Why adoption of causal modeling methods requires some metaphysics

forthcoming in *Routledge Handbook of Causality and Causal Methods*, eds. Russo and Ilari

H.K. Andersen
Simon Fraser University
handerse@sfu.ca

Abstract: I highlight a metaphysical concern that stands in the way of more widespread adoption of causal modeling techniques such as causal Bayes nets. Researchers in some fields may resist adoption due to concerns that they don't 'really' understand what they are saying about a system when they apply such techniques. Students in these fields are repeated exhorted to be cautious about application of statistical techniques to their data without a clear understanding of the conditions required for those techniques to yield genuine insight into the data. They are acutely aware that anyone can chuck some data into a software package and get what looks like an answer, even though these tests may not be well-defined for the data on which they can apparently be run. This is thus a healthy skepticism for uptake of causal modeling methods, which points directly to the need for a metaphysical understanding of causation in order to successfully use the modeling methods. Without a clear understanding of what the methods are committing to, including what it means to say that there exists a causal relationship, researchers have limited ability to identify potentially bad output, or to independently verify results.

Keywords: Causal Markov, Causal Faithfulness, causal Bayes nets, modelling, metaphysics

1. Introduction

Recent developments in causal Bayes net modeling techniques include the development of introductory texts such as Pearl, Glymour, and Jewell (2016) (see also Hernán and Robins 2010, Kleinberg and Hripcsak 2011). This primer is aimed at a range of practitioners across various sciences whose work involves statistical analysis. It is intended to facilitate adoption of causal modeling techniques by researchers who already use various statistical methods for analysis of their data. This audience is quite broad, including but not limited to biological sciences, medicine, and epidemiology. The development of such a textbook marks a genuine shift in discussions around causation and causal modeling. Kuhn (1962) claims that the establishment of a first paradigm involves or even requires textbooks that teach the new view, taking it as something to be used without arguing for it against competitors. The science of causation is at such a point in its development, having launched with Pearl (2000) and Spirtes, Glymour, and Scheines (2000), and expanded with a semantics in Woodward (2005).

Yet there has been from resistance to adoption of these methods from the very practitioners for whom the methods could provide a valuable additional handle on their complex and often very large datasets. In this paper, I will highlight a lacuna in existing resources for practicing scientists that stands in the way of more widespread adoption of causal modeling, with an example drawn from biomedical physiology. The reason to resist adoption of causal Bayes nets are currently presented stems from concerns that practitioners don't 'really' understand what they

are saying about a system when they apply such techniques and make claims about causal relationships. There is both a methodological and a metaphysical component to this concern, and I will argue here that there is a lacuna in the metaphysical resources available that stands in the way of causation finally launching as an entirely independent science.

Students in these fields are repeated exhorted to be cautious about application of various statistical techniques to their data without a clear understanding of the conditions that have to be met in order for those techniques to yield genuine insight into the data. Anyone can chuck some data into a software package, pick some analyses to run, and get what looks like an answer on the other end. But this 'answer' can be deeply misleading: many statistical tests may not be well-defined for the kinds of data on which they can still be apparently run. Learning what those statistical techniques require for appropriate application, when they might fail, and how to doublecheck their results via an independent method is part of graduate training in these fields and something of which practitioners must remain ongoingly cognizant.

This gives reason for a healthy skepticism about adoption of causal Bayes nets as part of one's repertoire of statistical methods (to be clear, this is not unique to causal Bayes nets; my example is drawn from professional experience working with researchers in biomedical physiology on this particular issue). I show that this is a methodological concern that points directly to a metaphysical issue: what at least some are looking for is a foundation for causation such that they can assess how their particular target systems are structured, to be able to choose appropriate analyses. I illustrate using the example of distinctness of variables, a requirement for application of causal modelling techniques that rely on the Causal Markov Condition, as a problem biomedical researchers will be especially likely to run into. What I call foundations of causation, or, the metaphysical view(s) that provide a basis for causal Bayes nets algorithms and interventionism generally, helps solve this problem, and is a missing resource for researchers.

