Wiley StatsRef:
Statistics Reference Online

Perspectives on Causality

By *Gianluca Manzo*[1] *and Lucas Sage*[2]

Keywords: *causal inference, computational methods, experiments, instrumental variables, mechanisms, potential outcomes*

Abstract: Statisticians and social and computer scientists tend to approach causality and causal inference with particular theories of causality in mind and defend tools that are supposed to support causal claims from the point of view of that theory. This entry explains why theoretical and methodological pluralism with respect to causality can benefit causal inference. To this aim, we first discuss various understandings of the concept of causality, and of mechanisms, and emphasize that none of them can be considered as intrinsically superior to the other. We then discuss typical design- and model-based identification strategies of causal effects from within the potential outcome approach, and point to the crucial role of untestable assumptions for defending causal claims within experimental and observational methods. Finally, we explain how computational tools like agent-based modeling can aid causal inference and argue that persuasive causal claims in fact require data and arguments produced by methods that are based on different assumptions and that incorporate different views of causality and mechanisms.

Books and articles on causality and causal inference are typically written from the point of view of a specific theory of causality. They present tools that are supposed to support causal claims from the point of view of that theory (for several examples, *see* Gelman[1]). Here we endorse a different approach. We first cover various understandings of the concept of causality, and of mechanisms, and emphasize that none of them can be considered as intrinsically superior to the other (Section 1). We then discuss typical design- and model-based identification strategies of causal effects from within the potential outcome approach, and point to the crucial role of untestable assumptions for defending causal claims (Section 2). Finally, we explain how computational tools like agent-based modeling (ABM) can aid causal inference (Section 3) and, in conclusion, we argue that persuasive causal claims in fact require data and arguments coming from different methods (for a systematic exposition of this perspective, *see* Manzo[2]).

[1]Sorbonne University, Paris, France
[2]ISTC-CNR, Rome, Italy

DOI: 10.1002/9781118445112.stat08510

1 Varieties of Views on Causality and Mechanisms

One of the major insights of philosophical scholarship on causality is that this concept can be given a variety of definitions (Cartwright[3]: p. 806). To put some order among them, an important distinction is between *dependence* (or *difference-making*) accounts of causality, and *production* accounts of causality[4]. The common intuition behind the former is that an event c is the cause of another event e, if had c not occurred, e would not have occurred. In contrast, the production view sees c as the cause of e, if c helps producing, bringing about, or generating e.

Social scientists proposed similar typologies. For instance, Goldthorpe[5] suggested to distinguish between causation as "robust dependence," as "consequential manipulation," and as "generative process." In the first case, X is seen as the cause of Y, if there is an association between the two, and that this association does not (completely) disappear after introducing another set of variables Z possibly related to Y (and/or to both X and Y). In the second case, a cause X is seen as a property that can be at least *in principle* manipulated, such that when appropriate controls are taken into consideration, interventions on X change Y. Finally, when causation is understood as a "generative process," X is seen as a trigger for a well-defined sequence of events that operates at a smaller scale than the association under scrutiny and that has the capacity to generate the effect of X on Y. Thus, Goldthorpe's first two concepts of causation clearly fall within Hall's category of "dependence" accounts of causation, while the third one, that is, "causation as a generative process," illustrates the "production" view.

Among scholars understanding causality in terms of dependence (rather than production) relationships, an additional important distinction is that between "forward" and "reverse causal inference" (*see* Gelman[1]: p. 955). In the former case, one seeks to quantify "what may happen if we do X"; in the latter case, one wants to answer the question of "what causes Y." That is why "what-if" versus "why" causal questions are also often referred as to respectively the "*effect-of-a-cause*" and the "*causes-of-an-effect*" approaches (*see*, for instance, Dawid *et al*.[6]). In Goldthorpe's typology, causation as consequential manipulation amounts to searching for the *effects of causes*, while causation as robust dependence searches for the *causes of effects*.

The distinction is important because a major evolution in contemporary causal reasoning across various disciplines is the diffusion of the potential outcome approach to causal inference, which precisely amounts to shift the focus of quantitative scholars from backward to forward causation (for a historical overview, *see* Imbens and Rubin[7]: ch. 2). Within the potential outcome framework, X is the cause of Y, if after being exposed to X, a unit of analysis manifests a change in Y. In theory, one would need to observe *the same* unit of analysis as being both exposed and *not* exposed to X and then interpret the difference between the two outcomes as the effect of the cause X. In practice, one can only observe a unit of analysis in one of the two states – a practical limitation known as the "*fundamental problem of causal inference*" (*see* Holland[8]: p. 947). To get around this, the potential outcomes approach resorts to the random assignment of units of analysis into "treatment" and "control" groups, where the former corresponds to units exposed to the purported cause X, and the later to units that are not. Thus, randomization makes it possible to interpret the control group as the counterfactual: what would have happened had the units not have been exposed to the treatment. The effect of a cause is then conceived as the average difference between the outcomes among those who were exposed to the treatment and those who were not.

