provide such truth-makers leaves causal claims unclear or problematic or at least involves a

dereliction of philosophical duty of some kind. Another FAQ (by ontologists<sub>2</sub>) has to do with the “relata” of the causal relationship—are these events?, processes?, tropes? Yet another issue is whether causal claims be “reduced” to claims that are non-causal – such as “Humean regularities”. Use of words such as “Fundamental”, “Reality” and “Ground” (especially when capitalized) and inquiries organized around these concepts are also often indicative that one is in the presence of an ontological<sub>2</sub> inquiry, as when a writer asks what “Causal Reality fundamentally consists in” or “Whether causation is one of the Fundamental Constituents of Reality”. Appeals to “intuition” or supposed intuition-like sources of information (even if the word “intuition” is not used and terms like “judgment” are substituted) to provide constraints on the results of investigation are also often indicative of an ontological<sub>2</sub> project, as when it is suggested that our intuitions suggest that causal relationships are “binary” and “intrinsic” relationships between events (cf. Paul and Hall, 2013). Finally, another important feature of ontology<sub>2</sub> is that the results of its inquiries are not supposed to be influenced by “merely pragmatic” factors having to do with human goals, interests or epistemic constraints—instead, insofar as there is a goal to such inquiry it something like the description of the most fundamental aspects of reality. Since, as I conceive of it of it, methodology is heavily influenced by such pragmatic factors, this is one important reason why it is more independent of ontology<sub>2</sub> /metaphysics<sub>2</sub>, than some suppose.

I concede that the distinction between ontology<sub>1</sub> and ontology<sub>2</sub> is not always sharp or obvious. (Indeed, it is hard to avoid the uncharitable suspicion that practitioners of the latter sometimes have an interest in deliberately blurring the distinction, in an effort to make their enterprise look more continuous with ordinary empirical science.) But the absence of a sharp dividing line does not mean that there are no clear cases. Papers appearing in *Science*, *Nature* or the *Physical Review*, even when they report the discovery of novel “entities” or structures such as the Higgs boson (and which thus might be regarded, if one wishes, as contributions to ontology<sub>1</sub>), are very, very different in terms of content, argumentation, and evidence appealed to than what is found in, say, Tooley 1987 or Paul and Hall, 2013 (much less Sider, 2011). If standard philosophical accounts of theory-testing, evidence, explanation and so on recognize no sharp difference between these enterprises (as is sometimes claimed by contemporary metaphysicians<sup>3</sup>) that reflects the inadequacy of those standard philosophical accounts rather than indicating that ontology<sub>2</sub> is just ordinary science, pursued at a higher level of abstraction.

### **3. Interventionism as Methodology, not Ontology<sub>2</sub>**

---

<sup>3</sup> See, e.g., Sider 2011 and Paul, 2012, both of whom claim that ontological<sub>2</sub> inquiry relies on the same methods as those allegedly employed in ordinary scientific research, such as appeals to “simplicity” and “inference to the best explanation”. Although I lack the space for discussion, I think that the ease with which these “methods” can be marshaled in defense of ontology<sub>2</sub> is a reflection of the fact they do not adequately characterize what is distinctive about various forms of science. Real science relies on much more specific and constraining strategies for assessing evidential and explanatory import.

I said above that methodology requires a specification of ends or goals we are trying to achieve in inquiry. In turning now to methodological issues having specifically to do with causal reasoning, it is thus crucial that we begin with a consideration of the goals or purposes distinctively associated with the causal concepts and reasoning strategies. Interventionists like me (cf. Woodward, 2003) think that among the goals distinctively associated with causal reasoning is the discovery of relationships that are exploitable for purposes of manipulation and control and the discussion that follows will be framed around this idea. However, there are certainly other possible candidates for goals associated with causal thinking—these might include, for example, finding compact and unified (or “strong” and “simple”, in the sense of the Mill-Ramsey-Lewis account of laws) representations of correlational relations. In principle, one might equally well apply a means/ends framework to justify methodological principles connected to the discovery of representations having this sort of feature.

The conception of methodology advocated in section 2 is a broad and expansive one. In connection with causal reasoning, it certainly includes, for example, algorithms for extracting causal information from passive observational data of the sort described in Spirtes, Glymour and Scheines, 2000, but it includes much more besides. I list immediately below some issues that arise in connection with causal reasoning that I see as methodological in nature or that can be fruitfully interpreted as issues in methodology. Where appropriate, I have tried to bring out the difference between these questions, interpreted methodologically, and questions of a more ontological<sub>2</sub>/metaphysical<sub>2</sub> nature, thus suggesting how the latter might be reconstrued in terms of the former.

- On what grounds, if any, is it justifiable to distinguish between cause and correlation—that is, what is the methodological justification for drawing this distinction and, given our methodological goals, how or on what basis *should* we draw it? A more specific version of this question is the following: Given a regression equation describing a correlation (3.1)  $Y = bX + U$  under what conditions does this have a “causal interpretation” – that is, what conditions must be satisfied before we are warranted in interpreting (3.1) as correctly claiming that  $X$  causes  $Y$  in accord with the quantitative relationship in (3.1)? This is a very frequently asked and fundamental question in statistics and econometrics. Of course one might also interpret this question in a more “ontological<sub>2</sub>/metaphysical<sub>2</sub>” way, as asking whether, ontologically speaking, causation “just is” correlation or whether fundamental reality contains only correlations or whether instead it contains causal relationships understood as different from or over and above mere correlation. However, one can also construe the question methodologically, as I have above—that is, as a question (or questions) about what goal or purpose is served by drawing the cause/correlation distinction in the way that we do, about what, if anything, would be lost if we did not make this distinction about the sorts of empirical tests might be used to distinguish true causal claims from mere correlations and so on.
- A more general but related question is this: what features must be present in a relationship for us to justifiably judge that it is causal? For example, must a “connecting process” between cause and effect be present? Are so-called double

prevention relations (in which  $e$  would occur in the absence of  $d$ ,  $d$  would prevent  $e$  if  $d$  were to occur, and the occurrence of  $c$  would prevent  $d$ , thus permitting the occurrence of  $e$ ) relationships in which  $c$  causes  $e$ ? Again, it is common to think of this as a question in the ontology<sub>2</sub>/metaphysics<sub>2</sub> of causation but one might also think of the question as a normative/methodological one: given our methodological goals, what rationale, if any, is there for insisting on the presence of a connecting process if a relationship is to count as causal or for distinguishing dependence relations in which a connecting process is present from those in which it is not? What, if anything, would be lost or missed if we regarded all relations of non-backtracking dependence as causal, whether or not a connecting process is present?

