Theoria, 2021, 36(2), 173-189
https://doi.org/10.1387/theoria.21757

THEORIA

ISSN 0495-4548 – eISSN 2171-679X

Mechanisms, good and bad[&]

(Mecanismos, lo bueno y lo malo)

Harold Kincaid
University of Cape Town (South Africa)

ABSTRACT: The claim that mechanisms are essential good science is widespread. I argue, however, that these claims are ambiguous in multiple ways. I sort out different version of the mechanism idea: (1) mechanisms that are horizontal —between cause and effect— and mechanisms that are vertical —they realize in lower-level terms causal properties—: and (2) different purposes or uses mechanisms may have. I then focus on the claim that various senses of mechanism are necessary for the confirmation of causal claims. The paper shows that mechanisms can be useful, essential, or harmful depending on context, using the now standard graphical causal structure framework. These conclusions also support the larger philosophy of science moral that methodological norms in science are often context specific and empirical, not a priori and universal.

KEYWORDS: mechanisms; causation; causal mediation; scientific norms.

RESUMEN: *La afirmación de que los mecanismos constituyen esencialmente "buena ciencia" está ampliamente extendida. Argumentaré, sin embargo, que esta afirmación es ambigua en varios sentidos. Distinguiré diferentes versiones de la idea de mecanismo: (1) mecanismos horizontales —entre causa y efecto— y mecanismos verticales —asientan propiedades causales en términos de bajo nivel—; y (2) diferentes fines o usos que los mecanismos pueden tener. A continuación, me centraré en la tesis de que los mecanismos son, en diversos sentidos, necesarios para la confirmación de enunciados causales. Sirviéndose del marco estándar hoy día —estructuras gráficas causales—, el artículo muestra que los mecanismos pueden ser útiles, esenciales o dañinos dependiendo del contexto. Estas conclusiones también respaldan una tesis más general en filosofía de la ciencia: que las normas metodológicas en ciencia son, con frecuencia, relativas al contexto y empíricas, y no a priori y universales.*

PALABRAS CLAVE: mecanismo; causación; mediación causal; normas científicas.

[&] This work builds on Kincaid (1996, 1997, 2011, and 2012b).

**** Correspondence to:*** Harold Kincaid. School of Economics, School of Economics Bldg, Middle Campus, University of Cape Town (South Africa, 7701) – Harold.kincaid@uct.ac.za – https://orcid.org/0000-0001-5032-815X

How to cite: Kincaid, Harold (2021). «Mechanisms, good and bad»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 36(2), 173-189. (https://doi.org/10.1387/theoria.21757).

Received: 2020-05-26; Final version: 2021-02-10.

ISSN 0495-4548 - eISSN 2171-679X / © 2021 UPV/EHU

The idea that mechanisms are importantly involved in good science is now widely asserted. However, the meanings of "mechanisms" and the rationales for them are still unclear in the literature. This paper sorts out theses and possible arguments and shows that mechanisms can be useful, essential, or harmful depending on context. It also offers some support for idea that methodological norms in science are often context specific and empirical, not a priori and universal. Thus, the goals of the paper are four: to clarify some of these claims about mechanisms, to show that strong claims about the necessity of mechanisms as evidence are implausible, to show that some weaker claims may be plausible but depend on contextual details, and to support the view that plausible methodological norms in science are often empirical and local.

The literature on mechanisms is enormous[1] and the potential issues many. I do not claim to address all the issues and corresponding commentary. To keep the topic manageable, I focus here on claims about evidence.[2] I also focus on practices standard in the social and behavioral sciences. However, the general framework I describe for analyzing the need for mechanisms as evidence could perhaps fruitfully be applied to claims about mechanisms and evidence in, for example, in medicine. But I will not support that claim here – it would deserve considerably more discussion.

Section 1 cites some of the standard claims in the literature about mechanisms and then Section 2 provides a matrix of meanings and rationales that demonstrates there are many different claims at stake and provides a specific set of questions that seem to be at issue when assertions are about mechanisms and evidence. Evidence about mechanisms in the form of mediating causes is the topic of Section 3. Section 4 looks at theses about mechanism as the realizing underlying details of causal properties. Section 5 finishes by assessing more moderate claims about mechanisms and identifies some situations where they are indeed plausible. Throughout the moral is that methodological claims about mechanisms in providing evidence are often local empirical issues which have to be evaluated case by case depending on the kinds of causal structures involved.

1. The Mechanisms Bandwagon

Enthusiastic claims that mechanisms are essential to good science currently abound. Philosophers are an important part of the chorus, but social scientists are advocates as well (Hedstrom and Swedberg, 1998). Both groups make strong and diverse claims on behalf of mechanisms.

Here are some typical examples:

> "[...] a plausible mechanism or theoretical connection is required to warrant the more general claim that "C causes E," probabilistic evidence needs to be accounted for by an underlying mechanism before the causal claim can be established, and to establish casual claims, scientists need the mutual support of mechanism and dependencies." (Russo and Williamson 2007, p. 159)

[1] See, for example, Craver and Tabery (2016) and Glennan and Illari (2018) for surveys.

[2] If explanatory power is evidence then the distinction between mechanisms for evidential reasons vs. explanatory reasons may call for added complexity. I do not think that inference to the best explanation actually adds additional evidence (see Day and Kincaid, 1990; Kincaid, 1996, Chapter 1) but this is a complication that is not necessary for the points that follow.

