**How interventionist accounts of causation work in experimental practice**

**and why there is no need to worry about supervenience**

Tudor M. Baetu

# Abstract

It has been argued that supervenience generates unavoidable confounding problems for interventionist accounts of causation, to the point that we must choose between interventionism and supervenience. According to one solution, the dilemma can be defused by excluding non-causal determinants of an outcome as potential confounders. I argue that this solution undermines the methodological validity of causal tests. Moreover, we don’t have to choose between interventionism and supervenience in the first place. Some confounding problems are effectively circumvented by experimental designs routinely employed in science. The remaining confounding issues concern the physical interpretation of variables and cannot be solved by choosing between interventionism and supervenience.

# 1. Interventionism and mixed-level causal models

The scientific defence of explanatory and predictive models aggregating what may be construed as ‘lower-level’ biological causes and ‘higher-level’ psychological and social determinants of a phenomenon or outcome rests on the notion that the elucidation of the causal structure of the world is a matter of experimental inquiry.[[1]](#footnote-1) If it is possible to intervene on a variable and measure a change in an outcome, then the variable is causally relevant to that outcome. It doesn’t matter that the variable in question is psychological or social rather than biological (Campbell 2008). The experimentalist’s argument should extend to interventionist accounts of causation, including the influential account developed by James Woodward (2003). After all, the notion that causal relevance is demonstrated by means of experimental interventions is the core tenet of such accounts (Holland 1986). Since the experimental scientist is free to test the causal relationships between any two variables, it seems only natural to conclude, as Woodward himself does, that in the context of an interventionist account of causation, “there is no bar in principle to mixing variables that are at what might seem to be different ‘levels’ in causal claims” (2008, 222).

It would seem, however, that this conclusion is problematic. Michael Baumgartner (2009) points out that, according to Woodward’s own account, interventions must satisfy a fixability clause and the requirements of an ideal intervention. Failure to satisfy either creates a confounding problem, since it leaves open the possibility that changes in variables other than the intended target of the intervention may be responsible for the observed differences in outcome.

In response to the challenge, Woodward (2015, 323) proposes that “it is inappropriate to control for supervenience bases in assessing the causal efficacy of supervening properties.” He justifies this claim by arguing that matters of definition and metaphysical necessity act as prior constraints on possible causal relationships between variables. The solution is therefore to exclude variables assumed to be standing in supervenience, definitional, nomological, identity and other non-causal relationships with the tested (independent) variable as potential confounders. This effectively entails a modification of interventionism to allow for inaccurate, or ‘fat-handed’ interventions targeting simultaneously variables standing in non-causal relationships.

In this paper, I argue that Woodward’s solution constitutes an unacceptable compromise undermining the empirical foundations of science. My rejection of Woodward’s solution is motivated by two considerations. First, the methodological consensus is that confounded studies cannot yield conclusive results because they fail to differentiate between rival explanations. And second, the concepts of definitional and metaphysical constraints proposed by Woodward are modelled after a spurious example. I further argue that Woodward is also mistaken in thinking that Baumgartner’s challenge posits an unsurmountable dilemma. Confounding problems arising from violations of the fixability requirement are circumvented by parallel experimental designs typically deployed in basic and clinical biomedical research. These experiments rely on comparability, rather than fixability, which means that no prior knowledge about the nature of confounders–be them causal, supervenience bases or something else–needs to be assumed or hypothesized. As for confounding problems arising from violations of the requirement for accurate interventions, they reflect an uncertainty about the physical interpretation of variables and can only be addressed by gathering additional information about the world.

In Section 2, I discuss some peculiarities of Woodward’s account, drawing a preliminary contrast with interventionism as commonly understood in experimental practice. In Section 3, I analyze the most common type of experimental design in biomedical research, focusing on comparability as an alternative to the fixability requirement. The main goals of Sections 2-3 are to emphasize the fact that causal inference proceeds via a systematic elimination of alternative explanations; and to show that the fixability assumption in Woodward’s account implies an experimental design subjected to limitations which do not affect more popular experimental designs employed in experimental research. In section 4, I discuss Baumgartner’s challenge and Woodward’s response. In Section 5, I detail the reasons for which I reject Woodward’s solution. In Section 6, I argue that Baumgartner’s challenge does not constitute a dilemma. Section 7 summarizes the main claims made in the paper.

# 2. Woodward’s interventionist account of causation

*2.1 Manipulability and ideal interventions*

Woodward’s interventionist account of causation rests on two key definitions, one specifying how causation relates to manipulability, the second what counts as a suitable intervention for manipulating a system in order to demonstrate causation:

1. “A necessary and sufficient condition for *X* to be a (type-level) direct cause of *Y* with respect to a variable set *V* is that there be a possible intervention on *X* that will change *Y* or the probability distribution of *Y* when one holds fixed at some value all other variables *Zi* in *V*. A necessary and sufficient condition for *X* to be a (type-level) contributing cause of *Y* with respect to variable set *V* is that (i) there be a directed path from *X* to *Y* such that each link in this path is a direct causal relationship; […] and that (ii) there be some intervention on *X* that will change *Y* when all other variables in *V* that are not on this path are fixed at some value” (Woodward 2003, 59); and
2. “*I* is an intervention variable for *X* with respect to *Y* iff 1. *I* causes *X*; 2. *I* acts as a switch for all the other variables that cause *X*. That is, certain values of *I* are such that when *I* attains those values, *X* ceases to depend on the values of other variables that cause *X* and instead depends only on the value taken by *I*; 3. Any directed path from *I* to *Y* goes through *X*. That is, *I* does not directly cause *Y* and is not a cause of any causes of *Y* that are distinct from *X* except, of course, for those causes of *Y*, if any, that are built into the *I−X−Y* connection itself; that is, except for (a) any causes of *Y* that are effects of *X* (i.e., variables that are causally between *X* and *Y*) and (b) any causes of *Y* that are between *I* and *X* and have no effect on *Y* independently of *X*; 4. *I* is (statistically) independent of any variable *Z* that causes *Y* and that is on a directed path that does not go through *X*” (Woodward 2003, 98).

*2.2 Preliminary clarifications*

Before continuing, some clarifications are needed. First, Woodward frames his account in terms of ‘possible interventions,’ which include both actual experimental interventions and counterfactual scenarios (i.e., what would happen, were an intervention to take place, even if the intervention is not actually feasible). For the remainder of the paper, I will consider only actual experimental interventions. However, the same analysis should apply to counterfactual scenarios (i.e., what would happen if an experiment satisfying certain requirements were to be conducted). For example, one may ask what would be the difference in respect to an outcome between two states of a system, one in which no intervention is conducted, the other in which an intervention was conducted while the background of all other variables that may have an impact on the outcome would remain constant throughout the duration of the experiment.