- What are some of the features that a causal claim should possess for it to be clear and unambiguous? How can we clarify causal claims that are unclear and make them more precise? This can be regarded as a kind of “semantic” project, but, as we shall see, it seems quite different from the more ontological<sub>2</sub>/metaphysical<sub>2</sub> project of trying to specify the semantics or meaning of causal claims by providing reductive truth conditions.
- What sorts of procedures for testing causal claims are reliable or warranted and what sorts of evidence is required to establish such claims? How does inference to causal conclusions from evidence and other assumptions work? When is it not possible, even in principle, to answer certain causal questions on the basis of a given body of evidence?

This list is very far from being exhaustive: the methodology of causal reasoning includes much more besides these issues (see Woodward, forthcoming) but it will be more than enough to occupy us in what follows<sup>4</sup>.

#### 4. Basic Ideas of Interventionism

With this as background, I turn now to some brief remarks introducing interventionism and then to some particulars concerning its role as a methodological thesis. The basic idea is that causal relationships are relationships that are potentially exploitable for manipulation and that, as a matter of methodology, it is illuminating and fruitful to associate causal claims with claims about the outcomes of hypothetical experiments – that is, as claims about what would happen to a candidate effect variable if

---

<sup>4</sup> This is perhaps the appropriate place to address a possible misunderstanding that has surfaced when I have presented this material orally. I do not mean to claim that appeal to an interventionist framework is the *only* way of addressing methodological issues like those described above. I claim only that this framework provides *a* fruitful way of thinking about methodology. I do not consider it to be an objection to a methodological framework that there may be other frameworks that lead to similar or additional (but consistent) conclusions.

one to intervene on a candidate cause variable. Here is a very simple version of an interventionist characterization of one particular notion of “cause”, where  $X$  and  $Y$  are understood as variables:

(M)  $X$  causes  $Y$  iff (i) it is possible to intervene on  $X$  and (ii) under some such possible intervention on  $X$ , changes in the value of  $X$  are associated changes in the value of  $Y$ .

The causal notion characterized here is a type-level notion that I have elsewhere called *total causation*. I have separated the rhs of (M) into two distinct clauses in order to make it explicit that (M) involves (i) a commitment to the *possibility* of intervening on  $X$  as well as (ii) a claim about what would happen to  $Y$  under such an intervention. I do not have a fully satisfactory account of what “possible” means in (i) but will assume in what follows that a necessary condition for intervention to qualify as possible is that be logically or conceptually possible.

Heuristically, an *intervention* on  $X$  with respect to  $Y$  causes a change in the value of  $X$  which is such that the value of  $Y$  changes if at all via a route (or routes) that goes through  $X$  and not in some other way. In the characterization of this notion in Woodward, 2003, interventions were taken to be (4.1) *hard* or *arrow-breaking* – that is, it was assumed that the effect of an intervention on some variable  $X$  is that the value of  $X$  comes entirely under the control of the intervention variable (it becomes a function of the intervention alone, and not of any other variable), so that all other endogenous causal influences on  $X$  are “broken”.

(4.1) is convenient for some purposes, but it is often methodologically fruitful to relax this requirement, as has been shown in recent work by Eberhardt (e.g., Eberhardt and Scheines, 2007). In particular, we may generalize the notion of an intervention so that:

(4.1\*) Interventions can be “soft” and non-arrow-breaking, in the sense that they just supply an appropriately exogenous and uncorrelated source of variation to the variable  $X$  intervened on, rather than breaking all other causal influences on  $X$ . Here “appropriately uncorrelated” means that the variation supplied by the intervention  $I$  should not be correlated with other causes of  $X$  or with causes of  $Y$  besides those that are on the route from  $I$  to  $X$  to  $Y$ .

On this conception, what is crucial to the notion of an intervention, is that the variation supplied to the variable  $X$  intervened on be such that it affect the candidate effect  $Y$  only through  $X$  and not via some other route. This condition can be met by soft as well as by hard interventions. We will make use of modification (4.1\*) below.

## **5. The Methodological Fruitfulness of Construing Causal Claims as Claims about the Outcomes of Hypothetical Experiments: Distinguishing the Target**

With this as background, I turn to some criticisms that have been advanced against interventionism. I believe that insofar as these criticisms have force, they

depend in large measure on construing interventionism as an ontological<sub>2</sub> / metaphysical<sub>2</sub> thesis. The criticisms are less telling if instead one thinks of interventionism as a methodological proposal.

The first criticism (cf. Strevens 2007, Baumgartner 2009, perhaps Glymour 2004) is that **(M)** is viciously “circular” because it claims to elucidate the notion of *X*’s causing *Y* by appealing to a notion (that of an intervention) which is obviously itself causal in character. An adequate account of causation, the critics claim, must be non-circular—it must be “reductive” in the sense that it explicates what it is for *X* to cause *Y* in terms of concepts (like “regularity”, “correlation” and so on) that are themselves entirely non-causal. Often this is put in terms of the complaint that **(M)** fails to provide “truth conditions” or “grounds” for causal claims, where the assumption is that these must take a reductive form. (cf. Hiddleston, 2005, Reutlinger, 2012)

The second worry is this: when an experimental manipulation of *X* can actually be carried out, it is perhaps plausible (the critic says) that **(M)** provides *one* way of *discovering* whether *X* causes *Y*. However, even in this case, the critic claims, this criterion is (at best) of purely epistemological significance. It doesn’t tell us anything about the ontology<sub>2</sub>/ metaphysics<sub>2</sub> (or semantics) of causation-- what causal claims mean or what causation “is” or anything like that. Moreover (the critic continues) when an appropriate experimental manipulation of *X* is not in fact (or cannot) be performed, it is even less clear how **(M)** could possibly be illuminating: **(M)** connects “*X* causes *Y*” to a counterfactual about what would happen if an intervention on *X* were to occur, but how can that counterfactual be of any use if we can’t carry out the intervention in question? **(M)** (the critic concludes) may describe a test that can sometimes (but by no means always) be used to determine whether *X* causes *Y*— do an experiment— but has no significance beyond this.

My discussion in Woodward, 2003 makes it clear that interventionism does not furnish a reductive theory of causation. I’m also willing to stipulate that interventionism has little or nothing to say about the ontology<sub>2</sub> of causation. But I don’t agree that interventionism is for this reason uninteresting or unilluminating. It may be ontologically<sub>2</sub> unilluminating, but I contend that it is *methodologically* illuminating.

