To appear in *Top-Down Causation and Emergence*, eds. J. Voosholz and M. Gabriel (Springer)

Causation as a high-level affair

Simon Friederich[1] and Sach Mukherjee[2]
[1]University of Groningen, University College Groningen
[2]Statistics and Machine Learning, DZNE, Bonn

Abstract: The causal exclusion argument supports the notion that causation should be thought of as a purely low-level affair. Here we argue instead in favour of high-level causation as a natural and meaningful notion that may even be *more* useful than causation at more fundamental physical levels. Our argument is framed in terms of a broadly interventionist conception of causation. Its essence is that causal relations at an appropriately high level can in a certain sense be less sensitive than those at a fundamental, microscopic level. This means that in settings where causal relations at the (micro-) physical level are not considered in the context of some suitable macro-level interpretation, statements concerning the low-level relations may be highly sensitive with respect to changes in background conditions. Using an example of accelerator experiments in particle physics, we consider what it means to characterize extremely sensitive low-level events as causal.

1. Introduction

Identifying cause-effect relations is central to disciplines such as psychology, medicine, biology, sociology and economics that focus on relatively high-level (as opposed to micro-physical) phenomena. According to some philosophers, however, causation is in fact a pure low-level affair, i.e. it is confined to the most fundamental level of elementary physics that exists.

This is, notably, the conclusion of the causal exclusion argument, championed by Jaegwon Kim (1998, 2005). It applies on the condition that the plausible doctrine of *non-reductive physicalism* is correct: the view that higher-level properties supervene on lower-level physical (micro-) properties, but are not identical to them. Physicalism entails that (micro-) physical properties are not affected by any non- (micro-) physical causes over and above their (micro-) physical causes. And supervenience entails that fixing the (micro-) physical properties of the world fixes all other properties. Accordingly---the argument roughly goes (there are various versions of it in the literature which use somewhat different assumptions)---unless one implausibly accepts that higher-level facts are causally overdetermined by lower- and higher-level facts, only the more fundamental physical properties can be genuinely causally effective. So unless one assumes that higher-level properties are actually themselves identical with lower-level properties, they cannot possibly be causally effective. If the causal exclusion argument is correct, to assume that there is any higher-level causation requires accepting interactionist dualism or austere reductive physicalism.

In this contribution, we argue for a quite different position: namely, that causation not only applies at higher levels, but that it can be a more natural notion at higher levels than at the more fundamental physical levels. As the backdrop of our argument we use a broadly interventionist conception of causation such as that invoked by Pearl (2009) and Woodward (2003). We note that our argument does *not* contradict the physicalist position: rather we focus attention on the interpretation and sensitivity of causal claims at higher and lower levels, but within a framework in which in the end any specific higher-level causal effect operates via some sequence of micro-physical events.

The structure of our argument is as follows: in section 2, we recapitulate extant interventionist responses to the causal exclusion argument. In doing so, we highlight that, even though these responses assert the existence of causal relations at the (micro-) physical level, we typically have no independent grip on the micro-physical variables between which those relations obtain beyond their playing a role in this argument.

Next, in section 3, we discuss the notion of sensitivity as it relates to causation and the level of hierarchy at which causation is invoked. We argue that causal statements at an appropriately high level can in a certain sense be less sensitive than those at a micro-physical level. Our argument for the robustness of high-level causal claims focuses on the background conditions with respect to which causal relations obtain. Following Lewis (1986) and Woodward (2006), we call causal relations that obtain for a wide range of background conditions *insensitive* and those that obtain only for a narrow range *sensitive*. As pointed out by Woodward, the more sensitive some purportedly causal relation is, the less appropriate it seems to describe it as genuinely “causal”. Finally, in section 4, we argue that, where statements concerning causal relations at the (micro-) physical level are not coupled with causal statements at some suitable macro-level---e.g. in accelerator experiments in particle physics---the causal relations tend to be highly sensitive.

Our point is a gradual one: we do not deny that it can be perfectly legitimate to apply causal vocabulary at the level of fundamental physics, as persuasively argued by Frisch (2014). Not does our argument presuppose a sharp hierarchy of clearly delineated levels of theories. With those caveats in mind, we argue that causal relations at a fundamental, microscopic level may be highly sensitive to background conditions and to that extent may obtain only in a somewhat weak and “boundary” sense. In the example of accelerator experiments in particle physics this sensitivity is to a degree that the experimental setup can be considered essentially a background conditions-controlling machine.