Second, Woodward operationalizes the notion of causation, equating causation with whatever a possible intervention experiment identifies as a cause, thus providing both a definition of what it means to be a cause and a method for determining causation. In what follows, I will focus on the methodological validity of causal tests as specified by Woodward, and not discuss the implications for definitions of causation. It must also be emphasized that there are methods for inferring causation which don’t require interventions at all (e.g., conditional probabilities, latent variable analyses, Granger causality), although they are sometimes regarded as providing lower quality evidence (Evidence Based Medicine Working Group 1992). Hence, I do not assume that interventions are necessary for testing for or defining causation.

Finally, science seldom deals with singular causation (Rubin 1974). What is compared in the vast majority of controlled experiments, are not individuals, but groups of individuals (animals, cells, DNA molecules, etc.). Moreover, measurement is a statistical concept. Even in the relatively rare cases where the same system serves as both test and control, more than one intervention and one measurement are performed, and the data generated summarized by a statistic. Thus, the causal tests and causal attributions discussed in this paper concern group-level causation, not singular events.

*2.3 The validity of the causal test*

The experimental design tacitly assumed in Woodward’s definitions consists in comparing two states of a system, as assessed by measuring an outcome of interest. The intervention changes the value of the variable whose causal relevance to the outcome is tested. This is the independent variable, or *X* in Woodward’s formulation. The outcome corresponds to the putative effect of the tested variable, that is, the dependent variable or *Y*. To demonstrate that the independent variable *X* makes a difference, or is causally relevant, to a measured difference in outcome *Y* between the two states of the system, three control strategies must be implemented:

1. An intervention on *X*, that is, an experiment or quasi-experiment (e.g., a natural experiment), must be conducted. This is the core feature of all interventionist accounts of causation. Manipulation–as opposed to mere observation of a difference in outcome between two conditions–is standardly required to establish the directionality of causation (i.e., changes in *X* cause differences in *Y* rather than the other way around) and rule out the possibility that the changes in *X* and *Y* are correlated due to a common cause. If the two are divergent effects of a common cause or if *X* is an effect of *Y*, then the intervention on *X* is not expected to have an impact on the outcome *Y*. However, if there is a causal pathway linking *X* and *Y* as upstream cause to downstream effect, then interventions on *X* are expected to result in changes in *Y*.
2. The background of all other variables that may have an impact on the outcome should remain constant throughout the duration of the experiment. This is the ‘fixability clause’ appearing in (M), which states that the values of all variables causally relevant to the outcome, minus those that may mediate the causal impact of the tested variable *X* on the measured outcome *Y*, should be fixed.[[2]](#footnote-2) If the background of relevant variables changes during the experiment, it is impossible to determine whether a difference in outcome is due to the variable manipulated in the experiment or to some other explanation. An alternative explanation amounts to an underdetermination issue termed a ‘confounding problem.’
3. The intervention–or quasi-intervention in the case of a natural experiment–should be accurate (unbiased, valid or ‘ideal’) in the sense that it should target only the variable under investigation. Accuracy is required to demonstrate the causal relevance of the tested variable to the difference in outcome. If the accuracy of the intervention cannot be demonstrated, it is still possible to demonstrate that the intervention itself is causally relevant (Winch and Campbell 1969). However, the causal efficacy of the intervention may be attributed to the fact that the intervention targets some other variable (a confounder) which influences the outcome (Chalmers et al. 1983). This condition corresponds to definition (IV.3) which, as Woodward (2015, 315) points out, “embodies a set of assumptions about what variables count as potential confounders one needs to control for in an experiment to determine whether *X* causes *Y*.”

Causal inference follows the general pattern of reasoning outlined in Mill’s method of difference (1843, Chapter VIII, § 2; Sekhon 2008), namely that of a comparative inference whereby those aspects of a situation known to be constant (controlled) between two situations are ruled out as possible explanations of differences in outcomes. The strategy is therefore to compare two experimental conditions, one in which an intervention targeting a putative causal variable is present (‘test’) and one in which the intervention is absent (‘control’). (i) is meant to rule out the possibility of a non-causal correlation between the tested variable and the observed differences in outcome, as well as the possibility of a reverse causation scenario. (ii) is meant to rule out explanations of a difference in outcome other than the intervention. (iii) is meant to further rule out explanations involving variables other than the designated independent variable which might have been affected by the intervention. If internal controls (i)-(iii) are implemented, there is good evidence to rule out alternative explanations and therefore conclude that the tested variable is the causal difference-maker responsible for the difference in outcome.[[3]](#footnote-3)

*2.4 The experimental design*

Woodward takes the definitions (M) and (IV) to capture general desiderata of an interventionist account of causation. A closer inspection, however, reveals two peculiarities associated with a rather unusual experimental design. The first peculiarity, which it shares with some experiments in clinical research, is that it relies on the rather stringent and experimentally demanding fixability condition stated in (M). In experimental practice, this condition needs to be satisfied when the same system serves as both test and control, in which case the inference refers to a comparison between ‘before’ and ‘after’ conditions. For instance, something akin to Woodward’s fixability clause is required when implementing single-subject cross-over experiments (Langreth and Waldholz 1999). This design is suitable for assessing potential treatments for chronic and stable conditions, such as chronic pain and psychiatric disorders. The stability of a condition is taken to be an indication that its determinants remain constant over time. A common example of such procedures is a ‘sandwich’ experimental design, where a condition is measured in the same patient before a treatment is introduced, during the treatment, and after withdrawal of the treatment.

The main strength of this design lies in the fact that it minimizes confounding that may arise due to differences between individual systems. The main limitation of the design is that it cannot be used to demonstrate causation in systems that evolve significantly over the timeframe of the experiment. For instance, it is impossible to determine whether an outcome such as recovery from the common flu is due to the treatment’s efficacy or to the natural progression of the disease. Likewise, absence of change in an outcome does not necessarily mean that a treatment is inefficacious, as the latter may have had a positive impact by stopping or delaying progression to more severe stages of the disease. In general, biological systems are known and expected to change over time and to be affected by treatment/exposure history (‘carry-over effects’). It is therefore important to emphasize that the fixability clause limits the applicability of any experimental design that requires it, a limitation which accounts for the fact that such designs remain relatively rare in clinical and basic science.