To motivate this claim, I begin with an empirical observation that reports a frequently made normative claim: researchers in a number of different disciplines claim that it useful or illuminating to connect causal claims and the results of hypothetical experiments in something close to the way described by **(M)**. For example, the *potential response* framework developed by Rubin, Holland and others (e.g., Rubin, 1974, Holland, 1986) and now widely used in statistics, econometrics (see references below), and elsewhere in social science is organized around construing causal claims in just this manner. Moreover, researchers adopting this approach are quite explicit that they regard it as illuminating even when the experiment associated with a causal claim is merely “hypothetical” or “possible”— that is, when the experiment is not in fact carried out and even when, although the experiment is in some relevant sense possible, there are barriers of various sorts to actually carrying out, so that it remains (in this sense) “hypothetical”. Indeed, one prominent recent discussion (King et al., 1994) goes so far as to claim that in those



areas of the social sciences in which it is difficult to carry out actual experiments, the notion of experiment is “useful primarily in understanding non- experimental design” (p. 125).

Of course critics of interventionism may respond that these researchers are simply confused or suffering from some sort of false consciousness—they think that this connection with purely hypothetical experiments is useful and illuminating when it is not—but I would advocate a more charitable approach in which we try to understand, at both a methodological and perhaps a psychological level, how it is possible for this connection to be informative. A key idea which will serve as organizing principle in what follows is this: an interventionist conception of causation can be methodologically illuminating in virtue of its role in characterizing the *target* to which we are trying to infer when make causal inferences in non-experimental contexts: Roughly, in non-experimental contexts we are trying to infer what the results of a hypothetical experiment would be without doing the experiment. Construing the inferential task in this way has distinctive implications for the causal questions we should ask and the sort of information that is relevant to answering them.

To illustrate the basic idea as it appears in the econometrics literature, I draw on some remarks about causal inference from a recent influential econometrics text, Angrist and Pischke, 2009, although I might equally well have drawn on any one of a number of other sources. The first remark occurs in the context of a discussion of what it means to talk of the “causal effect” of different hospital treatment regimes on patient recovery:

Causality means different things to different people, but researchers working in many disciplines have found it useful to think of causal relationships in terms of ... potential outcomes [which] describe what would happen to a given individual in a hypothetical comparison of alternative hospitalization scenarios. Differences in these potential outcomes were said to be the causal effect of different hospitalization regimes on patient recovery. (2009, p.52)

Elsewhere Angrist and Pischke make it clear that their “hypothetical comparisons” are to be understood in terms of possible experiments, although not necessarily experiments that are actually carried out:

We hope to find natural or quasi-experiments that mimic a randomized trial by changing the variable of interest while other factors are kept balanced. Can we always find a convincing natural experiment? Of course not. Nevertheless, we take the position that a notional randomized trial is our *benchmark*. (2009, p 21, my italics)

In another context, speaking of an inference from non-experimental data concerning the causal effect of education on income, Angrist and Pischke quote the following remark from Ashenfelter (1991) a prominent researcher in this area:

How convincing is the evidence linking education and income? Here is my answer: Pretty convincing. If I had to bet on what an ideal experiment would indicate, I bet that it would show that better educated workers earn more (2009, p. 21)

My gloss on these remarks is the following: Angrist and Pischke (and Ashenfelter) take the true causal effect of education on income (or hospitalization regime on recovery) to be what would be revealed in an ideal hypothetical experiment in which education (hospitalization regime) is manipulated (by an “intervention-like change, as in a randomized experiment) and the associated change in income observed. This serves as a *benchmark* for what the true causal effect is – it is the *target* they are trying to discover or the *standard* for whether the inference is reliable. In other words, inferences to a causal effect from non-experiment data are seen as correct or reliable to the extent that such inferences tell us what the result of an appropriate hypothetical experiment would be. If some candidate inference procedure employed on non-experimental data delivers certain causal conclusions and a properly performed experiment is brought to bear on the same inference problem, with results that are different from those in the non-experimental inference procedure, it is the results of the experiment that provide the standard for correctness and the results of the non-experimental procedure that should be rejected as mistaken. These ideas about the role of hypothetical experiments in providing a normative standard are most naturally viewed as methodological claims rather than ontological<sub>2</sub> claims.

Angrist and Pischke also emphasize another way in which hypothetical experiments can figure as a regulative and clarifying ideal in causal inference:

If you can't devise an experiment that answers your question in a world where anything goes, then the odds of generating useful with a modest budget and nonexperimental survey data seem pretty slim. The description of an ideal experiment also helps you formulate causal questions precisely. The mechanics of an ideal experiment highlight the forces you'd like to manipulate and the factors you'd like to hold constant. (2009, p.5)

The methodological ideas in these passages are ideas that are naturally suggested by an interventionist construal of causal claims. In particular, we may interpret Angrist and Pischke as suggesting that associating causal claims with hypothetical experiments can help to:

- (i) Pick out the target information we are trying to discover when we engage in causal inquiry (the outcome of a hypothetical experiment) and in doing this also help us to clarify the original causal claim or make it more precise.
- (ii) Show that certain causal questions we may be tempted to propose are not answerable, at least with available data– not answerable either because they do not correspond to any possible experiment or because the actually available data cannot provide answers to questions about what outcome of the hypothetical experiment would be.

- (iii) Clarify and evaluate some of the methods used to infer to causal conclusions, particularly in the case of non-experimental data. Very roughly, the idea is that we ask whether the data are such that (in conjunction with appropriate other assumptions) they can be used to infer what the results of the associated hypothetical experiment would be if we were to perform the experiment, although in fact we don't or can't actually perform the experiment

To get a more concrete sense of how these ideas work, it will be helpful to return to the issue of what is required for a regression equation to have a causal interpretation. Suppose, as above, we have two correlated variables,  $X$  and  $Y$ , and that we regress  $Y$  on  $X$ , yielding a regression equation of form (5.1)  $Y = bX + U$  where  $U$  is a so-called error term and the coefficient  $b \neq 0$ , since  $Y$  and  $X$  are correlated. Recall that regression is a mechanical procedure that can be applied to any body of data in which  $X$  and  $Y$  are correlated (even if  $X$  does not cause  $Y$  or vice-versa) to produce an equation of form (5.1) describing that correlation. Angrist and Pischke take the question of when (5.1) can be given a causal interpretation to be equivalent to the following: "When can we think of a regression coefficient as approximating the causal effect that might be revealed in an experiment?" (2009, p. 51) In other words, their idea is that (5.1) (correctly) describes a causal relationship between  $X$  and  $Y$  if (5.1) correctly describes how the value of  $Y$  would respond to changes in the value of  $X$  in a hypothetical ideal experiment in which  $X$  is manipulated— thus illustrating the characteristic interventionist idea that the standard for judging whether (5.1) describes a causal relationship involves reference to an associated hypothetical experiment.