> "[...] evidence of mechanisms is crucial to establishing both efficacy and external validity." (Parkkinen *et al.*, 2018, p. 6)
>
> One "needs both evidence of correlation and evidence of mechanisms to establish a causal claim." (Parkkinen *et al.*, 2013, p. 4)
>
> The demand for mechanisms involves both "mechanisms (a complex arrangement of entities and activities, organised in such a way as to be regularly or predictably responsible for the phenomenon to be explained) and mechanistic processes (a spatio-temporal pathway along which certain features are propagated from the starting point to the end point)." (Parkkinen *et al.*, 2013, p. 11)
>
> "[...] all interpretation in terms of causality of macro-variable explanations based on variable correlations is considered unsatisfactory unless the action level mechanisms triggering the social phenomena can be identified." (Demeulaere, 2011, p. 17)
>
> "Efficacy is established if one can establish, in the study population, the existence of a correlation and the existence of a mechanism that can explain this correlation." (Parkkinen *et al.*, 2013, p. 24)
>
> It is "broadly correct" that "a causal hypothesis in medicine can be established only by using both statistical evidence and evidence of mechanism." (Gillies, 2019, p.133)
>
> "[...] the advancement of social theory calls for an analytical approach that systematically seeks to explicate the social mechanisms that generate and explain observed associations." (Hedstrom and Swedberg, 1998, p. 1)

The demand for mechanisms is not new. Elster (1989) asserted 30 years ago that mechanisms are needed in the context of debates about functional explanations in the social sciences to avoid confounding.[3] Currently, the demand for mechanisms is extended to the social and behavioral sciences and medicine and sometimes is an all-purpose claim about good science in general.

I should make clear several points about this literature and my project here. Not everybody who talks about mechanisms makes all these claims; there are more subtle views advanced and I shall mention them at various points. Also, my goal is not to do a survey of the literature. It is to sort out some confusions still present, to show that claims about mechanisms have to be evaluated according to empirical context, and to provide some concrete illustrations of this claim.

2. *Approaching the Plurality of Mechanism Claims*

As can be gleaned from the above quotations, there are numerous different and often logically independent claims and rationales about mechanisms. Here is a list of some of the different questions involved that need to be distinguished:

a) What scientific virtues do mechanisms provide or embody? There are distinct claims in the literature about whether mechanisms are used to provide evidence or are used to provide adequate explanations. Even if I have a well-confirmed ex-

[3] And on my view was properly put in its place not long after that (Kincaid, 1996, 1997).

planation, I might still think mechanisms are necessary to make the explanation deep, deep enough, deeper than some other well-confirmed explanation, and so on. Then looking at scientific practice as dynamic, one might claim as some have that searching for mechanisms is the best way to promote scientific progress, however that is specified.

b) To what degree are mechanisms needed? Are they essential? Always useful if not essential? Sometimes useful? "Usually" useful? If not essential, when and why are they useful?

c) At what level are mechanisms supposed to be given? It helps to distinguish between horizontal and vertical mechanisms (Kincaid, 1996; Kastner and Andersen, 2018).[4] Horizontal mechanisms roughly are those at the same level as the cause and effect. The simplest cases of horizontal mechanisms are intervening or intermediate causes or variables of the fashion: cause à mediator à effect, the notion invoked by Gillies above. Vertical mechanisms refer to the component elements and structure that explain the properties, activities, etc. of the whole they make up – this is the notion of mechanism in the Machamer *et al.* tradition (2000) cited above. These distinctions are still simplifications. In particular, we will see below that when it comes to evidence, horizontal mechanisms can importantly be not just mediating, "in between" causes, but they can also be other causal factors that influence the cause, mediators and/or the final effect. Also, vertical and horizontal mechanism can no doubt also be combined into a mixed, multilevel account.

d) What sort of scientific product do we want mechanisms to help with? If we talk of causes, do we want them for determining the presence of causal effects or causal effect *size*? These are not the same. Do we want them to explain, confirm, etc. token events, e.g. the result of this RCT, or to explain, confirm, etc. generalizations? Models?

e) How exactly does evidence about mechanisms integrate into our total evidence? It could just be "additive" – just one more piece of evidence, it could trump or have more weight than other pieces of evidence, it could serve specific inferential functions, such as ruling out confounders, and so on.

More parameters could be identified. However, this list already suggests numerous different possible claims about mechanisms, many of them independent and some interdependent in various ways. Proceeding without being specific about which of these many claims are being discussed, defended, or criticized is asking for confusion. Furthermore, several of these elements can mean different things in different contexts and/or admit of subtypes and thus produce still further, different questions. I narrow the window somewhat in what follows by addressing only the epistemological issues raised.

With the above set of questions in view, it is helpful to have an explicit approach to methodological norms in science – of assertions like those above about how science should

[4] There is an orthogonal distinction which takes mechanisms to be some deeper cause that somehow makes standard causal claims possible – a background structure as it were (Mahoney 2016). I am not sure what this comes to as suggestive as it is. It may be related to the idea of a necessary cause – one which is required but insufficient by itself. Throughout my discussion I am focusing on standard sufficient causes —ones that have an effect on their own— not because I think the social world is always like that but because such considerations add more complexity than I can take on here.

work. In what follows I adopt a view that might be labelled "contextualism" (Williams, 1999; Kincaid, 2005). This approach is not the analytic philosophy doctrine that wants to give the necessary and sufficient conditions of our use of "know". Rather, contextualism in my sense is an antifoundationalist, pragmatist idea that emphasizes among other things that what we count as evidence depends on what else we know and what we are trying to do. Judgments about good science then are ultimately empirical claims.