The second peculiarity of the experimental design tacitly assumed in Woodward’s account is borrowed from causal modelling (Spirtes et al. 1993). Woodward doesn’t only require that an unspecified causal background is kept constant for the duration of the experiment, but he further assumes that this background and its structure are specified. This requirement marks an important point of divergence from typical controlled experiments in the life sciences. In most cases, what scientific experiments test is a categorical (qualitative) hypothesis of the form ‘*C* is (or not) causally relevant to differences in outcome *E* given an unknown (and therefore unspecified) background of other determinants of *E.’* This is because research in basic science is largely dedicated to the elucidation of the unknown causal background by piecing together bits and pieces of information about causal dependencies between manipulated and measured variables (Baetu 2019; Bechtel and Richardson 2010; Craver 2007). Alternatively, clinicians usually want to directly test the causal efficacy of a treatment in humans, irrespective of putative details of the pathophysiological mechanisms involved (Howick 2011).

The epistemic context Woodward has in mind is quite different. The starting point is not a hypothesis about the causal relevance of a variable vis-à-vis a difference in outcome, but a hypothesis postulating a more complex causal structure, usually one in which the relationships between causes and effects are quantified (a structural causal model). At the very least, a directed causal graph representing the putative causal structure of a system is first hypothesized, then tested by measuring changes in an outcome variable *Y* following an intervention on the independent variable *X* while all other variables (nodes) in the graph that are not on the *X–Y* path are fixed. This means that what is effectively tested in Woodward’s experimental design is not a single causal dependency relationship, but rather the whole causal structure postulated by a causal graph. The obvious strength of this approach is that it makes possible the testing of hypotheses postulating more complex causal structures. As Pearl et al. point out, “a researcher who has scientific knowledge in the form of a structural equation model is able to predict patterns of independencies in the data […]. Conversely, […] observing patterns of independencies in the data enables us to say something about whether a hypothesized model is correct” (2016, 35). On the downside, the approach is epistemically demanding, as at least some prior knowledge of a system is needed, and subject to limitations imposed by the assumptions underpinning causal modelling.[[4]](#footnote-4)

# 3. Controlled experiments in the life sciences

*3.1 The experimental design and the validity of the causal test*

Woodward frequently relies on examples in which the test and control conditions are distinct systems, such as Randomized Controlled Trials in which two distinct populations of patients are compared [e.g. (2003, 95-96)]. Conceptually speaking, no contradictions arise by imposing a fixability condition in such experiments. However, when a comparison is drawn between two distinct systems, neither knowledge, nor constancy of the causal background is required. The general idea is to compare distinct systems deemed identical in all relevant respects minus a manipulated variable. The main strength of a parallel experimental design (one in which a test and a control experiment run side by side) relies on the fact that an allowance is made for the possibility that the causal background may change over time.

Once again, to demonstrate that a variable is causally relevant, alternative explanations must be ruled out:

1. The possibility that the variable is the effect or a mere correlate of the outcome must be ruled out. This is achieved by conducting an experiment (or quasi-experiment) in which the variable is manipulated, as opposed to merely observing situations in which a variable takes different values.
2. The causal test is valid only if the test and the control conditions are comparable in all relevant respects except for the variable manipulated in the experiment. Failure to ensure comparability raises the possibility that some other difference between the two conditions (a confounder) is responsible for the observed difference in outcome. Comparability demands: (ii.1) that possible confounder-variables take the same values in the test and control conditions at the onset of the experiment; and (ii.2) that the experiment is shielded from unequal external interferences other than the intervention targeting the test, but not the control condition. In other words, comparability is meant to ensure that the test and control systems start in the same state and evolve in the same way except for whatever changes are brought about by the intervention on the independent variable and its downstream effects. In contrast to the fixability condition figuring in Woodward’s analysis, no prior knowledge of causal structures is assumed or hypothesized and it is not required to keep the values of confounder-variables constant for the duration of the experiment. An allowance is made for the possibility that the values of potential confounders may change during the experiment, due to external inferences or to the natural progression of the system, and that these changes may affect the measured outcome. However, whatever these changes are, their impact is expected to be identical in the two arms of the experiment.[[5]](#footnote-5)
3. Finally, an accurate (unbiased, valid, ‘ideal’) intervention is required to rule out explanations involving variables other than the designated independent variable which might have been triggered by the intervention.

*3.2 The comparability and accuracy requirements*

In the experimental practice of the life sciences, comparability (ii) has two components, one referring to the living systems under investigation, the other to the replication of the same experimental background in the test and control conditions. Experimental practices deployed to ensure the latter include the standardization and operationalization of the techniques of measurement and intervention. A parallel experimental design in which test and control are simultaneously deployed side by side is commonly adopted in order to ensure that causal interferences external to the experimental setup have an equal impact on both arms of the experiment (ii.2).

Ensuring the comparability of biological systems at the onset of the experiment (ii.1) is far more challenging. In basic science, the preferred strategy is to systematically remove differences between biological systems in order to generate isogenic/homozygous strains and cell-line clones, which are then maintained in standardized living conditions (Ankeny 2001). Further precautions are taken at the onset of an experiment to ensure that test and control biological systems are ‘synchronized’ (e.g., cells are at the same stage of the cell cycle).

The accuracy of the intervention (iii) is ensured by subjecting techniques of intervention to a process of validation meant to demonstrate that the technique targets only the variable under investigation and no other variables that may contribute to differences in outcomes. Test validity is in part demonstrated by including additional positive and negative controls. For example, it is common practice to perform placebo interventions (pipetting, mixing, centrifuging, incubating) in the control arm of the experiment. Such interventions are meant to ensure that the relevant difference maker is not some generic lab procedure, such as gently shaking the cells, but rather the investigated variable, say, a virus, which is added by gently shaking the cell suspension.

In clinical research, as well as other fields dealing with natural or heterogeneous populations, causal comparability cannot be assumed. In Randomized Controlled Trials, causation is inferred by estimating the probability of generating differences in outcome in virtue of random allocation (Fisher 1947; Hill 1955). By itself, statistical inference only serves to evaluate the likelihood of a strictly statistical explanation (e.g., a random allocation effect) of the difference in outcome. If a statistical explanation can be ruled out, it is possible to infer the probable existence of a cause other than random variation, although this does not say anything about the identity of this additional cause. In as much as it can be further demonstrated that the experiment did not introduce any differences other than random variation inherent to the studied sample and the treatment intervention, the additional causal contribution can be identified as the treatment administered to the test group. This requirement is satisfied by ensuring that the allocation of patients to test and control groups (the experimental intervention) is not biased, which is achieved by randomizing the study (Baetu 2020; Shadish et al. 2002, Ch. 8).[[6]](#footnote-6)

# 4. Controlled experiments and supervenience

*4.1 Supervenience as a confounding problem*

What does the metaphysical concept of supervenience have to do with experimental methodology? According to an argument by Baumgartner (2009), if the mind supervenes on the brain, then there is no possible intervention satisfying (M) and (IV). As discussed in Section 2, fixability refers to the fact that, to assess the causal relevance of a variable, all other variables relevant to the measured outcome should be kept constant. However, if psychosocial variables supervene on biophysical ones, it is difficult to see how supervening states could be manipulated while keeping their supervenience bases fixed. A second confounding problem arises in relation to the requirements of an ideal intervention, which should be accurate (unbiased, valid)–that is, specifically target only the variable tested for causal relevance. However, once again, if psychosocial variables supervene on biophysical ones, it is impossible to intervene on supervening states without also intervening on their supervenience bases. The fact that it is impossible to intervene on a supervening state without intervening on its supervenience basis entails that internal controls (ii) and (iii) cannot be implemented, meaning that experiments are confounded. Hence, causal tests fail to generate conclusive evidence for causation.