As a terminological note, I will be referring to some specific kinds of what I will call techniques, such as the application of causal Bayes nets algorithms to datasets in order to generate possible directed acyclic graphs (DAGs) compatible with the data. I will also refer to causal modelling more loosely: by this, I mean the ongoing modeling practices by which researchers develop a full causal model, which includes a far more expansive range of activities. The term 'model' is thus polysemous in this discussion. Sometimes it is used narrowly to refer only to DAGs with assigned weights on edges; sometimes it is used broadly to include any of the practices or products of modeling including graphical causal models and possibly also including other representational tools such as e.g. sets of parameterized equations or systems of differential equations used to model causal structure in the world. In the broader sense, causal Bayes nets techniques are one stage in causal modeling more broadly.

2. Reasonable skepticism about causal Bayes net modeling techniques

In some sciences, experimental and other methods of getting data can occur in ways that make it especially hard to wrangle the datasets. The sheer size of the datasets can be part of the issue (e.g. Leonelli and Tempini 2021, Chin-Yee and Upshur 2019). The number of variables involved can become quite large and difficult to integrate (e.g. Canali and Leonelli 2022). Many statistical

techniques can discern correlations between variables, but this might still leave researchers with a very large number of possible configurations of the underlying system(s) (see, inter alia, Spirtes 2001). Causal Bayes net algorithms would seem like a great solution to this problem: when applied to a set of variables and the data drawn from the system in question for those variables, they offer a set of directed acyclic graphs (DAGs) that are compatible with the observed conditional and unconditional dependencies and independencies.

As a field, biomedical physiology is a great candidate for these methods. Researchers in these units have backgrounds from a variety of sciences, including medicine, anatomy, kinesiology, epidemiology, gerontology, and more. Researchers work on overlapping parts of various systems: one might work on the capillary structure of the kidney, while another works on the entire renal system. Anatomical structures like capillaries in the kidney require methods such as blood flow measurements at a fine-grained spatial and temporal scale, while the role of the kidney in the renal system will involve longer time scales and coarser size scales. Yet these are clearly overlapping parts of understanding how the kidney works, what its structure(s) are, and how those structures allow it to serve the roles that it does (e.g. Zehra, Cupples and Braam 2021, Cupples 2007). Being able to concentrate on key interventions based on a smaller set of possible DAGs would be a boon.

Yet at least some practitioners who have been offered an especially clear chance to learn and implement causal Bayes net modeling have resisted doing so.[1] Why would they not immediately adopt established methods that could benefit their work? I will start from their reasoning, in one department, and then generalize. Their concerns are reasonable, and I want to focus on what those reasons are, not on the relative proportion of practitioners who might espouse such reasons.

Researchers were confident they could learn how to implement the methods in question (especially because, in the test case from which I am drawing these broader generalizations, a faculty member who had been specifically trained in this, by a major figure working on these techniques, was offering the training). Many were skeptical about using these methods because they did not understand exactly what was ‘going on under the hood’, but not just in the software. They were concerned that they did not 'really' know what it meant to say that there was, or was not, a causal relationship between variables. There are three main concerns to highlight here. The first is that they were not sure *when* these methods would apply, or fail to apply. The second is that they were not sure what *commitments* they were making about their systems of study by using these methods or by offering causal conclusions, including very basic "I don't feel like I really know what causation *is*" concerns. The third is that they did not know how to *check* the

[1] These concerns were raised by faculty and graduate students in the Simon Fraser University Department of Biomedical Physiology and Kinesiology (BPK). One faculty member in the Department did graduate work on causal modeling in biomedical physiology with Jewell (e.g. Pearl, Glymour, and Jewell 2016), and has been urging colleagues to incorporate this methodology into their own work, even offering trainings to interested faculty. Colleagues have resisted this, and their reasons given in discussion for resisting adoption of the new methods are the basis of this paper. Also of note is that, in the Graduate Methods seminar, graduate students were immediately interested in and actively using concepts from the new mechanism literature (e.g. Machamer et. al. 2000) when it was introduced, and thought these obviously helpful in collaboration with colleagues for coordinating work at different levels of anatomical organization. In contrast, they were unable to see applicability of causal Bayes nets algorithms to their own research, even when given a very sympathetic introduction to it, and expressed reservations to the effect that it simply did not suit their target systems.

plausibility of the results of the method by some other means. This third concern centers on the fact that they felt they had no clear independent method of verifying the results of a given analysis, because they were only being presented with one view of what causation is, namely, that implicit in the causal Bayes nets (such as interventionist semantics).