Contextualism applied to debates over mechanisms suggests that there will be few defensible general scientific norms about them. Methodological norms, like all claims, are defeasible and ultimately empirical. Given the holism of evidence and theory, when and where those norms apply depends on context. Relatively general formal rules for inference are to be applauded if we can get them. In some specific situations we can do so, as for example we can do sometimes for causal inference as I will discuss below. Yet, broad and substantive morals about good science are unlikely, and that holds for morals about mechanisms. I develop this claim throughout the paper.

A natural set of questions for thinking about the role of mechanisms in confirmation flows from the contextualist approach above.[5] Assume I am asking whether I should believe that H, given some new evidence E. Making that decision would involve asking at least the following questions:

1. How much prior evidence do I have that H?
2. How much does my claim that H presuppose about mechanisms?
3. How much evidence do I have about purported mechanisms?
4. How well does H fit with E?
5. What alternatives are there to H, and how well do they answer 1. - 4.?

These questions could be given a Bayesian gloss as getting at p(H), p(not H), p(E/H) and p(E/not H), but that rendering is unnecessary. These questions can equally be seen as just instantiations of the principle of total evidence. Details will have to be fleshed out case by case with the parameters noted above, e.g. it needs to be specified whether we are talking about horizontal or vertical mechanisms, if the causal claim is one of effect or effect size, and so on. The questions obviously have to be weighted if we were pretending to be precise. But for a rough judgment, it is fairly obvious that mechanisms can play a large role or a small, depending on our answers.

In the mechanisms literature a common situation concerns the status of a causal claim based on a data set of correlations. Smoking correlates with lung cancer in a specific set of observations. At the same time there are various pieces of evidence relevant to a possible mechanism linking smoking and lung cancer. The issue then is the relative role of the correlational and mechanistic evidence. That requires answering the five questions above where H is a causal claim and we must decide how our answers combine. This formulation of the issues is the focus here.[6]

---

[5] I sketched this framework some time ago (Kincaid, 1996) and more recently others have come to roughly similar conclusions, though with less clarity I believe (Illari, 2011).

[6] Causation here is sufficient causation where causes individually and independently contribute effects. More complex types of causation where there are necessary conditions and causal interaction —moderation of one causal effect by the level of another— are not considered but should be in a longer, more complete discussion.

3. *Horizontal Mechanisms as Essential for Evidence about Causal Effect Claims*

Keeping in mind the distinctions and questions just discussed, I turn now to look at one common version of the demand for mechanisms. These demands have to specify how much mechanistic evidence counts and what kind of mechanism claim is being invoked. Recall that we are investigating only claims about evidence.[7]

The strong claims for mechanisms as found in Russo and Williamson (2008), Parkkinen *et al.* (2018), or in Gillies (2019) basically seems to assert that the answer to the fourth question should be one of strong evidence and that an affirmative answer is a necessary condition to establish that C causes E or, perhaps, is a necessary condition to establish effect size (though this difference goes unnoted). Thus, it does not matter how positively we answer the first three questions, if we do not have strong evidence for a mechanism, then we do not have good evidence that C causes E.[8]

To evaluate this demand for mechanisms, we have to fill in further details as we showed in the previous section. Though the literature is consistently unclear on some key parameters, the core claim is:

> HME: Horizontal mechanisms taken as intervening or mediating variables are essential to establish a causal relation between correlated factors C and E.

This claim, while more explicit than some formulations in the literature, is still unclear. It is quiet on whether this is about establishing a causal relation or a causal effect size or both? It is also unclear on how mechanisms are needed – what role do they play in the inferential process? The claim is also unclear because the literature vacillates on "essential," sometimes qualifying with "usually." We will return to these ambiguities later. For now, I focus on the claim that mechanisms are essential evidence to show that an effect exists.

What is the argument for HME? Often, it is just stated as an obvious basic methodological principle or "thesis" (Russo and Wiliamson, 2007). Frequently, it is suggested that mechanisms are needed to avoid confounding when our data is correlational. For example, Trampusch and Palier (2016, p. 437), two political scientists, say that finding mediating factors is "predominantly defined as a method aiming to identify or test hypotheses on causal mechanisms in order to compensate for weaknesses in correlational analysis. The method is even sometimes viewed as the 'only method.'"

HME is implausible, as I show now. There are two serious regress problems that confront HME. Moreover, it contradicts both good informal and formal causal reasoning.

There are two obvious related regress problems for HME. The first is that a mediator is defined as something that causes the final effect. So, we have another causal relation when we identify a mediator between C and E – do I then have to show how it, the mediator,

---

[7] I am furthermore limiting my discussion directly only to what might be called internal validity. As the quotes show, it is asserted that mechanism are essential for external validity, i.e. generalizing from one causal situation to another. I think the arguments given here hold of that claim as well but that is beyond my purview here.

[8] There are also suggestions about sufficiency: if we have strong evidence about mechanisms and correlational data supporting C's effect on E, then we have strong evidence for the claim of C's effect on E. This will be indirectly addressed later.

causes the final effect? The obvious question is where do we stop and why? We seem to be forced to provide indefinitely many intervening mechanisms.

The force of the regress here is that there is an apparent claim that for any assertion A causes B, the mechanism in the form of an intervening variable M has to be provided. However, that then entails that A causes M and M causes B. These seem to be causal claims to which the mechanism requirement would thus apply and hence we need another set of mechanisms. Of course, in practice we find reason to stop the regress. However, that just means we do not apply a methodological norm about mechanisms universally, but that there are contexts, background knowledge, purposes of investigation, etc. involved.