This conclusion can be explained as follows. A confounder offers a rival explanation of the observed difference in outcome between test and control. If other variables determine the measured outcome–be it in virtue of causal influence, supervenience, nomic necessity or whatever other way–, then the causal test will be inconclusive since the possibility that something other than the independent variable is responsible for the difference in outcome cannot be ruled out. Or, on the assumption of supervenience, test and control differ in three respects: the outcome of interest; the independent psychosocial variable tested for causal relevance; and the biophysical supervenience basis of the psychosocial variable in question. The difference in outcome can be explained by either a change in the psychosocial variable and/or a change in the biophysical supervenience basis. Since the experiment fails to discriminate between possible explanations, it is bound to remain inconclusive.

*4.2 Woodward’s solution to the supervenience confounding problem*

As pointed out in Section 2.3, the experimental design assumed in (M) and (IV) diverges from typical designs in experimental science in as much as it is meant to test causal graphs representing the putative structure of a system rather than the causal relevance of a single variable given an unspecified, but comparable, causal background. Consistent with this design, Woodward’s initial reaction to Baumgartner’s challenge is to remark that a hypothesis postulating non-causal relationships is not a bona fide causal graph and therefore such a hypothesis cannot be tested by conducting an experiment satisfying the desiderata of his interventionist account of causation. Thus, “one might exclude as illegitimate any graph containing variables that violate INF” (Woodward 2015, 326).[[7]](#footnote-7)

I think this response is correct. Any scientific experiment is designed to test a clearly specified hypothesis. It shouldn’t be surprising that an experimental design meant to test one type of hypothesis is unsuitable for testing some other type of hypothesis. In the end, however, Woodward (2015, Section 6) does consider hypotheses postulating graph-like structures containing both causal and non-causal relationships between variable-nodes and, contrary to his initial suggestion, attempts to test such structures by means of intervention experiments.

In principle, two methodologically sound testing approaches are available at this point. One option is to posit by hypothesis structures containing both causal and non-causal relationships and devise a methodological valid experiment for testing these hypotheses. The problem is that science has yet to develop and validate experimental designs capable of discriminating between accidental associations, correlations, causal dependencies and non-causal supervenience dependencies. The other option is to rely on prior knowledge about non-causal relationships between variables and attempt to screen them out after experimental data has been generated. Here too, we encounter a difficulty: there is scarcely any prior knowledge to rely on.

Experiments in neuroscience consist in manipulating biological variables in order to test their causal relevance to psychological variables (Bickle 1998; Craver 2007). Such experiments allow researchers to identify causal determinants of psychological outcomes, order these determinants along causal pathways and propose mechanisms causing psychological states and associated behavioural responses. However, nothing here demonstrates that psychological states supervene on biological states.[[8]](#footnote-8) This is because these experiments only test for causal relevance, not for supervenience. Hence, they provide conclusive evidence that biological activity is causally relevant to psychological outcomes, but no evidence that psychological states supervene on biological activity.

Of course, absence of evidence is not evidence for absence. These experiments don’t prove that psychological states don’t supervene on biological activity either. Still, the net result is an absence of prior knowledge of supervenience relationships. The same applies to life sciences in general, which tend to be primarily experimental disciplines relying heavily on statistical methods of data analysis for uncovering correlations and on controlled experiments for identifying the causal structures linking correlated variables. Other disciplines are more generous when it comes to prior knowledge of supervenience relationships. For instance, the statistical-mechanical formulation of the ideal gas law is often interpreted as the claim that temperature is the average molecular kinetic energy of the molecules composing a gas. If this interpretation is correct[[9]](#footnote-9), it may be conceded that, at least in some cases, prior knowledge of supervenience is available. Even so, for this solution to be practicable, methodological guidelines specifying which supervenience relationships are sufficiently well justified and how supervenience artefacts should be screened out must be outlined.

Woodward doesn’t choose either of these two methodologically sound, but difficult to implement strategies. Instead, he attempts to carve a way between them. He accepts by hypothesis that supervenience relationships exist and proposes that, “in assessing the causal impact of *X* on *Y*, it is not legitimate to consider what would happen to *Y* under combinations of manipulations of the values of *X* and other variables (besides *Y*) that set these to combinations of values that violate the relationships taken to hold definitionally among those variables” (2015, 329). However, this kind of screening presupposes prior knowledge about non-causal constraints, which immediately raises questions about scientific evidence for such constraints.

Woodward’s solution to the problem of the missing evidence is simply to deny that such evidence is needed. He argues that there are obvious cases when non-causal determinants of an outcome must be treated as necessary constraints on a causal structure and not as confounders to be experimentally controlled. He illustrates this approach by means of an example (2015, 327-28): The blood concentration of high-density cholesterol (*HD*)–or, to be more precise, high-density lipoprotein complexes containing cholesterol–lowers the probability of cardiovascular disease (*D*). The blood concentration of low-density cholesterol (*LD*)–again, packaged in lipoprotein complexes–increases this probability. Now, what is most easily measured is the total concentration of blood cholesterol (*TC*), which, Woodward tells us, is defined as *HD* + *LD*. *TC* too is causally relevant to disease, although its effect will depend on the precise composition of *HD* and *LD*. Since *TC* is by definition the sum of *HD* and *LD*, demanding that *TC* and *HD* are fixed while intervening on *LD* inevitably entails that *LD* is fixed as well, leading to the counterintuitive result that *LD* is not causally relevant to *D*. This result can be avoided if we stipulate that, “in order to characterize the effect of *LD* on *D*, it is not legitimate to consider what would happen to *D* under a ‘manipulation’ that changes the value of *LD* while *HD* and *TC* are held constant” (2015, 329). Are we entitled to such a stipulation? According to Woodward, yes. That *TC* = *HD* + *LD* is a matter of self-evident logical necessity constraining all possible causal relationships between variables.