The SPSS problem was specifically mentioned by several faculty members. SPSS was an early statistical software package that introduced many to the vagaries of careful usage. The 'running' of the analysis in SPSS-like methods involves selecting a test to run on data, but without having to do any of the calculations oneself, using what seemed like a very slick and sophisticated user interface. It was so easy to use, though, that almost any statistical test could be applied to almost any data, even when those tests were not actually well-defined on the kind of data on which it was run. What looked like real results could turn out to be garbage, and this was opaque from the user-interface experience. The same problem can arise for any methodology that involves algorithms and statistical analyses applied to data by inputting data to a software package and then receiving results, where practitioners lack a sufficiently clear understanding of the analyses being run, in any of these three regards identified above.

A generalized version of this problem can be a liability for those in the process of learning these statistical techniques. It is a major part of graduate training in these fields to learn what the statistical methods used in the field are, when they can be applied, what they can show, where they break down, and how to select among several options to get at various aspects of a phenomenon under investigation. Training to become a practitioner in the field involves learning what features a set of data must have to be amenable to specific analyses, and how those features of the data are situated with respect to the systems from which the data is drawn. The data are not considered in isolation of the system(s) being studied; they are an imperfect window on those systems, and must be interpreted carefully for what they reveal about the target(s) of investigation. The training sets students up to learn how to interpret the results of a statistical analysis as revealing something useful. These cover the first two concerns noted above.

A key part of this training is learning how to independently verify the results of an analysis. If some result seems fishy, maybe outside the expected range by just enough to raise a red flag, running the same analysis again will not yield insight as to whether this is a valid result or if it involves artefact or error. Graduate students are trained about how to approach such cases using some other method of analysis or estimation in order to triangulate in on some range that results should be expected.

Taken together, good training in statistical methods involves knowing enough about how the data was taken from the system, what features the system may have that affect or bias this, what is required of the data for particular analyses to be appropriate or revealing, and how to take the results of those analyses of the data and interpret it as revealing important features of the system in question. All of this depends on a deep understanding of what it is one is attributing to the system in question based on the results of statistical analysis

This further differentiates into two areas of knowledge required to implement a technique like causal modeling, each of which involves a different kind of expertise but both of which are required for applying a particular technique to a particular system. The first is knowledge of how

the statistical methods work, generically. The second is knowledge of how this set of data was drawn from the system, and what knowledge researchers already have about the system's structure. This second kind of knowledge is system-specific. Researchers working on capillary structure in the kidney and its role in the renal system do not start from a tabula rasa in terms of background knowledge about its causal structure. What they have is surely partial, but it still helps in setting up how the measurements are designed to draw data, and are a key part of the 'first pass' at the variables that will end up in the system of causal variables for which a DAG might be eventually found.

3. Research likely to violate conditions like Causal Markov and Faithfulness

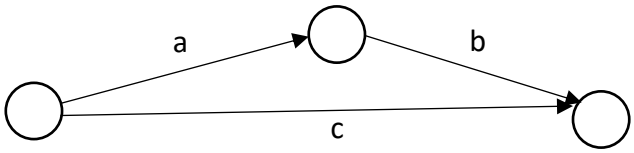
In order for causal Bayes nets methods to return reliable sets of possible DAGs compatible with a given dataset for a set of variables, both Causal Faithfulness (CF) and Causal Markov (CM) conditions must be met. There are further conditions as well, often dependent on or tied to these conditions, such as modularity of causal relationships among variables in a system with respect to intervention, or distinctness of variables so that variables in a single system don't double-count instances. Then, there are further conditions such as acyclicity. All these causal modeling techniques are thus conditional: *if* specific conditions are met, then these techniques will are justified, and return reliable results.

Biomedical physiology, as an example, situated where biological sciences intersect with medicine, is the kind of field that would benefit a great deal from adoption of these modeling techniques. And, as noted above, they have legitimate reasons to be cautious about adoption of this methodology. What makes their work special in this regard, though? There are distinctive features of the systems they study that in fact raise real complications for the use of causal Bayes nets methods. Given how carefully these methods were derived and justified, why should some research areas differ with respect to the requirements for application?

One major reason these techniques are difficult to apply in fields like biomedical physiology (and there are many more that have similar systems of study; this is just an example to illustrate the point) is that these fields are disproportionately likely to be studying systems with exactly the kinds of causal structure that violate the preconditions for application of causal Bayes nets.