I emphasize for future reference that the question Angrist and Pischke are addressing is naturally understood as an "interpretive" question in the sense that it has to do with what it *means* (in one perfectly good sense of "means") for  $X$  to cause  $Y$  or what one is *committed to* in claiming that  $X$  causes  $Y$ . But their answer to this question does not take the form of providing truth conditions or grounds in anything like the form that is likely to satisfy metaphysicians<sup>5</sup>.

It might seem, paralleling the more general objection to interventionism described earlier, that there is an obvious objection to Angrist's and Pischke's suggestion about when a regression equation has a causal interpretation— it is circular in a way that

---

<sup>5</sup> Or at least this is so to the extent that the demand for truth conditions or grounds is taken, as it commonly is by metaphysicians, to be the demand for an account of claims like (5.1) in terms of what is taken to be ontologically fundamental or primitive and some such. The category of what is ontologically fundamental or non-derivative may or may not make sense, but I assume that no one thinks that claims about hypothetical experiments are ontologically fundamental in the relevant sense. Thus Angrist and Pischke fail to provide truth conditions for causal claims in the metaphysician's sense of truth conditions. Put slightly differently, if one accepts Angrist's and Pischke's gloss on (5.1) as illuminating, then what this suggests is that one can explicate or clarify the content of causal claims while failing to provide truth conditions in the metaphysician's sense. The paragraphs that follow in the main text attempt to explain how this is possible.

renders it completely unilluminating. After all (it might be said) to know what the result of manipulating  $X$  in an idealized experiment, don't I have to already know whether  $X$  causes  $Y$  (and, further, what the quantitative causal relationship, if any, between  $X$  and  $Y$  is)? If so, how can it possibly be illuminating to invoke "what the outcome of a hypothetical experiment on  $X$  would be"?

There are a number of things that might be said in response to this worry, but, as a point of departure, let me note that one thing the invocation of hypothetical experiments does in this context does is captured by (i) above: this invocation directs attention to one possible goal or target or body of information one is trying to discover in contrast to others. In particular, there are (as the quotation above from p. 52 of Angrist and Pischke reminds us) many different possible accounts or proposals about what conditions a relationship has to satisfy to count as causal. For example, as already noted, some writers hold that a "connecting process" between cause and effect is required. Other writers hold that a necessary and sufficient condition for  $X$  to cause  $Y$  is that  $X$  and  $Y$  be correlated when one controls for (conditions on) all other available information (or perhaps all information temporally prior to  $Y$ ) – indeed this is the implicit assumption behind a fair amount of social science practice, and behind many philosophical theories of "probabilistic causation", as well as tests for so-called Granger causation in econometrics. Still others hold (roughly) that a necessary and sufficient condition for  $X$  to cause  $Y$  is that  $X$  is a non-redundant part of a condition  $K$  that is nomically sufficient for the occurrence of  $Y$  where "nomic sufficiency" means there is a fundamental law linking  $K$  to  $Y$  (cf. Paul and Hall, 2013, pp. 14-15). One of the things that **(M)** and the interpretation of causal claims in terms of hypothetical experiments does is to provide a criterion for (or proposal about) causation that aims to *distinguish* causal relationships from these other sorts of relationships. Although this is more obvious in the case of some of these alternative proposals than others, **M** differs extensionally from each of the above alternatives in the relationships it regards as causal. For example, double prevention relations obviously can satisfy **M** even if no connecting process is present. And although I lack the space for detailed discussion, it is arguable that the relationship between correlated measurement outcomes in an EPR-style experiment involving two particles in a singlet state can satisfy the non-redundant nomic sufficiency requirement above, but fail to satisfy **M**, since one cannot use the measurement outcome on one wing of the experiment to manipulate the outcome on the other—this is not a relationship that even in principle is exploitable for manipulation. Notice also the important point that **M** can play this sort of distinguishing or target specifying role without being "reductive" in the sense of providing a translation of causal claims into claims that are entirely non-causal in content. Even if we agree that an adequate ontology<sub>2</sub> of causation must be reductive, it does not follow that to play the methodological role of target described above, **M** must be reductive.

## **6. The Methodological Fruitfulness of Construing Causal Claims as Claims about the Outcomes of Hypothetical Experiments: Clarifying the Content of Causal Claims**

Another respect in which the association of causal claims with the outcomes of hypothetical experiments can be illuminating (and which can help to make apparent how

a non-reductive account like **M** can help to clarify the content of causal claims) is this: Causal claims are often advanced in forms that are unclear or indeterminate or underspecified—this is particularly likely to be the case when they are represented in a simple “*Cs* cause *Es*” format, as in “smoking causes lung cancer”. Associating causal claims with hypothetical experiments in the manner described in (**M**) helps to make such claims more determinate, clear, and precise—it does so by making it explicit that they are to be understood in terms of one particular hypothetical experiment (which we specify) rather than another such experiment. Among other things, making the hypothetical experiment associated with a causal claim explicit requires that we be more precise than we often are about which variables are being claimed to be causally related and what their possible values are. In particular, within an interventionist framework, a precise characterization of a causal claim requires that we specify at least two possible values for the cause variable (since we think in terms of an intervention changing one of these values to the other) and then indicate how changing the cause variable from one of these values to others leads to changes in the value of the effect variable.

As an illustration, consider the claim (cf. Glymour, 1986) that

(6.1) Smoking five packs of cigarettes a day causes a substantial increase in the probability of lung cancer.

One (uncharitable) way of associating this with a hypothetical experiment is to interpret (6.1) as claiming

(6.1\*) Any intervention that changes whether someone smokes five packs a day to some smaller number of packs (e.g. 4.9 packs) will cause a substantial change in the probability that person develops lung cancer.

Another, more charitable interpretation of (6.1)—probably closer to what is likely to be intended by someone asserting (6.1) -- is to interpret it as claiming that

(6.1\*\*) An intervention that changes whether someone smokes five packs to that person not smoking at all causes a substantial change in the probability of that person developing lung cancer.

(6.1\*) and (6.1\*\*) are different claims about the outcomes of different hypothetical experiments. It is likely that they have different truth values: (6.1\*) is likely false and (6.1\*\*) is likely true. Someone asserting (6.1) can thus clarify what is meant by indicating which of (6.1\*) or (6.1\*\*) is the intended interpretation.

As a second illustration, consider the claim

(6.2) Being a woman causes one to be discriminated against in hiring.

Interventionists are inclined to regard (6.2) as unclear and to think that it can be made clearer or disambiguated by making it explicit just which claim about the outcome of a hypothetical experiment is intended. From an interventionist perspective, the basic problem with (6.2) as it stands is that the notion of a manipulation of or intervention on

“being a woman” or “gender is unclear” – there are a number of different things that might be meant by this claim.