2. Intervention and causation in complex high-level systems: micro-properties that realize macro-properties

The interventionist conception of causation (see Pearl, 2009 and Woodward, 2003) emphasizes the role of external manipulation (“intervention”) on a system and in particular the notion that what is special about a directed causal relationship between variables A and B is that intervention on A can alter B. This view has close links to the experimental sciences, where designed experiments are used to carry out interventions on systems for which the scientist may not have a complete microscopic description or any plausible route to such a description.

Interventionist concepts have played a key role in contemporary developments in causal analysis in the field of machine learning (see for example Pearl, 2009 and Peters *et al.*, 2017). The interventionist framework can be formalized in terms of random variables (RVs) with a key notion being that the joint probability distribution (over the RVs in the system) can be changed under intervention on the system. Note that the RVs in this formalism need not be at the micro-physical level and indeed the probabilistic interventionist formalism is in a way agnostic to level.

A *causal graph* is a directed graph with vertices corresponding to RVs in the system of interest and edges corresponding to causal relations. There are many variants of causal graphs and their semantics and interpretation are the focus of an active scientific literature (see e.g. Hyttinen *et al.*, 2012 and Kocaoglu *et al.*, 2017), but an important point is that causal edges in such graphs refer to causal relations whose micro-physical details may be left implicit. Thus, in a graph with vertices {smoking, lung cancer, blood pressure}, an edge from smoking to lung cancer (not via blood pressure) would mean that the effect of smoking on lung cancer is causal in an interventionist sense and furthermore direct in the sense of not operating only via its effects on blood pressure, but it would be left implicit that the effect of smoking on lung cancer operates via myriad micro-physical events.

Shapiro (2012), partly in joint work with Sober (Shapiro and Sober 2012), as well as Woodward (2015) argue that, in the framework of interventionism about causation, the causal exclusion argument is to be rejected. If M1 and M2 are higher-level variables such that M1 causes M2, one can think of more fundamental micro-physical variables P1 and P2 such that M1 and M2 are coarse-grained versions of them, respectively, where P1 causes P2 just as M1 causes M2. As Shapiro and Sober emphasize, in this case, just as intervening on M1 is a way of (indirectly) intervening on M2, intervening on P1 is a way of (indirectly) intervening on P2. The functional dependence between M1 and M2 on the one hand and P1 and P2 on the other may be very different from each other, but the two causal relations in general coexist.

We agree with Shapiro and Sober on this point but wish to point out that, for realistic cases of variables M1 and M2 (where, say, M1 is the variable "smoking" and M2 is the variable "cancer") the variables P1 and P2, when imagined in the language of, say, quantum field theory, are not literally given and may not be even approximately expressible in practice. Moreover, inasmuch as we can characterize the conditions in which "Smoking causes lung cancer obtains", we do so in higher-level language. The very idea of variables P1 and P2, as well as the idea of a robust causal relation between them are in this sense "parasitic" on our understanding of a robust higher level causal relation.

In the following section we will further explore the distinction between robust ("insensitive") and sensitive causal relations.

3. Sensitive and insensitive causation reviewed

The distinction between sensitive and insensitive causation appeals to the background condition with respect to which causal relations obtain. A causal relation is sensitive if even rather small changes in background conditions disrupt it. Otherwise it is insensitive. The distinction is highlighted as of crucial importance stage in an illuminating article by Woodward (2006). Woodward traces the underlying idea back to Lewis (1986) and introduces it as follows:

> The counterfactual dependence of effects on their causes is such an obvious feature of many examples of causation that it is easy to miss the fact that there is another feature having to do with counterfactual structure that plays an important role in such examples. This feature has to do with the sensitivity of the causal relationship (and, more specifically, the sensitivity of certain of the counterfactuals associated with it) to changes in various other factors. Broadly speaking, a causal claim is sensitive if it holds in the actual circumstances but would not continue to hold in circumstances that depart in various ways from the actual circumstances. A causal claim is insensitive to the extent to which it would continue to hold under various sorts of changes in the actual circumstances. The sensitivity of counterfactuals is understood similarly. (Woodward 2006, p. 2)

We note that sensitivity of a causal effect can be thought of in terms of a causal graph on a set of variables *V*, augmented with additional "background" vertices *V'* for variables not considered to be part of the system under study but which may influence one or more of its variables *V*. In that case, a causal effect between variables in the system (i.e. members of *V*) can depend on the configuration of the (background) variables *V'* and the entire system can be studied within a unified probabilistic formalism (see for example Pearl and Barenboim, 2014). Sensitivity then amounts to the strength of this coupling and could be quantified using mathematical tools.