This renders concrete the concerns raised in the previous section. In order to apply some method with well-founded confidence, researchers need to be able to *check* that these conditions are met. In order to check that they are met, they need to know what they should be looking for in a system. And, they must be studying systems that are amenable to use of these methods; in other words, not study systems that are likely to have causal structures that violate the conditions, or use variables or systems of variables that violate the conditions. The skepticism in the previous section can be re-stated as concerns that researchers don't know how to check that these conditions are met, and don't know *exactly* what they are saying when they say there is, or is not, a causal relationship, because they don't know exactly what causation *is*.

Consider the Causal Faithfulness condition first. The faithfulness condition is, informally, the assumption that there are no precisely counterbalanced causal relationships in the DAG such that there is a causal pathway from two or more variables to an effect variable, but the effect variable is probabilistically independent of those. An example of what is called a triangle violation of CF:



where a + b = -c

The variable X will be probabilistically independent of Z, because X brings Z about via the pathway involving Y with exactly the same weight that X suppresses Z along the direct pathway. This will 'fool' the algorithms that start from probabilistic relationships among variables in a dataset and infers to the DAGs that are consistent with those probabilistic relationships. The set of candidate DAGs that are returned will not contain the correct DAG. Using the algorithm to narrow down the possible DAGs as a way of directing resources more efficiently on a system that violates CF will then direct researchers into a fruitless set of interventions to distinguish among candidate DAGs, none of which will end up being correct.

The original presentation of CF (Spirtes, Glymour, Scheines 2000) claimed that violations of CF would be vanishingly small, because of the improbability of getting exactly the same weights with opposite valence along two separate pathways to an effect variable. If all paths were arbitrarily weighted with equal probability of falling along the range [0, 1], this would be true.

Fields like biomedical physiology, however, study systems that involve many complex, often hidden, homeostatic systems. Physiology is full of examples where some parameter, such as blood pressure, needs to be kept within a given range, and so a variety of mechanisms are balanced precisely so as to do so. These homeostasis-maintenance relationships will violate CF. Worse, researchers may not realize they are there, such that they can be on the look-out for them. This makes researchers use of causal Bayes nets for systems with potential CF violations difficult. In order to know that they have adequately accounted for CF violations, they would already have to have the kind of causal knowledge that researchers lack, and are trying to gain. And researchers already know that such violations are not only likely in physiological systems, but that they are likely, for especially important homeostatic values such as blood pressure, to have multiple layers of 'back-up' to maintain the value within a constant range (e.g. Andersen 2013). Some of these may not even be active until something 'knocks out' the usual mechanism for homeostasis (e.g. Mitchell 2008; see also Andersen 2012). Renal systems are an excellent example of this: "The kidney represents a unique demand-driven, interconnected resource distribution network that is responsible for body homeostasis maintenance over a broad range of conditions, including variations in blood pressure and fluid intake and loss." (Postnov et al 2022, p. 1/13).

There are work-arounds to this issue, and weaker assumptions that can be used in place of CF (see, inter alia, Choo et al 2023, Marx et al 2021, Weinberger 2018). These weaker assumptions, though, justify weaker inferences from data; using them also mitigates some of the advantages to adopting this method. Any time homeostasis is involved, we should expect to find CF violating

causal structures. This means that in fields like biomedical physiology, researchers are right to be skeptical of the resources being offered for causal modeling.

Next, consider the Causal Markov condition (CM). Informally, this condition ensures that the only persistent probabilistic dependencies between variables is due to some causal pathway connecting them. Put another way, causally unconnected variables will not be probabilistically dependent in the data. This condition is required in order to establish some of the most important parts of the formal framework, such as D separation.

From a philosophical perspective, the Causal Markov conditions looks almost tautologically true (Hausman and Woodward 1999). It would only be violated, in principle, if there were systematic, persistent, yet unexplainable correlations between two or more variables in the system. The chances of finding such a system seem to be vanishingly small, since it would require extraordinary coincidence, as long as one chooses variables appropriately.