If we agree with the argument thus far, then we may want to take a step further and endorse the claim that, just as definitional relationships are a matter of logical necessity, supervenience relationships between variables are a matter of metaphysical necessity positing similar constraints on possible causal relationships. The argument here is that “if it is ‘metaphysically impossible’ to change the value of a supervening variable like *M1* while holding *P1* [its supervenience base] fixed, then the very fact that this is impossible is itself an indication that counterfactuals with this antecedent do not tell us about the causal effect (or the absence of such an effect) of *M1* on other variables” (2015, 335). Thus, Woodward concludes, interventionist accounts of causation should be modified to allow for ‘fat-handed’ interventions inaccurately targeting both supervening states and their supervenience bases.[[10]](#footnote-10)

# 5. A critical assessment of Woodward’s solution

*5.1 The problem of inconclusive (confounded) causal tests*

Woodward’s proposal faces two problems. First, the consensus in experimental science is that the internal validity of confounded studies is compromised (Leighton 2010; Shadish et al. 2002). As discussed in Section 4.1, under the hypothesis of supervenience, a difference in outcome can be explained by either a change in the psychosocial variable targeted by an experimental intervention and/or a change in its biophysical supervenience basis. Since the experiment fails to discriminate between rival explanations, it remains inconclusive. No amount of a priori arguing about what is legitimate or illegitimate to control for can possibly squeeze a conclusive result from a confounded study. A methodologically valid test for ruling out alternative explanations must be specified. For instance, if certain supervenience artefacts are to be screened out, then there must be independent evidence that the supervenience relationships in question exist, and a method of screening and demonstrating causation must be detailed.

*5.2 Definitional and metaphysical constraints are modelled after a problematic example*

Second, Woodward’s notion of metaphysical constraint is modelled after that of definitional constraint, which in turn is based on a conflation of conventional definitions and empirical regularities. What does the fact that *TC* is defined as *HD* + *LD* have to say about the physical structure of the world? Nothing. *TC* is whatever we want it to be. It could be ½ *LD* + *HD*¼ or any other function of *HD* and *LD*. Irrespective of how we define *TC*, it will still have a causal impact on disease in virtue of the fact that *HD* and *LD* are causally relevant to disease. But if *TC* is an arbitrarily defined variable, then the proposed graph-like structure detailing the contributions of *TC*, *HD* and *LD* to cardiovascular disease is a pseudo-hypothesis. It too says nothing about the physical structure of reality.

On the other hand, if we give the statement ‘*TC* = *HD* + *LD*’ the physical interpretation ‘all cholesterol is present in the blood as either high- or low-density lipoprotein complexes,’ things change dramatically. This is no longer a matter of definition, but a testable hypothesis about reality. If *LD* decreases, *HD* doesn’t increase as a matter of logical necessity, but in virtue of a metabolic process converting one form of cholesterol into the other. This is a bona fide causal process which, unlike logical necessity, is located in the liver and can be chemically and biologically manipulated. Under this interpretation, the variable *TC* refers to the metabolic pathways in which *LD* is converted into *HD* and vice versa. It is legitimate to hypothesize and test a causal structure in which *LD* causes *HD*, *HD* causes *LD* and both have an impact on *D*.

Whether a physical quantity, such as total cholesterol, is conserved cannot be a matter of definition, but is contingent on how the world is.[[11]](#footnote-11) More generally, it is difficult to see how an equation describing the relationship between physically interpreted or operationally defined variables can be a definition. A physical interpretation should anchor these variables to physical reality. Likewise, an operationalized definition should anchor them to a set of physical operations performed on physical things in the context of a physical experimental setup. If variables are linked to reality, then reality, not convention, dictates how they relate to one another. The conservation of mass, the laws of chemical combination and countless other equations in science are contingently and a posteriori true in virtue of experimental testing, not a priori true as a matter of definitional or metaphysical necessity.[[12]](#footnote-12)

# 6. Baumgartner’s challenge revisited

*6.1 It is not necessary to fix the background of confounders to demonstrate causation*

If we agree to reject Woodward’s amendments for the above reasons, it would seem that we are back to Baumgartner’s challenge. According to Woodward’s account, interventions must satisfy a fixability clause [(M) in Section 2.1; condition (ii) in Section 2.2]. But if psychosocial variables supervene on biophysical ones, supervening states cannot be manipulated while keeping their supervenience bases fixed. Thus, we must choose between: (a) maintaining the supervenience thesis while abandoning the fixability requirement; or (b) retaining the fixability requirement while discarding the supervenience thesis.

The dilemma can be defused by replacing the fixability requirement [condition (ii) in Section 2.1] with a comparability requirement [condition (ii) in Section 3.1]. In other words, the potentially undesirable result of having to consider what “would happen to *D* under a ‘manipulation’ that changes the value of *LD* while *HD* and *TC* are held constant” or “that in order for *M1* to have a causal effect on *M2* (or *P2*), there must be a possible intervention that changes *M1* while *P1* [the supervenience basis] is held fixed” can be avoided by conducting a garden-variety experiment comparing in parallel comparable test and control systems.

For example, under the experimental design described in Section 3.1, it is possible to test the causal relevance of *LD* vis-à-vis *D* without having to fix any elements of the causal background, such as *HD* and *TC*. Instead of hypothesizing a causal structure in which *LD* is embedded and demanding that parts of this structure are fixed (kept constant) for the duration of the experiment, this design requires that the test and control systems are initially identical (ii.1), and that they evolve in the same way except for whatever changes are brought about by an intervention on *LD* and its downstream effects (ii.2). If, following an intervention on *LD* in the test system, the values of *D*, *HD* and *TC* (or of any other variable) differ from those measured in a comparable control, the experiment demonstrates that these differences are attributable to the intervention on *LD*, which is the only respect in which test and control systems are different. Since the causal background is allowed to follow its natural progression, the experimenter can remain ignorant about the existence of specific background confounders. This includes other causes, supervenience bases or whatever else may act as a confounder. In this example, there is no need to know or hypothesize anything about *HD*, *TC* and the relationships between them. This completely evacuates any reference to definitional and metaphysical constraints.

*6.2 We don’t have to choose between the supervenience thesis and accurate interventions*

The second target of Baumgartner’s challenge is the requirement of an ideal (accurate, unbiased, valid) intervention [(IV.3) in Section 2.1; (iii) in Sections 2.2 and 3.1]. If psychosocial variables supervene on biophysical ones, it is impossible to intervene on supervening states without also intervening on their supervenience bases. It would seem therefore that we must choose between: (a) maintaining the supervenience thesis while abandoning the requirement for accurate interventions; and (b) retaining the requirement for accurate interventions while discarding the supervenience thesis.