Woodward illustrates the concept of sensitivity with an example as follows:

> The sensitivity of the counterfactual
>
> (2.1) If the event of the rock thrown by Suzy striking the vase were to occur, the shattering of the vase would occur
>
> has to do with whether (2.1) or its analogues would continue to hold under circumstances that differ in various ways from the actual circumstances. Put slightly differently, what we are interested in is whether (and which) counterfactuals of the form
>
> (2.2) If the rock thrown by Suzy were to strike the bottle in circumstances Bi different from the actual circumstances, the bottle would (still) shatter are true for various Bi.
>
> Some of the circumstances Bi for which claims of form (2.2) are true are so obvious that they will seem trivial. If Suzy's rock strikes the vase in Boston at the moment at which someone sneezes in Chicago, then presumably if that person had not sneezed but the world had remained relevantly similar in other respects, the bottle still would have shattered. Similarly, if we vary the color of Suzy's blouse or the price of tea in China at the time of the impact. (Woodward 2006, p. 5)

Woodward goes on to list other aspects of the background conditions which presumably can be varied without affecting the truth of (2.1), notably, when and where Suzy's striking and the subsequent vase shattering occur.

Here we consider these notions in type-causal contexts.

Example: "Exercise causes elevation of heart rate in humans" holds for a wide variety of background conditions. "Being born in August causes better grades than being born in July" may hold in rather specific conditions (societies where children born in August tend to be the oldest ones in their classes etc. etc.)

The distinction between sensitive and insensitive causation may be gradual and may not admit a clear classification into "specific" as opposed to more "general" background conditions. Furthermore, inasmuch as background dependence of causal relations can be quantified it is often multi-dimensional, i.e. a causal relation can be sensitive along one dimension and insensitive along another. For example, the causal claim "Being born in August causes better grades than being born in July" might obtain robustly for all children within a school system, independent of the specific school district, say, (and, so, be insensitive with respect to the "dimension" school district), but only be true about specific school systems, where, in any given school class, children born in August are on average older than children being born in July .

Two related aspects of sensitive causal relations matter for the present discussion: first, the more sensitive causal relations are, the less useful they tend to be because it is more difficult to exploit them. If some causal relation holds only for very specific background conditions, then, it may be challenging to guarantee that these background conditions indeed obtain or even to verify that they do. In those cases, one may have difficulties to exploit that causal relation in practice because one may not be able to ascertain that the background conditions obtain in which it holds.

Second, we tend to view sensitive causal relations---or at least *very* sensitive ones---as only borderline causal. For example, one may argue that “Being born in August causes better grades than being born in July” does not represent a strong causal statement: it is not really one’s birth month *per se* that influences one’s success; rather, it is the wider setting, including conventions about deadlines for school entry, which is influential. Note that from a strictly formal, interventionist perspective a highly sensitive causal relation remains causal in the sense of describing changes under intervention in the system, albeit under restrictive background conditions. Nevertheless, regarding highly sensitive causal relations as second class (and perhaps not even causal at all) is particularly natural in situations where one has no chance to ascertain whether the background conditions obtains with respect to which a given sensitive causal relation supposedly holds.

4. Is causation in fundamental physics (hyper-) sensitive?

In textbooks of fundamental physics, unlike in those of, say, the social sciences, causal notions are typically not very prominent. As pointed out by Woodward (2006), the typical problem in the social sciences – that one observes certain correlations but finds it difficult to judge whether and, if so, to what degree they are causal – does not prominently arise in fundamental physics.

Still, discoveries in fundamental physics involve the tracing of effects to their causes. In accelerator experiments, for example, inferences are made about the causes of certain detection events. Sometimes those events can be linked to decays of more energetic particles such as the Higgs boson, whose existence can then be indirectly inferred. Wüthrich (2017) gives an account of the 2012 discovery of the Higgs boson (ATLAS Collaboration 2012, CMS Collaboration 2012) as such a causal inference. In the meantime, not only the existence, but also the specific properties, of the Higgs boson have been inferred from combining data on detection events of proton pairs (ATLAS Collaboration 2014), lepton quartets (ATLAS Collaboration 2015), and bottom quarks (CMS Collaboration 2018) among others.