This perspective is the wrong angle to help researchers, though. Consider the kinds of issues faced by researchers working on the capillary structure of the kidneys, and its place in the renal system. There are multiple different researchers that focus on different parts of this complex phenomenon. Some of them will be working in fine-scale detail, at very small physical size scales and/or very short time scales, while others work at larger size scales, or, often the same thing but not always, higher levels of organization, compared to the first, such that the target of investigation of the first researcher is a component or part of the target of investigation of the second (e.g. O'Malley et al 2014). Often enough, this means that the data taken by one researcher will involve different variables than another researcher, even to the point of having two variables that measure closely related but not quite identical versions of some quantity.

When two researchers are working on overlapping but non-identical parts of the same phenomenon, it is easy for philosophers to say, CM is almost tautologically true if the variables are chosen correctly. Yet the question faced in the lab is, what is the correct way to choose the variables? How should these overlapping variable sets be combined, and when will it be safe to apply the algorithm to get a set of compatible DAGs? This matters a great deal because of how it supports further requirements such as distinctness of variables. Accidental double-counting of instances by variables in the same system will violate CM.

A short explanation demonstrates why this is so (a longer version of how doublecounting violates CM can be found in Andersen 2023). If two variables sometimes each count the same token event as an instance of that variable taking one of its values, then, if those two variables are combined into a single system of variables, they will always be correlated, even when they have no causal pathway of any kind connecting them. This is because they are double-counting: one event happens, but it counts as an instance of two variables. Those instances will result in some non-zero correlation between the variables. This correlation will persist even when either of the variables are intervened on.

This violates the CM not because of deep reasons related to causation, but because of putting the wrong variables together into a single system. It is a modelling mistake, not a metaphysical issue. Yet it is exactly the kind of 'mistake' that researchers may not realize they are committing,

especially in fields like biomedical physiology, with overlapping size and time scales and integrated processes that don't have clearly defined start and end points. It highlights their concerns about the SPSS problem for causal modeling. It requires substantial resources to check that two data sets can be safely combined without violating CM. Centrally, it requires a clear understanding of what the Causal Markov condition is, what it implies for any given system, in order to adequately ensure that it is satisfied in a particular system of variables being constructed. As such, if researchers want to integrate different work on the same complex phenomenon, they require guidelines by which to tell when the algorithm can be used, and when it should be avoided, including a basic understanding of what causation is, such that a condition like Causal Markov falls out of it as a useable part of the method. They cannot tell if they have accidentally double-counted, or check using other methods whether their variables are adequately independent, until they have a solid foundation in what it means for there is be causal structure at all.

The response that there always exists such a way of individuating variables so as to satisfy CM, and thus, that CM is always satisfied if one is modelling correctly (e.g. Woodward 2016; see also Danks and Harrell, this volume), helps itself to a great deal of presupposition. It is a way of stating that there is an answer to the problem, without solving the problem. Knowing that there is, in principle, some way of selecting variables so that CM is satisfied does not give them anything substantive to answer the question of whether *this* way of individuating the variables does so. They have to solve the problem the hard way, not merely state that there exists a solution. Telling them that, if you do it right, the CM will be easily satisfied, is to avoid the difficulty that was identified earlier, that of being able to check if the conditions are satisfied. What is required to solve this problem is details about how to go about setting up variables so that they are distinct, for example.

Finally, while there isn't space here to fully explore this, there are also issues related to the 'A' in DAG. Some of these systems will involve cycles between variables. Techniques to deal with this involve tricks like converting the data these into time-indexed series (e.g. Runge 2018), or to consider a smaller time unit for modeling than the length of the time cycle (see also Ramsey et al 2010). Many (arguably, the overwhelming majority) of the systems studied in biomedical physiology are cyclical. For this first trick, converting data into time series that are indexed so that there are no cycles, note that requires researchers to only take and use data in very specific ways. It may be that doing the time-indexing in the way required to ensure acyclicity, in order to use causal Bayes nets modeling, will involve more work than just using the techniques they are already using. It is in principle possible, from a philosophical perspective (e.g., Spirtes and Zhang 2016; Leuridan and Gebharter, this volume). But that does not translate into being feasible or desirable from the perspective of actual researchers.