The second regress problem concerns how mechanisms get established in the first place. The literature talks of "mechanistic evidence" as though it was something that was *sui generis* and that does not require arguing from correlations. Thus, Russo and Williamson (2007, p. 159) say: "Evidence is constituted by two complementary elements: probabilities and mechanisms." Yet, that claim is false – experimental reasoning involves arguing from correlations or probabilities between interventions and outcomes. In fact, defenders of MHE find themselves sometimes arguing that RCTs need mechanisms, but RCTs are *experiments* in the real sense of the term. There is no correlation-free independent mechanistic evidence that can ground correlations – correlations, mechanisms and inference apparently are an interdependent package. Thus, the argument for HME cannot be about any causal claim simpliciter. If causal claims require mechanistic evidence instead of correlational evidence alone, but mechanistic evidence essentially uses correlation evidence itself, then that evidence needs further mechanistic evidence, and so on. Instead of the claim that correlational evidence always has to be supported by mechanistic evidence, a more guarded context-specific claim seems to be needed: this correlational data needs this specific background knowledge about mechanisms to infer a causal effect.

Aside from regress problems, HME also seems to contradict good causal reasoning, both common sense and formal. Common sense would seem to say that we can have good evidence that the flying ball broke the window or turning the ignition caused the car to start even if we do not know the intervening causes. In addition, our best current formal causal reasoning —that using causal graphs and deriving from them predicted dependencies and conditional independencies in the data (Pearl, 2009)— makes no requirement on mechanisms beyond the vacuous claim that causal accounts need to specify the causes involved. A causal graph like those depicted below in Figure 1 entails a specific set of independencies; data consistent with those relations provide support. The result is evidence for causal claims, and there is nothing in this well-worked out logic that r*equires* intervening variables.

So, general arguments for HME like the above seem not to work. Perhaps more specific considerations are compelling? As we saw, a major motivation for mechanisms by both philosophers and social scientists is to avoid confounding. Correlational evidence typical in many sciences can be a bad guide to causes because a correlation may just be the result of a common cause. Mechanisms are supposed to help solve this problem. We have already raised doubts about this assertion in that knowledge of mechanisms must likewise depend on correlational evidence, so the confounding is unlikely to be ruled out without further correlational evidence. But put that aside. Is it the case that considering mechanisms, currently in the sense of intervening or mediating causes, is essential to avoiding confounding in inferring causal claims?

Following my contextualist approach, I think this question is unanswerable in this form. We need to get more specific. I would suggest that "confounding" is at least a 3-place relation. It is common to talk about confounders just as objects in the world or as variables. But "confounding" is an epistemological notion. So, the claims are really of this sort: "Inferences about causal hypothesis H relying on specific inference method or rule R lead to error in the presence of factors of type F." As we will see, it can be shown (and has been) that this role of context can make a big difference in evaluating claims about mechanisms.

The standard approach in the social and behavioral sciences for looking for confounders —"controlling"— is statistical conditioning. The simple case best known to philosophers of science is "screening off." If the correlation between C and E in Figure 1 goes away when we condition on a potential confounder (CON) —when we look at C and E for values of CON— then the correlation between C and E may be spurious because CON is their common cause. This approach is embodied in the widespread practice of looking for confounders by using multiple regression or related tools. Standard practice across the social and behavioral sciences is to regress C on E and add in other suspected confounders. Conditioning on CON, the thought is, will thus control for them. In this guise, HME would find wide, though not universal, acceptance in the social and behavioral sciences, for example.

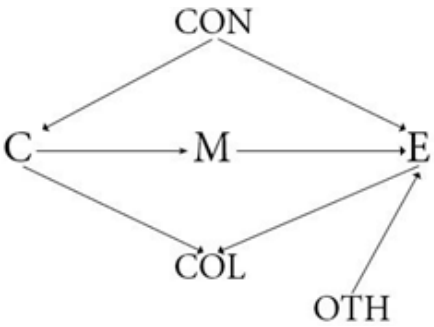

Figure 1

A directed acyclic causal graph where arrows indicate causes and causal direction and CON stands for confounding cause, M stands for a mediating or intervening cause, COL stands for a collider, and OTH stands for other independent causes.

Thus, the potential situation is that we have a data set with identified correlations between a set V of variables. Our background knowledge tells us that variable C in V is a likely cause of variable E. We then, following standard practice, model this with the linear equation:

$$E = bC \quad (1)$$

To test (1) we add a constant *a* representing the intercept on the y-axis, *e* which is an error term representing either measurement error and/or unknown causes, and then a vector of control variables Z to produce the multiple regression equation:

$$E = a + bC + Z + e \quad (2)$$

This sort of equation is the most dominant approach to quantitative analysis in the social sciences, with various statistical tests such as ordinary least squares being used to evaluate them. Multiple regression equations are everywhere in the social sciences.

To give a concrete example, there is a huge literature in economics on the determinants of economic growth. It works by taking national data from most countries in the world on a variety of variables along with measures of annual GDP. That data is then used for statistical tests based on multiple regressions like that described here (see Kincaid, 2021).

The mechanistic thought is that we escape the worries about such approaches that correlation is not causation because of confounding by specifying the real direct causal process relating E and C under controlled conditions. Unfortunately, HME in this form is not just wrong but actually sometimes epistemically pernicious. The bad effects here, recall, are for HME embodied in multiple regression methods, the work horse of much social scientific research. Over the past twenty years or so the potential problems have been raised and then repeatedly demonstrated (Pearl, 2009). The general practice of adding control variables to regressions beyond those thought to be the causes of interest results in identifiable errors in inference.