Woodward takes this to be a hard dilemma and chooses (a). As discussed in Section 5.1, this choice entails a flawed experimental design undermining the validity of causal tests, hence my rejection of (a). However, I am equally reluctant to embrace (b). Whether it is indeed the case that there cannot be differences in mental states without differences in physical states is an empirical matter. The suggestion that one could prove or refute a matter of fact by a purely conceptual leveraging–that is, in the absence of independent empirical evidence–is eminently dubious. Accordingly, my strategy is to challenge the validity of the dilemma.

Consider the following findings reported by Rainville et al. (1997): (1) Noxious heat stimulation results in both subjects reporting significantly higher pain ratings, as well as increased neuronal activity in the somatosensory (SI/II) and frontal cortices (anterior cingulate cortex, or ACC), as contrasted with neutral stimulation, which results in lower pain ratings and lower brain activity in these areas. (2) Patients with frontal lobotomies or cingulotomies sometimes still feel pain in response to noxious stimulation but report it as less distressing or bothersome; this is not observed in control subjects. (3) Hypnotic suggestion results in increases/decreases in pain unpleasantness ratings, with no changes in the perceived intensity of the pain sensations induced by noxious stimulation; perceptual dissociation is not observed under hypnosis alone or no hypnosis control conditions. (4) Hypnotic suggestion selectively targeting pain unpleasantness (as opposed to hypnosis alone and no hypnosis controls) results in increases/decreases in ACC, but not SI/II, activity.

Four controlled experiments are described above. It is generally agreed that experiment (1) demonstrates the causal relevance of noxious stimulation to pain ratings and brain activity. The experiment involves comparable subjects under comparable conditions (ii is satisfied), and there are no reasons to believe that an intervention on temperature, an external stimulus, is inaccurate (iii is satisfied). Experiment (2) provides evidence that ACC brain activity is causally relevant to pain unpleasantness ratings. From a methodological standpoint, this is by far the weakest evidence. The surgical interventions involved are not identical in all patients, who are all affected by other, very diverse, conditions in the first place. This entails a potential lack of comparability, further aggravated by small sample size, making this kind of studies particularly susceptible to confounding. As far as the accuracy of the intervention is concerned, it is well understood how surgical interventions work and what their biological targets are, although possible confounding due to placebo effects is not entirely out of question.

What about experiments (3) and (4)? Prima facie, the hypnotic suggestion intervention is psychological in nature, it seems to target a psychological state (consciousness) and has been used for more than a century to produce psychological effects such as perceptual dissociation. Experimentally speaking, the intervention is psychologically accurate since it differentiates between the relevant psychological variables (‘pain unpleasantness’ vs. ‘pain intensity’), which validates its use in experiment (4). Still, we may hesitate to infer causation because it is not known what psychological variables such as ‘pain unpleasantness’ ultimately refer to. Pain ratings are defined operationally, in terms of tests in which subjects are asked to rate or report on phenomenological dimensions of pain. Little is known about the inner workings of these measurement techniques (i.e., we don’t know how subjects come to rate or report their subjective experiences). The fact that hypnotic suggestion results in both psychological and biological changes adds further uncertainty. It could be that the intervention causes changes in ACC activity, as measured by an imaging technique, and these changes are in turn responsible for the ratings of the variable ‘pain unpleasantness.’ Consistent with this uncertainty, Rainville et al. don’t make any claims about the variables targeted by hypnotic suggestion and their possible causal relevance to the outcomes measured in experiment (4). Their discussion is focused entirely on the correlation between selective changes in the perceived unpleasantness of painful stimuli and ACC activity, which is taken to provide “direct evidence [i.e., more direct than evidence from lesion studies mentioned in (2)] of a specific encoding of pain unpleasantness in the ACC [as opposed to sensory dimensions, such as location, quality, and intensity of noxious stimulation, which are presumably processed by the somatosensory cortex].”

The reluctance to conclude that ‘pain unpleasantness’ is causally relevant to ‘ACC activity’ reflects an uncertainty about the physical interpretation of the variable ‘pain unpleasantness,’ the inner workings of reporting technique used to measure it, and the target of the hypnotic suggestion intervention. Now, according to Baumgartner’s analysis, this uncertainty ultimately reduces to two mutually exclusive metaphysical scenarios: either it will turn out that psychological states supervene on neural ones, or not. If the supervenience scenario turns out to be true, experiments (3) and (4) are confounded, since the hypnotic suggestion intervention always targets simultaneously two variables, ‘pain unpleasantness’ and its supervenience basis. If the supervenience scenario is false, then psychological variables, such as ‘pain unpleasantness,’ can be manipulated independently of biological variables, such as ‘ACC activity,’ and therefore tested for causal relevance.

Is Baumgartner correct? A priori, no. That supervenience is defined as any kind of determination relationship other than causation doesn’t entail that all possible physical interpretations fall into two mutually exclusive categories: supervenience or causal interpretations. By analogy, we can define the property ‘is corpuscular’ as any property other than ‘is a wave,’ yet it doesn’t follow from here that things in the universe have either the property of being waves or that of being corpuscles; a priori, things could have one, or the other, or both, or none of these properties.

For example, if two operationally defined variables, one psychological and one biological, refer to the same biological activity, as reductionists have it, then supervenience cannot be a relationship between distinct ontological items. Rather, supervenience describes a coreference scenario in which two variables measure the same ontological item. But, as far as we know, the measurement procedures in virtue of which the two variables are operationally defined are strictly causal in nature (i.e., the measured values of the two variables are observable effects reflecting changes in the ontological item measured). If so, then supervenience is physically speaking nothing else than a bifurcating causal structure. Consider now a non-reductive scenario in which a psychological something (entity, property, etc.) supervenes on a biological activity. To produce a change in the supervening psychological thing, one needs to change biological activity. Is the change instantaneously propagated from the supervenience basis to the supervening thing? Or is it propagated with a finite speed, as dictated by special relativity? If the former, then the psychological thing and the biological activity cannot be physically individuated any more than two entangled electrons can. If the latter, then supervenience is just a potentially novel form of causal interaction.

There are, of course, many other possibilities. The scenarios sketched above are solely meant to illustrate the fact that, at least under some physical interpretations, supervenience and causality are not mutually exclusive. In turn, this supports the conclusion that there is no dilemma. There is only uncertainty. Accordingly, there is no call to choose between current experimental methodology for demonstrating causation and the supervenience thesis. What we should do instead is gather more information about the world.[[13]](#footnote-13)

# 7. Conclusion

In a recent paper, Baumgartner points out that the supervenience thesis entails that experiments involving interventions on supervening states, such as psychosocial variables, are plagued by an unavoidable confounding problem. Woodward attempts to deflect this conclusion by amending his own interventionist account of causation to exclude non-causal determinants of an outcome as potential confounders. I reject his proposal because it amounts to a methodologically invalid test which cannot distinguish between rival explanations. I also reject the notion that certain confounding problems can be assimilated to definitional or metaphysical constraints on the causal structure of the world. Definitional constraints don’t have anything to say about the physical structure of the world, while metaphysical constraints should be open to empirical testing. Finally, I argue that contrary to Woodward’s assumption, Baumgartner’s challenge is not a dilemma. Experimental designs routinely deployed in science don’t require fixing or specifying background determinants of an outcome, thus circumventing one of the confounding threats highlighted by Baumgartner. The other confounding threat stems from an uncertainty about physical interpretation. Uncertainty is lack of information. It doesn’t entail that we must choose between current methodological standards and supervenience. I conclude therefore that there are no legitimate grounds to revise interventionist accounts as proposed by Woodward in his response to Baumgartner.