It is possible to compensate for some of these by modeling something less than or just short of a full cycle. This does require a great deal of pre-existing knowledge about the system. Yet even with a substantial amount of pre-existing knowledge about causal structure of cycles in the system of interest, it may simply not be possible to find a single cycle length such that modeling on a timescale shorter than that cycle length would ensure acyclicity in a DAG. This is because there are, for many such phenomena of interest, multiple nested time scales that modelers need to consider. Just for the renal system alone, there are time scales on the order of hours, on the order

of minutes, and many on shorter time scales, including cycles that don't have fixed lengths because they are tied to blood pressure. For example, a cycle involving one full heartbeat cycle will not have a single definitive length. Taking only the shortest of the time scales won't allow for the integration of data from different researchers that they rely on. And, it may require the very knowledge of the system they lack and are trying to gain. Put together, this means that ensuring the CM is met may require having the independent verification of other causal structure that researchers lack. Put together, this is good grounds for scepticism about the usefulness of this method.

4. The solution: foundations of causation

There are several interconnected different issues here. One is that practitioners have reason to resist using statistical methods for which they don't have a clear understanding of when it is appropriate to apply, or how to robustly verify the results of an analysis, or how to interpret the results of such an analysis in terms of its implications about the target of investigation.

A second issue, though, is that they don't have the resources to answer these questions for *causal* modeling in the way that they do for other statistical methods. The pedagogical problem can be avoided by developing resources for checking variable overlap, or identifying markers that might indicate violations of CF. One major issue can be avoided by improving materials for training, by making them more concretely applicable to the specific types of system practitioners in a given field will be applying them to.

As it stands, though, practitioners still find themselves short on the kinds of conceptual and foundational resources that would allow them to address the other, second, issue for causal modelling. Founders of causal Bayes net modeling, notably Spirtes, Glymour, and Scheines but also Pearl, insisted on an aggressively agnostic stance about what causation *is*, precisely so as to leave behind the existing philosophical discussions and avoid what they took to be philosophical baggage in order to proceed in developing the methodological framework. When Woodward (2005) offered an interventionist semantics for the causal Bayes nets modeling, it involved a determined avoidance of anything metaphysical. Even now, as texts are beginning to appear that introduce these techniques to graduate students and interested researchers, the focus is still on the formal justification instead of the practical, here is what to look out for and how to trouble shoot.

The first problem, that researchers lack the resources to know that they know how to use these techniques, cannot be fully solved for until the second problem is also solved. This second problem is what sometimes goes under the heading of causal metaphysics, what I call the foundations of causation (Andersen 2017). What researchers are looking for, to address the methodological problem, is essentially a full view of what causation *is*, in general and in the details, such that they can then use that to get more concrete and specific for their own particular systems, in order to find ways to check if the conditions hold of their system, to guide their selection of variables to ensure that conditions like CM and distinctness are met, to verify if the results are reasonable and identify potentially mis-fires that need to be re-done, and so on.

Recall the point about the conditional character of causal modeling: *if* specific conditions are met, then techniques like causal Bayes nets will return reliable results. These conditions might, as we saw, be fundamentally violated in the systems in question: many evolved systems, especially those involving homeostasis, involve causal structures that are highly calibrated over long evolutionary periods exactly in order to violate CF. There is no amount of finessing the way in which data is taken for those to not violate CF. It is not a failure on the part of the researchers engaged in modelling. It is just a genuine feature of those systems: these processes evolved exactly in order to have robustly counterbalanced causal relationships. Finding such CF-violating relationships may itself highlight a target for further investigation: why were these relationships selected to be balanced, in terms of what further values are controlled by small changes in the weights of those counterbalanced relationships. And it might be that, if one already knew the causal structure of a system, one could construct exactly the right set of variables to represent it while satisfying both CM and CF. But this is not the situation researchers are generally in.

The development of causal Bayes net modeling and other such techniques deliberately avoided the questions such as what causation *is*, in order to develop these methods. SGS claimed that causal faithfulness was more or less a guaranteed feature for purely mathematical reasons regarding the vanishingly improbable matching of two random numbers drawn from the real number line (2000, p. 71). There is nothing in that line of reasoning that recognizes how causation is a physical relation that might have further features than those given in the bare mathematical description. The bare-bones causal Bayes nets was then given a semantic interpretation by Woodward (2005, 2021) as interventionism. Famously, even while adding back some of the stripped-out philosophical content and physical significance, Woodward still avoided foundational or metaphysical questions about causation. It is striking how effectively interventionism about causation worked to launch this approach to causation as something practitioners in a wide range of fields could actually take up and put to use. There is genuine progress that has been made by severing the older philosophical conversation about causation from the newer developments in causation and causal modelling.