The errors are easy to see once pointed out, even if it took literally more than a century of statistical confusion to get there. For any given member of Z in (1), it might represent one of the following possibilities:

a) a common cause CON causing both C and E.
b) a mediating variable M between C and E.
c) a collider variable COL where both C and E cause COL.
d) another variable OTH that effects E and is neither a common cause, mediator, or collider.[9]

These are illustrated in the diagram in Figure 1. Figure 1 describes a "directed acyclic graph" because the arrows show the direction of causation and no causes circle back to themselves. This causal graph, *without controlling any of the variables*, entails dependencies or associations only between variables linked by arrows. The graph in Figure 1 only entails:

$$C \perp OTH$$

It is only when we control or fix variables in Figure 1 that new associations result or are eliminated; these are set by the logic of causal graphs.

Thus, in Figure 2 in the first case —(a)— CON is controlled for as represented by its shading. Now we know that:

$$C \perp E,$$

if these are the only variables. It is this situation that fuels the mechanistic thought that controlling variables will reveal the real causal process and eliminate confounding. Indeed, it does so but *only if* (a) is the case and no other variables are at issue except CON.

[9] Things are actually a bit more complicated for parents and ancestors of the three types of factors may be relevant; such detail is not needed for my point, however.

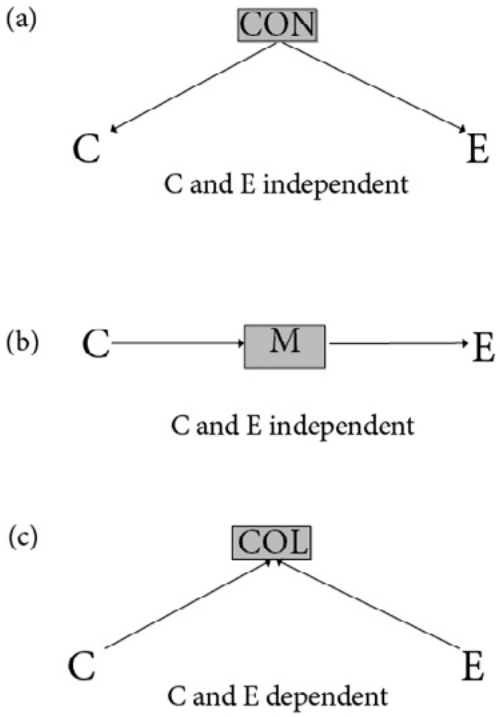

Figure 2

Illustration of common cause (a), mediating cause (b), and a collider (c) where these have been controlled for (indicated by gray shading) and the corresponding implications about dependency between the variables.

However, when the members of Z are either mediators or colliders (panels b and c in Figure 2), including Z leads to systematic errors in causal inference. In the collider case (c), holding COL fixed creates a correlation between C and E even if there were no causal relation as there is in Figure 1. In the mediator case, (b) of Figure 2, holding M fixed eliminates the correlation between C and E even if there is a real causal relation. So now $C \perp E / M$.

Moreover, these problems potentially extend even further. Automatic adding of possible mechanisms really opens up indefinite possibilities for making biased inferences about causes. If Z includes the effect of a mediator, then conditioning on that effect fixes M even if it is not in Z. Then the C-E correlation goes to zero just as it we had conditioned on M. Similarly, if Z includes the cause of a collider —its parent in the language of causal graphs— a spurious, noncausal correlation is created between C and E even though the collider itself is not controlled.[10]

[10] Actually, the possibilities for errors do not stop here. If I condition on a collider whose causes are U1 and U2 and U1 causes E and U2 causes C, I introduce a spurious correlation between C and E. This happens even if I know the M between C and E.

We can give this destructive conclusion about mechanisms a constructive spin. While it is wrong that we always have to have the mechanism for causal inference, it is true that we need to know what is *not* the mechanism when regression is our empirical tool; we need to know that our variables are not causal mechanisms between C and E and that C and E do not have a common effect. Knowing that much causal structure is important.

A similar story —that shows we do not need the mechanism— can be given about the need for mechanisms which unobserved common causes suggest (panel a in Figure 2). Unobserved causes can be potential confounders by creating spurious correlations by causing two otherwise independent variables. However, I can eliminate a possible confounded correlation by ruling out common causes while not knowing if any of them are real. One way I can do so is by holding constant —here, regressing on— a *suspected* confounder and showing that the correlation between C and E stands. Complete confounding of the C to E causal relation is thus ruled out for that possible situation, and I can then repeat the procedure for other worrying variables.

There are still other methods for dealing with common causes in this indirect way. Results in formal causal modeling such as instrumental variables or the use ancestral graphs (Angrist and Pischke, 2008 for instruments and Pearl, 2009 for both) are ways to deal with certain types of possible common causes that are not observed; it can be shown that in certain circumstances the causal inferences are reliable even if there are common causes and we do not know what they are. I will not go into details here. However, the moral in all these cases is that HME is implausible as a general methodological norm.[11]

4. Claims About Vertical Mechanism

We earlier distinguished horizontal from vertical mechanisms. Vertical mechanisms are the realizing detail, the microlevel components of a given whole. Much of the general literature on mechanisms (Machamer *et al.*, 2000) is as much about these part-whole or micro-macro connections. No doubt there is much of value in looking at these relations for a better understanding of scientific practice and in particular of explanation. The question here, however, is what role these lower-level details play in confirmation. Philosophers working in the mechanism tradition are less clear that vertical mechanisms are essential or primary in confirmation; their concern is largely about explanation and discovery. However, as I point out below, many social scientists have claimed that microfoundations are essential for any well confirmed theory. So, claims about necessary vertical mechanisms are important to consider.