Acknowledgements

I would like to thank the editors of this volume, as well as the anonymous reviewers for their comments on previous versions on this paper. This research was supported by SSHRC Grant # 430-2020-0654.

# Bibliography

Ankeny, R. 2001. "Model Organisms as Models: Understanding the ‘‘lingua franca’’ of the Human Genome Project." *Philosophy of Science* 68:S251–S61.

Baetu, T. M. 2019. *Mechanisms in Molecular Biology*. Edited by Grant Ramsey and Michael Ruse, *Elements in the Philosophy of Biology*. Cambridge: Cambridge University Press.

———. 2020. "Causal Inference in Biomedical Research." *Biology and Philosophy* 35:43.

Baumgartner, M. 2009. "Interventionist Causal Exclusion and Non-Reductive Physicalism." *International Studies in the Philosophy of Science* 23 (2):161-78.

Bechtel, W., and R. Richardson. 2010. *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*. Cambridge, MA: MIT Press.

Bickle, J. 1998. *Psychoneural Reduction: The New Wave*. Cambridge: MIT Press.

Campbell, J. 2008. "Causation in Psychiatry", in K. Kendler and J. Parnas (eds.), *Philosophical Issues in Psychiatry*, Baltimore: Johns Hopkins University Press, 196-216.

Chalmers, T. C., P. Celano, et al. 1983. "Bias in Treatment Assignment in Controlled Clinical Trials." *New England Journal of Medicine* 309 (22):1359-61.

Craver, C. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press.

deCharms, R. C., F. Maeda, et al. 2005. "Control Over Brain Activation and Pain Learned by Using Real-Time Functional MRI." *Proceedings of the National Academy of Sciences of the United States of America* 102 (51):18626-31.

Eberhardt, F., and R. Scheines. 2007. "Interventions and Causal Inference." *Philosophy of Science* 74:981-95.

Evidence Based Medicine Working Group. 1992. "Evidence-Based Medicine. A New Approach to Teaching the Practice of Medicine." *Journal of the American Medical Association* 268:2420–25.

Fisher, R. A. 1947. *The Design ol Experiments*. Fourth edition ed. Edinburgh: Oliver and Boyd.

Hardcastle, V. 1999. *The Myth of Pain*. Cambridge, MA: MIT Press.

Hill, A. B. 1955. *Principles of Medical Statistics*. 6th ed. New York: Oxford University Press.

Holland, P. W. . 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396):945-60.

Howick, J. 2011. *The Philosophy of Evidence-Based Medicine*. Oxford: BMJ Books.

Ioannidis, J. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2 (8):e124.

Jaynes, E. T. 2003. *Probability Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Langreth, R., and M. Waldholz. 1999. "New Era of Personalized Medicine: Targeting Drugs for Each Unique Genetic Profile." *Oncologist* 4 (5):426–27.

Leighton, J. P. 2010. "Internal Validity." In *Encyclopedia of Research Design*, ed. N. J. Salkind. Thousand Oaks, CA: SAGE.

Melzack, R. 2001. "Pain and the Neuromatrix in the Brain." *Journal of Dental Education* 65 (12):1378-82.

Mill, J. S. . 1843. *A System of Logic, Ratiocinative and Inductive*. London: John W. Parker.

Pearl, J. 2000. *Causality. Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Pearl, J., M. Glymour, et al. 2016. *Causal Inference in Statistics: A Primer*. Chichester: Wiley & Sons.

Rainville, P., G. H. Duncan, et al. 1997. "Pain Affect Encoded in Human Anterior Cingulate But Not Somatosensory Cortex." *Science* 277 (5328):968-71.

Rubin, D. . 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5):688-701.

Sekhon, J. S. 2008. "The Neyman-Rubin Model of Causal Inference and Estimation via Matching Methods." In *The Oxford Handbook of Political Methodology*, ed. J. M. Box-Steffensmeier, H. E. Brady and D. Collier, 271-99. New York: Oxford University Press.

Shadish, W. R., T. D. Cook, et al. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Spirtes, P., C. Glymour, et al. 1993. *Causation, Prediction and Search*. New York: Springer-Verlag.

Wager, T. D., L. Y. Atlas, et al. 2013. "An fMRI-Based Neurologic Signature of Physical Pain." *New England Journal of Medicine* 368 (15):1388-97.

Winch, R. F., and D. T. Campbell. 1969. "Proof? No. Evidence? Yes. The Significance of Tests of Significance." *The American Sociologist* 4 (2):140-43.

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

———. 2008. "Cause and Explanation in Psychiatry: An Interventionist Perspective." In *Philosophical Issues in Psychiatry: Explanation, Phenomenology and Nosology*, ed. K. Kendler and J. Parnas. Baltimore: Johns Hopkins University Press.

———. 2015. "Interventionism and Causal Exclusion." *Philosophy and Phenomenological Research* 91 (2):303-47.