One possible way of manipulating gender is to change an individual’s sex chromosomes immediately after conception (substituting an X chromosome for a Y or vice-versa<sup>6</sup>.) Interpreted with this particular intervention in mind, (6.2) might be understood as claiming that

(6.2\*) Changing an individual sex chromosomes so as to alter their gender immediately after birth would change that person’s probability of being hired for certain jobs.

A second, alternative construal of (6.2) which I would guess comes closer to capturing what most of those asserting intend to claim is this:

(6.2\*\*) Intervening to change an employer’s beliefs about the gender of an applicant will change that person’s probability of being hired.

One way of bringing out the difference between (6.2 \*) and (6.2\*\*) is to note that (6.2\*) would be true under a regime in which hiring is based entirely on the applicant’s merit and qualifications as long as different genders develop different interests and abilities that cause them to be differentially qualified for various jobs. If what is intended by (6.2) is a claim to the effect that hiring decisions involving women are not made on the basis of the applicant’s merit and qualifications and that qualified women are not hired because of their gender, (6.2\*) does not capture this. By contrast, (6.2\*\*) comes closer to capturing what is presumably intended by (6.2). Note also that in the case of (6.2\*\*), the variable which is viewed as the target of the intervention (and the cause) is “employer beliefs about gender” rather than gender itself. This illustrates how the interventionist framework forces one to be more precise about which variables are the intended causal relata. I emphasize that the framework accomplishes this by providing a *criterion* for when a candidate variable in a causal claim is such as to make the claim unclear or in need of clarification—this will happen when there is no unambiguous well-defined notion of manipulation associated with the variable in question or when there are different things that might be meant by a manipulation of the variable in question and different outcomes of such manipulations. Thus the interventionist framework yields the result that “gender” is probably not a good variable (for the reasons described above) and that causal claims concerning the effects of gender can be disambiguated by associating them with different claims about the outcomes of the manipulation of different candidate cause variables.<sup>7</sup>

---

<sup>6</sup> Some may claim that such a manipulation would change the identity of the individual involved. This claim plays no role in the argument that follows.

<sup>7</sup> An anonymous referee suggests that my remarks here amount to nothing more than the triviality that “you should be precise about which variables are the intended causal relata” and that this is something “you can endorse while denying interventionism in all of its forms.” I have tried to make it clear that interventionism goes beyond this triviality in providing at least some guidance about which variables are appropriate variables for

(6.2 \*\*) is a claim that might be (and in fact has been tested) by, for example, submitting otherwise identical resumes in which only the gender of job applicants has been altered. By contrast, quite different data would be required to determine whether (6.2\*) is true. In any case, the important point for our purposes is that (6.2\*) and (6.2\*\*) are non-equivalent claims which may well have different truth-values. It does not seem controversial that it would be worthwhile for someone asserting (6.2) to think through which of these possibilities he or she has in mind.

These examples provide illustrations of how associating causal claims with hypothetical experiments in the manner of (M) can force one to be more precise and explicit about what causal claims commit us to and how they might be tested, thus making sense of the fact, noted above, that researchers in many disciplines find this sort of association useful. Again, note that this sort of precisification (or clarification of meaning) does not proceed by providing a reduction of causal claims to non-causal claims, but can be methodologically useful despite this—the clarification comes from the specification of additional commitments and structure associated with the original claim, rather than by providing a reduction. We may think of these examples (and others discussed in this essay) as illustrations (in connection with causation) of that portion of methodology that has to do with the clarification and critique of concepts.

Part of the answer, then, to how interpreting causal claims as claims about the outcomes of hypothetical experiments can be illuminating despite its apparent “circularity” has to do with the kinds of considerations described in this section and the previous one. In addition, however, I believe that there is an aspect of this issue that has to do with the empirical psychology of human cognition and the ways in which humans “store” information. In particular, when one initially entertains a causal claim like “*X* causes *Y*” it is often the case that not all of the implications of this claim or all of the information relevant to testing it is explicitly integrated into one’s causal judgment. That is, one can entertain the claim that “*X* causes *Y*” without thinking, at least very clearly and explicitly, about just what would be involved in changing or manipulating *X* or how one expects *Y* to change under various possible manipulations of *X*. Nonetheless, the latter information may be information that one “possesses” in some sense, perhaps merely tacitly or implicitly, or at least it may be information that one might readily obtain

---

causal relata, when variables should be regarded as defective or “imprecise”, and so on. I have tried to emphasize that this guidance is at least suggested by the basic commitments of interventionism. By way of contrast, the restrictions described above (that there should be a well-defined notion of manipulation associated with each variable) does not follow as naturally from various other accounts of causation. For example, it seems perfectly possible for a Lewisian to introduce a “miracle” that changes someone’s gender, gender can be an INUS condition for some outcome etc. I will also add that in general I do not think that it is good enough in formulating causal claims to be “precise” about which variables are involved. This is partly because one needs to be “precise” in the right way but also because some choices of variables are (even if these can be made precise) defective from the point of view of formulating causal claims. There are a number of other plausible criteria governing variable choice besides the one mentioned in the text above, some of which are discussed (in some cases with an interventionist motivation) in Woodward, forthcoming b.

via further reasoning or empirical investigation. Associating “ $X$  causes  $Y$ ” with a hypothetical experiment forces one to explicitly incorporate this information into one’s causal judgment. There thus can be a sense that one has genuinely learned something new by representing this association<sup>8</sup>.

### **7. More on The Methodological Fruitfulness of Construing Causal Claims as Claims about the Outcomes of Hypothetical Experiments: Unanswerable Causal Questions**

I suggested above that in addition to clarifying and disambiguating causal claims, associating causal claims with hypothetical experiments can also help us to see that certain causal queries are unanswerable, either in principle or with available data—they are unanswerable because the data cannot be used to tell us what the outcome of the associated hypothetical experiment would be. As an illustration<sup>9</sup>, consider the following causal question: does starting school at a later age cause children to do better in school—e.g., does starting school at age seven rather than six lead to better educational outcomes, in the sense that children learn more effectively from a year in school (as first graders) at age seven than at age six? This is a prima-facie sensible question which, if answerable, would have important policy implications. We might try to make the question more precise and operational by framing it as the question of whether a later age of starting school for a child leads to higher test scores at some later time. This in turn suggests a hypothetical experiment: randomly select some children to start first grade at age seven and others to start first grade at age six, expose them to a year of first grade schooling, and then compare their test scores at the end of first grade. However, thinking in terms of this hypothetical experiment draws attention to an obvious problem: in addition to whatever effect school starting age has on test scores, there is also a maturation effect on test scores—children who start school at age seven will on average do better on tests simply because they are older, independently of any effect of start age on school achievement. This may suggest that what we need to do is to somehow hold fixed or correct for this maturation effect so that we can assess whether, independently of this, start age has an effect on achievement—that the relevant hypothetical experiment is one in which the effects of maturation are controlled for, while start age is manipulated, all children receive a year of schooling, and the effect on achievement is then observed. A moment’s thought, however, should suggest that this hypothetical experiment is impossible, at least as long as we take maturation to be measured just by chronological age. The problem is that it is a mathematical identity that start age ( $S$ ) = chronological age ( $M$ ) – time in school ( $T$ ). We are supposing that  $T$  is fixed by the experiment at one

---

<sup>8</sup> Thus in thinking of (**M**) in this way, we are not construing it as a purely descriptive claim about what people always or usually have immediately in mind when they make causal claims— we are not claiming that they always associate causal claims with hypothetical experiments and so on. On the contrary, we recognize that they often do not spontaneously make this association, but argue that they *ought* to, as a way of making causal claims more clear and precise.