The obvious parallel strong thesis about vertical mechanisms is:

> VME: Vertical mechanisms taken as lower-level or realizing factors are essential to establish a causal relation between correlated purported factors C and E.

[11] I have given arguments about mechanism in establishing that there is a causal effect. Looking at mechanisms as necessary for effect size raises similar issues that I do not have space to detail here. However, it is obvious, for example, that total causal effect size is biased up from controlling for a collider and is biased down in controlling for a mediator.

The intuitive story supporting VME comes from the undoubted progress in science by explaining traits at one level by seeing what brings them about at another level. The gas laws as macroscopic claims, for example, are explained by facts about the state of the molecules that compose them. The functioning of cells is explained via their molecular components. By explaining macroproperties, the realizing details tell us how macrolevel causes work. At least that is the thought.

Lurking in the shadows here are a variety of reductionism issues. I will skirt them as too fraught for the space here in hopes of noting some less complex considerations. VME is subject to some of the same difficulties that confronted HME. Both regress problems are applicable. Reminiscent of the "its turtles all the way down" joke, we can ask what level of underlying detail is needed and why. As we will get to shortly, it is plausible that understanding the realizing micro details of macro causes may increase our confidence in macro claims. However, VME asserts that such details are *necessary*. So, either we need an explanation why some level is basic or why the regress is not a problem. We are confronted with the prospect of quantum mechanics being essential for macroeconomics. I am not claiming that there may not be ways of stopping or averting such regresses, simply pointing out that it is a problem that needs addressed.

The second regress mentioned above surfaces as well. "Mechanistic evidence" is not some separate source over and above observed correlations and real or possible interventions. A nice example comes from debates over "microfoundations" in economics (King, 2014). Economists have been skeptical of macroeconomic causal claims for reasons like those motivating strong mechanism theses among philosophers: the macro level causal connections seem too tenuous on their own. Econometric models of aggregate data are subject to a seeming unending list of potential missteps.

However, one standard microfoundations for macroeconomics describes the behavior of *firms* and *households*. But how do we know about them? We rely almost entirely on correlations in observational data, with the best inferences often coming when nature produces those correlations by "natural experiments" – by something mimicking an actual experimental intervention. Economists probably have in the back of their minds finding yet a deeper, more secure turtle. At one time, it was thought that fundamentals of individual choice were a priori introspective truths. That would stop the regress. However, we now know that no such foundation exists. There is no escaping reliance on correlation of observed changes, no special "mechanistic evidence" at the lower level that somehow is more certain than observed changes and correlations.

VME also conflicts, like HME, with common sensical and formal causal knowledge. On the everyday side, we know that flying rocks break windows in full ignorance of molecular details. On the formal side, we know that in our some of our best worked out causal semantics and modeling, realizing detail can be just what we do *not* need.

I use again the results from graphical analyses of causal claims and evidence. Central to the semantics of causal graphs —to tracing their implications about independencies between variables— is the assumption that individual variables are separate or modular in that we can analyze the effect of setting them —possibly hypothetically— to a particular value. That means in effect that we remove all the causal arrows coming into the variable to be set. So, we transform the causal graph in

Figure 1 into that of Figure 3 – we wipe out the arrow from C to M and fix M at some realized value *m*. This is the central core of graphical causal modeling.

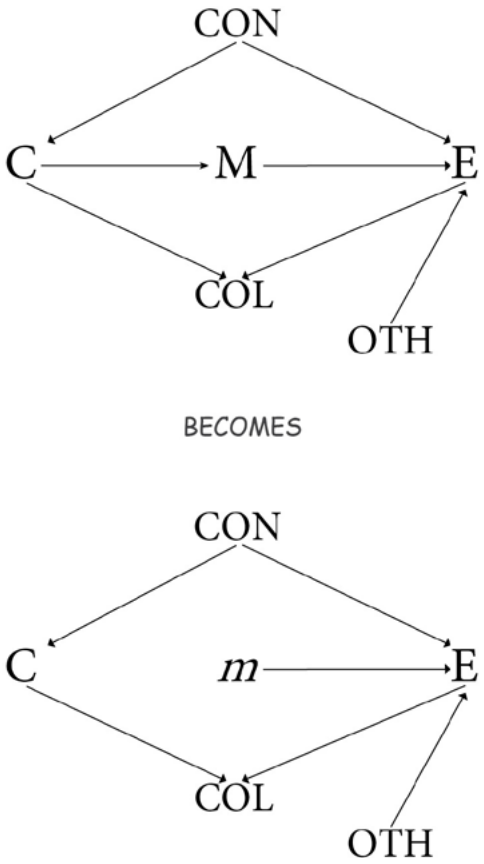

Figure 3

The causal relations described in Figure 1 have been transformed by an intervention, real or hypothetical, that removes the causal relation from C to M. The causal arrow in the graph has been "wiped out" and M replaced by *m*, an instance or token realization of the variable M. The corresponds to Pearl's *do* operator and is essential for the semantics of drawing inferences from graphs. Logically related variables cannot be so transformed.