Yet the example from researchers in biomedical physiology show that there are limits to how far causal modelling can be developed and launched in the absence of foundations. And those foundations are precisely the kind of philosophical commitments that were eschewed before. In terms of the practical problem faced by researchers considering adoption of these methods, what would be ideally helpful, from their perspective, would be something like checkable conditions that are explicitly defined in an almost recipe-like fashion to allow for researchers to assess how likely their target system is to, for instance, have the kinds of homeostatic equilibrium maintenance mechanisms that violate CF, or, for them to check that they are not combining data taken from different levels of the same phenomena that are liable to violate distinctness, etc. These definitions of extensional independence and distinctness, along with prototypical examples of when they are fulfilled and when they would be violated, along with the same for Causal Markov and Causal Faithfulness, would be a marked improvement in convincing researchers to adopt these methods.

Having a solid metaphysical foundation has methodological consequences that researchers can actually use. Consider the situation for the claim that cause and effect must be distinct. How would we even adjudicate whether causal relata 'really' are distinct or not? We can give a

pragmatist answer to this. If we take distinctness on board as a requirement, then we are able to rely on the Causal Markov condition to be much more likely to hold of our system; if we reject distinctness as a requirement, then we also lose the Causal Markov condition. Given that there can be real inferential power in using these methods, we thereby have practical reasons to endorse a metaphysical view where causal relata should be distinct.

Researchers sceptical of adopting causal Bayes nets methods are not offering a devastating, in-principle critique of the methods. The problem is more to highlight that, in order to actually *apply* these techniques in practice, some additional guidance is required about what the CM means, and what distinctness is, such that they could use this to construct the variables they will use in the first place. Developing the foundations of causation, by laying out the metaphysical and ontological features, not just the epistemic ones, would allow researchers to apply this way of thinking to their systems in order to make informed choices about variables that measure phenomena at different time scales and different size scales or levels of organizations.

Establishing a metaphysical or foundational understanding of causation and conditions like CM, CF, distinctness, modularity, and more, is exactly what is needed to provide the answers to address these kinds of skeptical concerns. Scientists studying systems like this know that some of the causal variables at the smaller size and time scales at least partially overlap with variables at larger size and time scales. A clear understanding of what distinctness and extensional independence are, and how they support conditions like CM, can be used explicitly to navigate integration of variables measured at different scales. A clear understanding of what violations of CF look like, in terms of graphical structure, can be used to assess likely 'hot spots' of violations in target systems under investigation, so that researchers can selectively choose to use these methods in appropriate circumstances. This helps satisfy the concern about knowing when the methods might apply, but also helps researchers interpret the results, in terms of what it is saying about their target system(s). Knowing what causation is, such that one is attributing specific causal structure to a system, is part of responsible scholarship for researchers, and it is still a lacuna in the resources to which they have access.

5. Conclusion

Philosophers of science working on causation have sometimes been reticent to engage in more metaphysical debates about what causation is. It is impressive how much development of causal modeling methods, broadly construed, has been accomplished by bracketing many of these issues. This stage of the development, though, is coming to an end. In order to continue this trajectory in which the study of causation branches out of philosophy and becomes a free-standing field, more is required than a purely methodological focus. No matter how well-justified the techniques are, they have to be put to use by researchers who require more than just the justification of the methods. They have to have the more foundational knowledge that allows them to put these methods to good use without falling prey to the next iteration of the software package problem.

Acknowledgements: Much thanks to William Cupples and to graduate students in the BPK Methods courses, as well as to the audience for the 2019 BPK Research Day Keynote, for great discussions and questions. Thanks to Samantha Kleinberg, Bert Leuridan, Federica Russo, and Phyllis Ilari for helpful comments on a draft version of this chapter. I am grateful for the opportunity to live and work on the unceded territory of the Squamish, Musqeam, T'sleil Watuth, and Kwikwetlem First Nations.

References:

Andersen, H., 2012. Mechanisms: what are they evidence for in evidence-based medicine?. *Journal of evaluation in clinical practice*, *18*(5), pp.992-999.