<sup>9</sup> Again drawn from Angrist and Pischke, 2009, pp. 6-7.



year, and hence one cannot in addition hold  $M$  fixed while changing  $S$ . Thus, it appears that there is no way even in principle of doing an experiment which would disentangle the effects of maturation and of age of starting school on achievement, at least if we are confined to the data we have been considering and have no other way of measuring the relevant variables. If causal claims must be interpretable as claims about the outcome of possible experiments, we know that no inferential procedure, however sophisticated, will answer the causal question with which we began. In this case, interventionism shows its methodologically usefulness by clarifying what can be learned about causal relationships from certain sorts of data. In particular, the *impossibility* of performing a certain experiment turns out to be methodologically informative.

## 8. The Methodological Relevance of Interventionism to Causal Discovery

I want now to shift to a different topic, which I think also illustrates the methodological usefulness of interventionism. As noted above, **M** allows for the possibility that one can learn about causal relationships from many different sources, including passive observations not involving interventions. In cases in which one learns that  $C$  causes  $E$  on the basis of passive observations, what **M** implies is that one should think of oneself as having learned that  $E$  would change under some intervention on  $C$ , but without actually performing the intervention in question. Contrary to what one might initially suppose, this idea has genuine methodological bite because it suggests that if one's evidence is not sufficient to establish such claims about what would happen to  $E$  under interventions on  $C$ , one's evidence is not sufficient to establish that  $C$  causes  $E$ . In particular, it suggests a general strategy of evaluating proposed causal inferences on the basis of whether they provide evidence that allows conclusions to be reliably drawn about the outcomes of appropriate hypothetical experiments. (Recall the remarks from Angrist and Pischke above in which exactly this evaluative strategy is advocated.)

Certain procedures for inferring causal conclusions such as the use of instrumental variables or regression discontinuity designs are readily justifiable on the basis of the consideration just described, while other procedures are not. I will briefly illustrate this by reference to instrumental variables. Suppose that the problem we face is to estimate  $b$  in a context in which the data generating process is represented by (8.1)  $Y = bX + U$ ,  $X$  and  $Y$  are our candidates for cause and effect variables, respectively, and  $U$  is an error term that is *correlated* with  $X$ . Assume also that the context is one in which the equation is intended to have a causal interpretation. As is well known, when  $X$  and  $U$  are correlated, we cannot use the simplest possible estimator for  $b$ , involving ordinary least squares. We can, however, reliably estimate  $b$  if we can find an instrumental variable  $Z$  for  $X$  with respect to  $Y$ . Such an instrument  $Z$  is a variable that (i) is associated with  $X$ , (ii) is independent of  $U$ , and (iii) is independent of  $Y$  given  $X$  and  $U$ . When these conditions are satisfied, it is widely accepted that the use of the instrumental variable  $Z$  allows us to estimate  $b$  and leads to reliable causal inferences. (See, e.g., Angrist and Pischke, 2009). In particular, if  $Z$  is an instrument for  $X$  with respect to  $Y$  meeting the above conditions, then  $b^* = \text{cov}(Y, Z) / \text{cov}(X, Z)$  is an estimator for  $b$ , interpreted as the *causal effect* of  $X$  on  $Y$ . Thinking of causal claims within an interventionist framework helps us to understand why under conditions (i)- (iii), use of an instrumental variable for estimating a causal effect is a reliable procedure: under these

conditions,  $Z$  functions in a way that is intervention-like in its relation to  $X$  with respect to  $Y$ . In particular,  $Z$  is in effect a “soft intervention” in the sense described in section 4. The underlying intuition is this: under the above conditions, any variation in  $Z$  which is associated with  $Y$  must be due to the causal influence of  $X$  on  $Y$ , rather than being due to some other source, such as a confounding variable, which is precisely the condition for  $Z$ 's being a soft intervention. Thus the covariation between  $Y$  and  $Z$  (normalised by  $Cov(X,Z)$ ) provides information about what would happen to  $Y$  were an intervention on  $X$  to be performed, which is just the information required for causal interpretation according to the interventionist.

Other causal inference procedures such as regression discontinuity designs can be evaluated in a similar way, also by asking whether they provide reliable information about the results of a hypothetical experiment. It is thus mistaken to claim, as critics have sometimes asserted (e.g. Russo, 2012), that interventionist accounts have nothing to say relevant to the methodological assessment of testing or inference procedures for causal claims in non-experimental contexts.

## **9. The Relation Between Methodology and Ontology<sub>2</sub> Revisited: The Role of Truth Conditions**

Despite the remarks above, I suspect that many metaphysicians will continue to believe that ontology<sub>2</sub>/metaphysics<sub>2</sub> has an important role to play in methodology. After all, it may be said, virtually by definition, ontology<sub>2</sub>/metaphysics<sub>2</sub> concerns what is most “fundamental” or what “grounds” everything else. How could this fail to be relevant to methodology? One answer is simply that what is most “fundamental” from the point of view of metaphysics may not be what is most important or useful from the point of view of methodology. In support of this assessment, consider the very large literature outside of philosophy (in statistics, econometrics, epidemiology etc.), referenced in passing above, that pursues methodological issues concerning causation, apparently very successfully, but also apparently without addressing ontological<sub>2</sub>/metaphysical<sub>2</sub> issues at all. Of this course it might be responded that this literature is misguided (or could be improved by appropriately incorporating ontological<sub>2</sub>/metaphysical<sub>2</sub> considerations), but this is a claim that needs to be argued for in detail, rather than simply asserted.