However, including realizing variables may violate these semantic conditions because there is a logical or analytic connection between a variable quantifying over a whole and one quantifying over the parts that make it up. I cannot set the one to a particular value without setting the other. This means I cannot use the semantics of causal graphs to infer things about these variables that I need to determine causal effects. A different but related problem comes from the statistical regression models used. If two independent variables in a multiple regression are perfectly correlated, the value of the regression coefficients are undefined. Multiple regression finds the value of dependent variable when one independent variable is changed while all other variables stay the same. If two independent variables are logically related, that cannot happen. When properties of the whole and properties of the

parts realizing them are both independent variables we may be trying something not possible.[12]

These problems with VME arise in using two widespread testing procedures across the social and behavioral sciences. They do not show that vertical mechanisms always cause errors, only that they can in some standard circumstances. Again, the need for mechanisms depends on local detail.

5. Moderate Claims about Mechanisms

We have seen that HME and HVE are quite implausible as general theses. I look in this last section at more modest epistemological claims about mechanisms.[13] Generalizations are generally difficult: the force of evidence about mechanisms usually depends on specifics.

Following the five questions of Section 2 but talking now of hypotheses in the abstract rather than causal claims and of evidence rather than data sets of correlations, we might ask about hypothesis H and mechanism M:

1. How much evidence do I have for H, not considering knowledge of M?
2. What does H presuppose about M?
3. How much evidence do I have about M?

These questions are in a sense trivial, since all they ask for is combing evidence and considering logical or theoretical connections among claims. Details have to be fleshed out case by case with the parameters noted above in Section 2, e.g. it needs to be specified whether we are talking horizontal or vertical mechanisms, if the causal claim is one of effect or effect size, and so on.

The general morals about mechanisms from these questions seem clear:

> *Mechanisms can count for much*: If H makes strong assumptions about M, if H is only weakly supported on other grounds, and if we have strong evidence that M cannot be as H assumes or strong evidence that M is as H presupposes, then evidence about mechanisms is epistemically important.
>
> *Mechanisms can count for little*: If H makes weak assumptions about M, if H is well supported on other grounds, and if our knowledge of M is limited, then evidence about mechanisms is not epistemically important.

---

[12] These problems were noted in the causal modeling literature some time ago. A parallel but independent discussion has occurred more recently among philosophers of science in debates over interventionist notions of causes and the explication of constitutive relations between wholes and realizing mechanisms (Kastner and Andersen, 2018). There are interesting open issues raised by this literature which I do not take up here.

[13] Defenders of strong claims about mechanisms sometimes slip in qualifiers such as "usually." I have no idea how to specify what the population of uses would be in order to calculate the frequency of need for mechanisms and no idea what criteria to use to decide which percentage use would constitute "usually," so I am not going to try to sort out such hedges. On my view a "usually" qualification suggests that not much is being said and that we need to get more specific and spell out concretely where and why mechanisms do and do not matter.

These morals are, of course, disappointing if you are yearning for substantive methodological norms or rules. To the question whether mechanisms are essential, these rules just say that "it all depends...". This is no surprise on the contextualist view, however. That does not mean there are no interesting questions about mechanisms nor that there are never circumstances where they are essential. The call instead is for specifics and concrete detail if we want to make a plausible case.

We have already seen, in a negative sort of way, specific situations where knowledge about mechanisms can be important, depending on what the causal world is like. There are numerous further interesting circumstances where mechanism matter a lot. I want to finish by detailing some examples.

The graphical causal framework suggests further ways adding mechanist detail can increase confidence in results. For example, mechanisms can help increase the strength of evidence because they raise the number of independent causal processes that need to be explained by a hypothesis of confounding. Suppose I propose that C causes E and defend that assertion on the grounds of correlation between the two in a given data set. The usual worry is that there is some third variable CON that, as the name suggests, tricks us by being the common cause of C and E which in fact are not causally related. However, suppose further that I identify a possible mediating variable M and show that C and M are correlated in the data as are M and E. Now the ante is upped for the friend of CON. CON must now be the common cause of all three variables. Another set of correlations has to be found. Proving CON is now a more stringent affair. This point illustrates the common insight that a more detailed theory faces a tougher test than a looser one and correspondingly gets more credit when it passes.

There are no doubt other situations where we can show that knowledge of mechanisms in one of its many forms adds to evidential strength in some contexts, as the defenders of mechanisms have emphasized and illustrated. Steele (2008) and Jiménez-Buedo and Squitieri (2019) have discussed situations where macrovariables may be less well measured or more complicated to observe and thus where lower-level mechanism potentially add value. Social scientists, particularly political scientists (Trampusch and Palier, 2016; Bennett and Checkel, 2015), recently have been emphasizing "process tracing" as a concrete embodiment of the mechanistic idea. Case study work motivates the thought that showing how one thing leads to the next is important; social scientists claim that it can be done without the large -N type of correlations discussed above. The closest political scientists get to a logic of process tracing invokes Bayes theorem (Mahoney, 2016). Not surprisingly, however, all the work goes into giving evidence for the priors and likelihoods. How that evidence is supposed to be spelled out is still a very open question.

6. Conclusion

The debate and literature on mechanisms has raised a number of important issues. My argument has been that discussing mechanisms in general and in the abstract can only get so far. Many different things are meant by mechanisms; their rationales are various. Here I have taken steps to sort out differences in claims and differences in how they are supported. No doubt more work can be done in this direction.