Andersen, H., 2013. When to expect violations of causal faithfulness and why it matters. *Philosophy of Science*, *80*(5), pp.672-683.

Andersen, H.K., 2017. Patterns, information, and causation. *The Journal of Philosophy*, *114*(11), pp.592-622.

Andersen, H.K., 2023. Running Causation Aground. *The Monist*, *106*(3), pp.255-269.

Chin-Yee, B. and Upshur, R., 2019. Three problems with big data and artificial intelligence in medicine. *Perspectives in Biology and Medicine*, *62*(2), pp.237-256.

Choo, D., Gouleakis, T. and Bhattacharyya, A., 2023, July. Active causal structure learning with advice. In *International Conference on Machine Learning* (pp. 5838-5867). PMLR.

Cupples, W.A., 2007. Interactions contributing to kidney blood flow autoregulation. *Current opinion in nephrology and hypertension*, *16*(1), pp.39-45.

Danks, D., and Harrell, M. (this volume, 4b).

Hernán, M.A. and Robins, J.M., 2010. *Causal inference*. https://grass.upc.edu/en/seminar/presentation-files/causal-inference/chapters-1-i-2/@@download/file/BookHernanRobinsCap1_2.pdf accessed August 7, 2023.

Hausman, D.M. and Woodward, J., 1999. Independence, invariance and the causal Markov condition. *The British journal for the philosophy of science*, *50*(4), pp.521-583.

Kleinberg, S. and Hripcsak, G., 2011. A review of causal inference for biomedical informatics. *Journal of biomedical informatics*, *44*(6), pp.1102-1112.

Kuhn, T. S. 1962/2012. *The structure of scientific revolutions*. University of Chicago press.

Leonelli, S. and Tempini, N., 2021. Where health and environment meet: The use of invariant parameters in big data analysis. *Synthese*, *198*(Suppl 10), pp.2485-2504.

Leuridan, B., and Gebharter, A. (this volume, 7a)

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

Marx, A., Gretton, A. and Mooij, J.M., 2021, December. A weaker faithfulness assumption based on triple interactions. In *Uncertainty in Artificial Intelligence* (pp. 451-460). PMLR.

Mitchell, S.D., 2008. Exporting causal knowledge in evolutionary and developmental biology. *Philosophy of Science*, *75*(5), pp.697-706.

O'Malley, M.A., Brigandt, I., Love, A.C., Crawford, J.W., Gilbert, J.A., Knight, R., Mitchell, S.D. and Rohwer, F., 2014. Multilevel research strategies and biological systems. *Philosophy of Science*, *81*(5), pp.811-828.

Pearl, J., Glymour, M. and Jewell, N.P., 2016. *Causal inference in statistics: A primer*. John Wiley & Sons.

Postnov, D., Marsh, D.J., Cupples, W.A., Holstein-Rathlou, N.H. and Sosnovtseva, O., 2022. Synchronization in renal microcirculation unveiled with high-resolution blood flow imaging. *elife*, *11*, p.e75284.

Ramsey, J.D., Hanson, S.J., Hanson, C., Halchenko, Y.O., Poldrack, R.A. and Glymour, C., 2010. Six problems for causal inference from fMRI. *neuroimage*, *49*(2), pp.1545-1558.

Runge, J., 2018. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(7).

Spirtes, P., 2001, January. An anytime algorithm for causal inference. In *International Workshop on Artificial Intelligence and Statistics* (pp. 278-285). PMLR.

Spirtes, P., Glymour, C.N. and Scheines, R., 2000. *Causation, prediction, and search*. MIT press.

Weinberger, N., 2018. Faithfulness, coordination and causal coincidences. *Erkenntnis*, *83*(2), pp.113-133.

Spirtes, P. and Zhang, K., 2016, December. Causal discovery and inference: concepts and recent methodological advances. In *Applied informatics* (Vol. 3, No. 1, pp. 1-28). SpringerOpen.

Woodward, J., 2005. *Making things happen: A theory of causal explanation*. Oxford university press.

Woodward, J., 2016. The problem of variable choice. *Synthese*, *193*, pp.1047-1072.

Zehra, T., Cupples, W.A. and Braam, B., 2021. Tubuloglomerular feedback synchronization in nephrovascular networks. *Journal of the American Society of Nephrology: JASN*, *32*(6), p.1293.