To expand on this point, consider how various ontological<sub>2</sub> claims about causation in the philosophical literature might be brought to bear on the methodological issues described above. Suppose, for example, that a completely reliable oracle tells you that causal claims are grounded in relations of necessitation between universals or involve “mutual manifestation” relations among powers (cf. Mumford, 2009). How exactly does this help with such projects as clarifying unclear causal claims, specifying reliable procedures for inferring causal conclusions from non-experimental data, clarifying when a regression equation has a causal interpretation, and so on? Asking these questions suggests that these ontological<sub>2</sub> projects seem to be addressed to very different issues than the methodological issues described above.

This point seems to me to hold even for ontological<sub>2</sub> projects that attempt to make more direct contact with current science such as the common suggestion that the “truth conditions” for causal claims are to be found in fundamental laws of physics that “underlie” or “ground” those claims. Suppose, for the sake of argument, that this is

defensible as a piece of metaphysics. Again we may ask whether this helps with methodological problems, at least as these arise in areas of science that are concerned with causal reasoning outside of fundamental physics. There are at least two general grounds for skepticism. First, in the case of most of the methodological problems described above, no one has any idea how the relevant underlying laws (even if they are known) might be brought to bear. Suppose that one is interested in the causal effect of education on income and is looking for an instrumental variable that might be used to estimate this effect (or, more generally, is simply interested in understanding the rationale for employing an instrumental variable). No one has a clue how to “translate” variables like “education” and “income” into the language of fundamental physics or what physical laws link these translated variables or how to specify an appropriate instrumental variable in this language. Certainly researchers interested in the causal effect of education on income don’t proceed by looking for such underlying truth-makers. Of course one can insist that if education has a causal effect on income, the appropriate underlying laws linking physical specifications of these variables must be “there” and that if these two variables are merely correlated no such law will be “there”, in this way attempting to capture the difference between causation and correlation in this case. But, as nearly as I can see, this contention, even if correct, is completely unhelpful from the point of view of methodology, given our ignorance of the underlying physics associated with this causal claim. A similar point seems to apply in connection with the other methodological issues discussed above— for example, how can facts underlying physical laws help to clarify the content of unclear or ambiguous causal claims, given the epistemic unavailability of these underlying facts?

It may seem tempting to suppose that, although in many cases it is difficult or impossible to obtain information about the truth-makers (fundamental laws etc.) underlying upper level causal claims, it is nonetheless true that if we somehow had such information, it would be a good scientific strategy to make use of it to “ground” procedures having to do with testing and reasoning about upper level causal claims. In other words, as an “in principle” matter, if we had information about underlying details, couldn’t we use it to provide a really secure and reliable foundation for upper level causal claims? In fact, even this is far from obvious. As Wilson (forthcoming) has emphasized, in areas of science outside of fundamental physics, a common key to successful inquiry is to make the reliability of theorizing and reasoning as *independent* (epistemically and methodologically) as possible of underlying fundamental physical details, precisely because these are so epistemically inaccessible and so difficult to reliably model. In other words, it is often good method *not* to make one’s inferences and theorizing hostage to whether one has got the details of the underlying physics right. Thermodynamics is a paradigm of a successful science that follows this sort of approach: complex systems with astronomically many degrees of freedom at the level of fundamental physics are characterized in terms of just a few macroscopic parameters whose interrelations are in many cases surprisingly independent of the details of their microscopic realizations. (So much so that much of classical thermodynamics would remain as it is, even if matter at the fundamental level were continuous, as long as it continued to satisfy certain generic constraints.) As Wilson emphasizes, this strategy can produce results that are far more epistemically secure than strategies that require bottom-up modeling from fundamental physics (even if, contrary to what is typically the case, we have enough apparent

information to attempt such a bottom-up approach). I believe that a similar moral holds for methodology: as a simple illustration, consider that a properly designed experiment showing that treatment with a drug causes recovery from a disease can yield secure causal knowledge even if the experimenter has no idea what the fundamental physical laws underlying this causal claim are. In other words, the experiment can provide a route to secure causal knowledge that not only does not “go through” this underlying information but which *insulates* or protects the experimenter from having to know it. A similar conclusion holds for procedures like instrumental variables for inferring causal conclusions from non-experimental data. Thus not only is it the case that in actual fact we are very rarely in a position to even attempt to derive upper level causal claims from underlying physical facts and are forced to rely on other procedures (such as experimentation) instead, but it is also the case that use of those procedures is methodologically superior from the point of view of reliability.

I suggest that we thus arrive at the following upshot. The metaphysically oriented philosophical literature contends that causal claims (and counterfactuals associated with them) cannot be “barely true”; and that these require a specification of “grounds” and “truth makers” before they can be legitimately employed. The idea is that by providing such truth-makers one somehow clarifies or renders respectable claims that would otherwise be problematic. However, as far as methodology goes, exactly the *opposite* conclusion often seems to be supported. When methodological discussion in econometrics, statistics and so on turns to the clarification of causal claims or principles of causal reasoning there is little or no invocation of truth-makers in the metaphysician’s sense. Instead clarification generally takes the non-reductive forms described above—association of causal claims with hypothetical experiments, specification of non-reductive principles linking causal claims to evidence and so on. So, ironically, it is strategies that fail to provide truth-makers in the metaphysicians’ sense and that are non-reductive that turn out to be the useful and clarifying ones.

## 10. Reconfiguring Ontological<sub>2</sub> Questions as Methodological Questions

My discussion so far has argued for the relative independence of methodological considerations from ontology<sub>2</sub>. However, as I have intimated, it is certainly possible to bring these two topics closer together, roughly by reconstruing or reconfiguring ontological<sub>2</sub>/metaphysical<sub>2</sub> questions as questions about methodology. A recent paper by Hitchcock (2012) provides a beautiful illustration of this. Hitchcock’s concern is the role of time in the characterization and individuation of events figuring in causal relationships. As a matter of ontology/metaphysics, this tends to be put in terms of the question of whether the time at which an event occurs is essential to it— for example, if Billy and Suzy throw rocks at a bottle, with Suzy’s rock striking and shattering the bottle at time  $t$  but where Billy’s throw would have shattered the bottle at time  $t + d$ ,  $d > 0$  if Suzy’s throw had not occurred, is the bottle’s shattering at  $t$  the same event as the bottle’s shattering at  $t + d$ ? This is a question that has received different answers from different metaphysicians. However, one can also understand this issue in methodological terms, as concerning how we should characterize the variables that figure in causal relationships given that our goal is the perspicuous representation of what would happen under various possible interventions. When the question is put this way, then, as Hitchcock persuasively

argues, there are compelling methodological reasons to represent the shattering of the bottle at  $t$ , and its shattering at  $t+d$  by different variables—i.e., they should be represented as different events. One of the many advantages of this methodological turn is that it replaces the vague and difficult-to-answer ontological/metaphysical question about whether time is essential, with a more sharply formulated question, where it is also made clear what turns, in terms of methodological consequences, on answering this question one way rather than another.