9979 words

1. For example: the diagnosis and pathology of psychiatric conditions encompass physiological and psychological symptoms, risk factors and causal determinants; epidemiological studies reveal that many diseases are influenced by diverse variables, such as genes, diet, lifestyle and socioeconomic status; biopsychosocial models of disease emphasize multiple levels of medical intervention; biopsychosocial models of pain and fear are explained in terms of interactions among biological, psychological and social variables. [↑](#footnote-ref-1)
2. Woodward reiterates the fixability requirement in his response to Baumgartner: “M requires that for *X* to be a direct or contributing cause of *Y*, it must be true both that it is possible to intervene on *X* (there must “be” such an intervention) and that under this intervention, *Y* changes *when other specified variables be held fixed*” (Woodward 2015, 312, emphasis added). [↑](#footnote-ref-2)
3. The methodological literature repeatedly emphasizes that the logic of causal inference is one of systematic elimination of alternative explanations: “Determining whether there is a causal relationship between variables, *A* and *B*, requires that the variables covary, the presence of one variable preceding the other (e.g., *A* → *B*), and ruling out the presence of a third variable, *C*, which might mitigate the influence of *A* on *B*” (Leighton 2010, 622). Or again: “The temporal structure of the experiment ensures that cause precedes effect. Whether cause covaries with effect is easily checked in the data within known probabilities. The remaining task is to show that most alternative explanations of the cause-effect relationship are implausible” (Shadish et al. 2002, 249). [↑](#footnote-ref-3)
4. Following Pearl (2000), Woodward (2003, 95-96) assumes that an intervention fixes the values of the independent variable, thus simulating the effect of a random allocation intervention. This requirement is rather restrictive. Most experiments in biomedical research don’t fix variable values. For instance, overexpressing a gene doesn’t fix the concentration of a gene product. Notwithstanding, by choosing comparable test and control systems, researchers routinely demonstrate the causal relevance of gene expression to biological activity. The assumption of graph acyclicity is also a major idealization in the case of molecular mechanisms, which involve chemical equilibria and feedback structures. The implication here is that Woodward’s formulations of (M) and (IV) are too narrow, since they tie the definitions of manipulation and ideal intervention to causal modelling assumptions which cannot be satisfied in many experiments. This is not a fatal shortcoming for interventionism, but more a question of finetuning. For example, Eberhardt and Scheines (2007) provide a formal treatment of ‘soft interventions,’ which, unlike Woodward’s ‘hard interventions,’ change the value of a variable without breaking the causal arrows into that variable. [↑](#footnote-ref-4)
5. Fixability is a special case of comparability. If the test and control conditions correspond to an ‘after’ and a ‘before intervention’ state of a system, comparability dictates that the confounder-variables must take the same values in the ‘before/control’ and ‘after the intervention/test’ conditions, which is to say that they must stay constant over time. [↑](#footnote-ref-5)
6. Randomization is said to equate groups on the expected value of all variables at pretest, which simply means that test and control groups will differ only by chance. [↑](#footnote-ref-6)
7. INF stands for ‘independent fixability’ and presupposes the fixability clause stated in (M). [↑](#footnote-ref-7)
8. Consider, for example, Ronald Melzack’s (2001, 1378) proposal that pain “is produced by the output of a widely distributed neural network in the brain,” dubbed the ‘neuromatrix,’ which “is the primary mechanism that generates the neural pattern that produces pain. Its output pattern is determined by multiple influences, of which the somatic sensory input is only a part, that converge on the neuromatrix.” According to this proposal, pain is produced by a complex biological mechanism, which immediately suggests that pain cannot non-causally supervene with that which produces it. Thus, the recently discovered pattern of fMRI activity predicting whether a subject will report a heat stimulus as being painful or not (Wager et al. 2013) refers solely to the mechanism causing pain, not pain itself. It tells which structures in the brain should be monitored to measure pain or targeted by interventions to alter pain experience, but it tells us nothing about the supervenience of pain on a biological state. [↑](#footnote-ref-8)
9. Causal interpretations paradoxically coexist side by side with supervenience ones. For instance, physics textbooks are also in the habit of defining temperature as a measure of the average translational kinetic energy of the molecules of a gas. Under a causal-realist interpretation of measurement of the sort typically endorsed in experimental science, this definition entails that temperature is a measurable effect of kinetic energy in a specific experimental setup. [↑](#footnote-ref-9)
10. “[…] the requirement that in order for *M1* to have a causal effect on *M2*, there must be a possible intervention that changes *M1*while *P1* is held fixed and under which *M2* changes, is inappropriate. […] the requirements in the definition (IV) are understood as applying only to those variables that are causally related to *X* and *Y* or are correlated with them but not to those variables that are related to *X* and *Y* as a result of supervenience relations or relations of definitional dependence. Call this characterization of interventions (IV\*) and an intervention meeting these conditions an IV\*-intervention” (Woodward 2015, 333-34). [↑](#footnote-ref-10)
11. Since the human body can metabolize cholesterol in many other ways, *TC* is not, strictly speaking, a conserved quantity. Woodward’s example probably refers to Friedewald equation (*TC* = *HD* + *LD* + 20% triglyceride blood concentration), which is a method for estimating *LD* given measurements of the other variables. The practical value of the equation lies in the fact that it can act as a substitute for direct, but costlier, measurements of *LD*. The method is known to yield inaccurate estimates for certain patients, such as metabolic syndrome patients, hence the recommendation to opt for the direct test. Moreover, pharmacological agents designed to inhibit cholesterol synthesis (statins) are known to interfere with metabolic pathways regulating the balance between *LD* and *HD*. If a relationship between variables of the sort illustrated by Friedewald equation were indeed a definitional constraint, we would have to conclude that hereditary metabolic abnormalities and enzyme inhibitors can override matters of logical necessity. [↑](#footnote-ref-11)
12. I am not arguing that experimental practice is, can or should be insulated from prior assumptions. For instance, a statistical model embodies substantive claims about how data is generated (e.g., deterministically or probabilistically), thus offering a putative explanation of why data scatter in a particular way (Jaynes 2003). The key point, however, is that these assumptions are not dogmatically assumed to be true, but are adopted as working hypotheses, the consequences of which can be eventually tested (e.g., consistency or inconsistency with actual variations of measured values, the fact that measurement precision can or cannot be increased by improving measurement techniques and experimental setups, etc.). A similar argument can be made about thresholds of statistical significance. Changes in methodological standards in response to the replicability crisis (Ioannidis 2005) demonstrate that conventions are revised in light of empirical results. In contrast, the definitional and metaphysical constraints advocated by Woodward do not admit empirical testing. [↑](#footnote-ref-12)
13. In doing so, we may find out that some psychological variables don’t refer or don’t constitute natural kinds. As noted above, the variable ‘reported pain’ is multidimensional and some of these dimensions can be manipulated independently. This is compatible with the view that ‘pain unpleasantness’ may be associated with sensory dimensions of pain, such as ‘pain intensity,’ to the same extent visual experiences are associated with auditory ones (Hardcastle 1999). We may also find something entirely new and surprising. For example, deCharms et al. used real-time functional MRI (rtfMRI) to train subjects to control ACC activation. This is what they found: “When subjects deliberately induced increases or decreases in rACC fMRI activation, there was a corresponding change in the perception of pain caused by an applied noxious thermal stimulus. Control experiments demonstrated that this effect was not observed after similar training conducted without rtfMRI information, or using rtfMRI information derived from a different brain region, or sham rtfMRI information derived previously from a different subject. […] These findings show that individuals can gain voluntary control over activation in a specific brain region given appropriate training, that voluntary control over activation in rACC leads to control over pain perception, and that these effects were powerful enough to impact severe, chronic clinical pain” (2005, 18626). [↑](#footnote-ref-13)