I have tried to assess a subset of those possible claims about evidence. To do that I have looked at causal claims and recent advances in causal modeling in the social and behavioral sciences. Doing so allows some claims around mechanisms to be made more precise and decidable. Horizontal mechanisms and vertical mechanisms as necessary conditions for confirmation are very implausible. Less demanding claims that give mechanisms an important role in some contexts can sometimes be clearly specified and can be plausible. Progress in thinking about the role of mechanisms in science may best come from turning to details of scientific practice such as these.

Acknowledgements

Two anonymous referees made very helpful comments and as did the audience at Fourth Meeting in Philosophy, Probability and Scientific Methodology: Statistical evidence and evidence of mechanisms at the University of Valencia in 2019.

REFERENCES

Angrist, J. and Pischke, J. (2008). *Mostly harmless econometrics*. Princeton: Princeton University Press.

Bennett, A. and Checkel, J. (2015). *Process tracing: From metaphor to analytic tool*. Cambridge: Cambridge University Press.

Clarke, B., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi,* 33(2), 339-360

Craver, C. and Tabery, J. (2015). Mechanisms in science. In E. Zalta (ed.). *Stanford Encyclopedia of Philosophy* (Summer 2019 edition). https://plato.stanford.edu/archives/sum2019/entries/science-mechanisms/

Day, T. and Kincaid, H. (1994). Putting inference to the best explanation in its place. *Synthese,* 98, 271-295.

Elster, J. (1989). *Explaining technical change*. Cambridge: Cambridge University Press.

Gillies, D. (2019). *Causality, probability, and medicine*. London: Routledge.

Glennan, S. and Illari, P. (2018). *The Routledge handbook of mechanisms and mechanical philosophy*. London: Routledge.

Hedstrom, P. and Swedberg, R. 1998. (eds.) *Social mechanisms: An analytic approach to social theory*. Cambridge: Cambridge University Press.

Hedstrom, P. and Swedberg, R. (1998). Social mechanisms: an introductory essay. In P. Hedstrom and R. Swedberg (eds.). *Social mechanisms: An analytic approach to social theory* (pp. 1-31). Cambridge: Cambridge University Press.

Jiménez-Buedo, M. and Juan Carlos Squitieri (2019). What can mechanisms do for you? Mechanisms and the problem of confounders in the social sciences? *Philosophy of the Social Sciences,* 49(3), 210-223.

Illarji, P. (2011). Mechanistic evidence: Disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science*, 25(2), 139-157.

Kastner, L. and Anderesen, L. (2018). Intervening into mechanism: Prospects and challenges. *Philosophy Compass*, 13(11). DOI: 10.1111/phc3.12546

Kincaid, H. (1997). *Individualism and the unity of science: Essays on reduction, explanation, and the special sciences*. Lanham, MD: Rowman and Littlefield.

Kincaid, H. (1996). *Philosophical foundations of the social sciences: Analyzing controversies in social research*. Cambridge: Cambridge University Press.

Kincaid, H. (2005). Contextualism, explanation and the social sciences. *Philosophical Explorations* 7, 201-219.

Kincaid, H. (2011). Causal modeling, mechanism, and probability in epidemiology. In P. Illari, F. Russo, and J. Williamson (eds.). *Causation in the sciences* (pp. 70-91). Oxford: Oxford University Press.

Kincaid, H. (2012a) (ed.). *Oxford handbook of the philosophy of the social sciences*. Oxford: Oxford University Press.

Kincaid, H. (2012b). Mechanisms, causal modeling, and the limitations of traditional multiple regression. In H. Kincaid (ed.). *Oxford handbook of the philosophy of the social sciences* (pp. 46-64). Oxford: Oxford University Press.

Kincaid, H. (2021). Making progress on causal inference in economics. In H. Kincaid and D. Ross (eds.). *A modern guide to philosophy of economics*. Cheltenham, UK: Edward Elgar.

King, J. (2014). *The microfoundations delusion: Metaphor and dogma in the history of macroeconomics*. Cheltengam, UK: Edward Elgar.

Machamer, P.K., L. Darden, and C.F. Craver, (2000). Thinking about mechanisms, *Philosophy of Science*, 67, 1-25.

Mahoney, J. (2016). Mechanisms, Bayesianism, and process tracing. *New Political Economy*, 21(5), 493-499.

Parkkinen, V.-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M.P., Norell, C., Russo, F., Shaw, B., Williamson, J. (2018). *Evaluating evidence of mechanisms in medicine: Principles and procedures*. Amsterdam: Springer.

Pearl, J. (2009). *Causality*. Oxford: Oxford University Press.

Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21, 157-170.

Sperber, D. (2011). A naturalistic ontology for mechanistic explanations in the social sciences. In P. Demeulenaere (ed.). *Analytical sociology and social mechanisms* (pp. 64-77). London: Routledge.

Steel, D.P. (2008). *Across the boundaries: Extrapolation in biology and social science*. New York: Oxford University Press.

Trampusch, C. and Palier, B. (2016). Between X and Y: how process tracing contributes to opening the black box of causality. *New Political Economy*, 21(5), 437-454.

Williams, M. (1999). *Groundless belief*. Princeton: Princeton University Press.

Harold Kincaid is Professor in the School of Economics at the University of Cape Town. He is the author and editor of numerous articles and books in the philosophy of social science and philosophy of science more generally. His current research interests are in causal modeling and in experimental economics, especially of risk time attitudes, social preferences, and addiction.

Address: School of Economics, School of Economics Bldg, Middle Campus, University of Cape Town, South Africa, 7701. E-mail: Harold.kincaid@uct.ac.za. ORCID: orcid.org/0000-0001-5032-815X