One might pursue a similar strategy in connection with a number of the examples discussed above. As one illustration, consider the remarks in Section 6 on clarifying the content of causal claims by associating them with specific hypothetical experiments. As readers will likely have noticed, this strategy maps in a straightforward way onto the familiar philosophical idea that causal claims can be thought of as having a contrastive structure—that one should think of a causal claim of overt form “ $C$  causes  $E$ ” as having something like the underlying structure:  $C$  rather than  $C^*$  causes  $E$  rather than  $E^*$ . Here this contrastive structure in effect specifies (at least partially) the hypothetical experiment associated with the original claim—intervening to change the cause from  $C$  to  $C^*$  changes the effect from  $E$  to  $E^*$ . One might think of the contention that causal claims should be understood contrastively in this manner as an ontological (that is, ontological<sub>2</sub>) or metaphysical claim, which is how it is understood in Paul and Hall, 2013—see especially, pp. 21-23. Construed in this way, metaphysicians will have many objections, deriving from, e.g., conflicts with “intuitions” to the effect that causation is an “intrinsic” and “binary” relation between events and from the worry that the contrastive construal makes the “metaphysical character of the causal relation.. determined by human interests” (Paul and Hall, p. 22). Suppose, however, one interprets the question of whether causal claims “are” contrastive not as a metaphysical/ontological thesis but rather (along the lines followed in section 6) as a methodological proposal: that one can very often clarify the content of a causal claim and make it clearer by specifying the intended contrastive structure. Understood in this way, the basis on which “contrastivism” is to be assessed shifts in an important way, since the standard is now methodological fruitfulness rather than consistency with metaphysical intuition. Moreover, in contrast to the disputes surrounding contrastivism as a metaphysical thesis, the issue of whether construing causal claims contrastively is methodologically useful seems like an fairly straightforwardly answerable question—we can look to the additional information apparently provided by contrastive construals, actual practice of those involved in causal reasoning in various domains and so on.

As a second illustration, consider an issue raised in passing above—whether double prevention relations are “causal”. Again one may view this as a question about the basic ontology<sub>2</sub>/metaphysics<sub>2</sub> of causation (which is how it is usually understood) but one can also approach it in a more methodological spirit, as having to do with what grounds, if any, there might be for distinguishing double prevention relations (which lack a connecting process) from other non-backtracking, intervention-supporting dependency relations which do involve a connecting process. Woodward (2006, 2011), drawing in part on empirical results in Lombrozo (2010), pursues this sort of project, arguing that dependency relations involving a connecting process are typically more stable and more “exportable” to new situations than dependency relations lacking this feature and that since this sort of stability is something we value, it makes sense, methodologically, for us

to distinguish, in some way, between dependency relations for which connecting process are present and those for which they are absent. As an added bonus, this approach also suggests that there are good methodological reasons for distinguishing among double prevention relations with respect to their degree of stability, with more stable double prevention relations being regarded as more paradigmatically causal. Again I would argue that if we are willing to re-construe issues about the status of double prevention in this way, they become much more tractable and it becomes easier to see what turns on answering them one way rather than another—something that is missing (in my opinion) when we simply ask whether double prevention relations are “really” causal.

Although I think that in these cases and many others, there is much to recommend this route of reinterpreting metaphysical questions as methodological ones, I acknowledge that most metaphysicians will not find this course appealing—it involves giving up too many of their characteristic concerns and assumptions. I nonetheless put it forward in an irenic spirit, as one possible way of engaging with problems that bear some similarity to those that have concerned metaphysicians, but with different standards for their solution.

### References

- Angrist, J. and Pischke, J-S. (2009) *Mostly Harmless Econometrics*. Princeton; Princeton University Press.
- Baumgartner, M. (2009) “Interdefining Causation and Intervention”, *Dialectica*, 63, 175-194.
- Eberhardt, F. and Scheines, R. (2007) “Interventions and Causal Inference”. *Philosophy of Science*, 74:981-995.
- Glymour, C. (1986) “Comment” *Journal of the American Statistical Association* 81 964-66.
- Glymour, C. (2004) “Critical Notice of *Making Things Happen*” *British Journal for the Philosophy of Science* 55: 779-790.
- Hiddleston, E. (2005) “Review of Woodward, *Making Things Happen*” *The Philosophical Review* 114: 545-547.
- Hitchcock, C. (2012) “Events and Times: A Case Study in Means-Ends Metaphysics” *Philosophical Studies* 160: 79-96.
- Holland, P. (1986) “Statistics and Causal Inference” *Journal of the American Statistical Association* 81: 945-60.
- King, G., Keohane, R. and Verba, S. (1994) *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton: Princeton University Press.

- Lombrozo, T. (2010) "Causal-Explanatory Pluralism: How Intentions, Functions, and Mechanisms Influence Causal Ascriptions" *Cognitive Psychology* 61:303–32.
- Mumford, S. (2009) "Causal Powers and Capacities", in *Oxford Handbook of Causation*, Beebe, H., Hitchcock, C. and Menzies, P. (eds), Oxford: Oxford University Press: 265-78.
- Paul, L. (2012) "Metaphysics as Modeling: The Handmaiden's Tale" *Philosophical Studies* 160: 1-29.
- Paul, L. and Hall, N. (2013) *Causation: A User's Guide*. Oxford: Oxford University Press.
- Reutlinger, A. (2012) "Getting Rid of Interventions" *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 43: 787–795.
- Rubin, D. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies" *Journal of Educational Psychology* 66: 688–701.
- Russo, F. (2012). "Correlational Data, Causal Hypotheses, and Validity". *Journal of the General Philosophy of Science*, 42, 85–107.
- Sider, T. (2011) *Writing the Book of the World*. Oxford: Oxford University Press.
- Spirtes, P, Glymour, C. and Scheines, R. (2000) *Causation, Prediction and Search*. Cambridge, MA: MIT Press.
- Tooley, M. (1987) *Causation: A Realist Approach*. Oxford: Oxford University Press,
- Strevens, M. (2007) "Review of Woodward, Making Things Happen" *Philosophy and Phenomenological Research* 74: 233–249.
- Wilson, M. (Forthcoming) *Physics Avoidance*.
- Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*. New York: Oxford University Press.
- Woodward, J. (2006) "Sensitive and Insensitive Causation" *The Philosophical Review*.115: 1-50.
- Woodward, J. (2012) "Causation, Interactions Between Philosophical Theories and Psychological Research" *Philosophy of Science* 79:961-972.
- Woodward, J. (Forthcoming a) "A Functional Theory of Causation" *Philosophy of*

*Science.*

Woodward, J. (Forthcoming b) “The Problem of Variable Choice”