That decays of Higgs particles can – directly or indirectly – result in the creation of such pairs can relatively straightforwardly be derived from the Standard model of elementary particle physics using the formalism of quantum field theory. In actual physical experiments, however, the inference from detecting such pairs and quartets of particles to such a decay having taken place requires great care. Part of the reason for this is that the background conditions potentially affecting the detection events must be carefully controlled. The decay (assuming that it has indeed happened) must take place in a high quality vacuum so that the trajectories of the decay products are accurately predictable, other potential causes of the detection events can be excluded, the detectors must be carefully setup and calibrated, and the numerical analysis that separates detection events from “noise” must be carefully designed. Because of these complexities, particle accelerator and detector design and construction are now scientific (and engineering) fields in their own right, as is the extraction of genuine particle “signals” from noise in high energy physics experiments.

One way of understanding these complexities is to regard them as manifestations of the very high sensitivity of causal relations at fundamental physical levels: the decay of a Higgs boson causes the imminent detection of particles at specific angles with specific velocities, but it does so only against a very specific fixed background that must be rigorously controlled, both concerning the spatio-temporal setup (which is predominantly an engineering challenging) and concerning the numerical analysis (which is predominantly a theoretico-computational challenge).

This point can be illustrated by highlighting that there are also about 10 000 decays of Higgs particles per day in the atmosphere of the Earth, created by cosmic rays. But apart from the fact that the locations of these decays are impossible to predict, the background conditions against which they occur are not “under control” in that it is impossible to discern particles that result from them from irrelevant statistical noise.

Experiments in high energy physics are in a sense “background conditions controlling machines” that enable discoveries about fundamental physical objects. Beyond those experiments, causal reasoning plays at best a limited role in fundamental physics. Generally, types of earlier events and types of later events may be counterfactually linked such that intervening on the former would lead to changes in the latter. But in general, what occurs at the fundamental physical level in a limited spatio-temporal region depends counterfactually on everything that occurs in that region’s backward light cone, and if there is quantum entanglement with objects outside the region, even on space-like separated events. It is no wonder that, outside of highly controlled particle physics experiments, causal notions are usually not much employed in fundamental physics.

References

ATLAS Collaboration, 2012: "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC", *Physics Letters B*, 716:1-29.

ATLAS Collaboration, 2014: “Measurements of Higgs boson production and couplings in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector”, *Physical Review D*, 90:112015.

ATLAS Collaboration, 2015: “Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector”, *Physical Review D*, 91:012006.

CMS Collaboration, 2012: "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". *Physics Letters B*, 716:30-61.

CMS Collaboration, 2018: "Observation of Higgs boson decay to bottom quarks". *Physical Review Letters*, 121:121801.

Frisch, Mathias, 2014: *Causal Reasoning in Physics*. Cambridge: Cambridge University Press.

Hyttinen, A, Eberhardt, F. and Hoyer, P (2012). Learning linear cyclic causal models with latent variables. Journal of Machine Learning Research, 13(1):3387–3439.

Kim, Jaegwon, 1998: *Mind in a Physical World.* Cambridge: MIT Press.

Kim, Jaegwon, 2005: *Physicalism, or Something Near Enough.* Princeton: Princeton University Press.

Kocaoglu, M, Shanmugam, K. and Bareinboim, E (2017). Experimental design for learning causal graphs with latent variables. In Advances in Neural Information Processing Systems 30, pages 7018–7028.

Lewis, D. K. (1986). Postscript C to “Causation”: Insensitive Causation. In: Philosophical Papers, vol 2., pages 184–188. Oxford: Oxford University Press.

Pearl, J. 2009. *Causality: Models, Reasoning and Inference*. Cambridge: Cambridge University Press.

Pearl, J. and Bareinboim, E. (2014). External validity: From do-calculus to transportability across populations. Statistical Science, 29(4), 579-595.

Peters, J., Janzing, D. and Schölkopf, B, 2017. *Elements of causal inference: foundations and learning algorithms*. MIT Press.

Shapiro, Lawrence A., 2012: “Mental manipulations and the problem of causal exclusion”, *Australasian Journal of Philosophy*, 90:507-524.

Shapiro, Lawrence A. and Sober, Elliott, 2012: “Against proportionality”, *Analysis*, 72:89-93.

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

Woodward, James, 2006: “Sensitive and insensitive causation”, *The Philosophical Review*, 115:1-50.

Woodward, James, 2015: “Interventionism and causal exclusion”, *Philosophy and Phenomenological Research*, 91:303-347.

Wüthrich, Adrian, 2017: “The Higgs Discovery as a Diagnostic Causal Inference”, *Synthese,* 194:461-476.