From a methodological point of view, randomized experiments are seen as the prototypical method to implement the potential outcomes approach (Gelman[1]: p. 956) but a major implication of this perspective clearly was the attempt to reinterpret multivariate statistical methods for observational data as a tool aiming not so much to identify "robust dependences" – to go back to Goldthorpe's distinctions – as to render comparable the outcomes of group subjects that were not randomly assigned to the treatment state of interest (Hernán and Robins[9]: ch. 15). That is why the potential outcomes

Wiley StatsRef: Statistics Reference Online, © 2014 – 2024 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08510

framework is now often seen as a "unified framework for the prosecution of causal question" (Morgan and Winship[10]: p. 3).

Despite its ambition of generality, the potential outcomes approach is intrinsically rooted into a specific understanding of causality, namely a counterfactual perspective, which is a form of dependence accounts of causality in above-mentioned Hall's terminology. From a production perspective, this restriction of causal reasoning to counterfactual dependences is questionable. Indeed, many social scientists interested in mechanism-based explanations tend to tie causal inference to the construction of generative models clarifying how the dependence connection of interest could arise (*see*, among others, Boudon[11]; Hedstrom[12]). Even some statisticians prefer to "restrict the term [causality] to situations where some explanation in terms of a not totally hypothetical underlying process or mechanism is available" (Cox[13]: p. 297). To this, followers of the potential outcomes framework retort that mechanism-based explanations in fact can be rigorously tested within a counterfactual framework (Morgan and Winship[10]: ch. 10).

The problem here is that different conceptions of mechanisms enter the picture, which complicates the dialogue between scholars animated by different understanding of causality. In particular, from within a dependence perspective, a mechanism is seen as "(…) a causal relationship involving one or more intervening variables between a treatment and an outcome" (Knight and Winship[14]: p. 282). In this sense, a "mediating mechanism M unpacks the black-box of a treatment to outcome relationship by elaborating on how the causal effect is brought about (via M)"[15]. In contrast, from a production perspective on causation, a mechanism "(…) describes the relevant entities, properties, and activities that link them together, showing how the actions at one stage affect and effect those at successive stages" (Machamer *et al*.[16]: p. 12). In this sense, a mechanism cannot be reduced to a network of intervening variables but should be explicitly modeled as a dynamic system of interacting units whose behaviors generate a sequence of micro-level changes that are supposed to create the dependence connections of interest[17, 18].

Once this variety of views on causality (and on mechanisms) is taken into account, one should become "(…) more cautious about investing in the quest for universal methods for causal inference" (Cartwright[3]: p. 806). Being pluralistic may help to see intuitions behind different concepts of causality as complementary, and to start thinking about causal inference from an "evidential pluralism" perspective[19] according to which diverse methods can produce different types of data and arguments that are equally important to support persuasive causal claims.

2 Causal Inference, Empirical Data, and Assumptions

An obstacle that must be overcome to defend this view comes from the hierarchies that many establish between experimental and observational methods that can help to discover causal connections compared to computational methods like ABM that *cannot*. In a nutshell, the presumed superiority of the former methods for causal inference comes from the fact that they rely on empirical information whereas computational models like ABM are based entirely on substantive and formal assumptions (for a particularly clear statement, *see* Diez Roux[20]: p. 101).

Randomized control trials (RCTs) and instrumental variables (IVs), respectively, one design- and one model-based strategy to identify causal effects from within the potential outcomes approach (*see* Morgan and Winship[10]: pp. 30–33), can be used to illustrate the following point: contrary to this widespread view, experimental and observational methods are also based on untestable assumptions, which implies that the causal claim at hand is, within these methods too, contingent on a complex mix of limited empirical information and theoretical arguments.

DOI: 10.1002/9781118445112.stat08510

2.1 Randomized Control Trials (RCTs)

From a dependence perspective on causality, to establish that X causes Y, one needs to rule out all possible confounders, that is, all possible factors affecting both X and Y. Yet, it is impossible to know all these factors. RCTs are regarded as "the failsafe way to generate causal evidence" in many disciplines (Antonakis *et al*.[21]: p. 1086) because, by randomly allocating units of analysis to different groups – the treatment (where the putative cause X is present) and the control (where the putative cause X is absent) – they ensure *by design* that the effect of all confounders is ruled out. As nothing else than randomness is responsible for units being in the treatment or in the control group, any observed difference in the probability of Y between the two groups could have been caused only by the treatment. The average difference between the treatment and the control group can therefore be inferred to have been caused by X.

As noted by Deaton and Cartwright[22, p. 2], RCT's appealing comes from the fact that randomization to the treatment seems to make the method requiring "(…) minimal substantive assumptions, little or no prior information, and to be largely independent of 'expert' knowledge (…)." But this fails to be the case in practice. In particular, as an RCT cannot say anything about the effect of the treatment on any particular subjects, assumptions are needed to handle the way the effect of the putative cause X on Y varies across subgroups of the target population (Cartwright[23]: pp. 16–17) as well as the possible temporal heterogeneity of these effects within subjects (Sampson *et al*.[24]: pp. 13, 18–19). Turn-arounds for tackling treatment response heterogeneity obviously exist (*see* Manski[25]: pp. 63–76), but these solutions require additional assumptions (*see* Morgan and Winship[10]: pp. 425–427), which, as admitted by the method's proponents themselves, are "(…) credible to the degree that someone thinks it so" (Manski[26]: p. 48).

RCTs also require assumptions that are simply empirically untestable. The most important of them is the "stable unit treatment value assumption" (SUTVA), which requires the absence of possible interference among the units' potential outcomes as well as the absence of hidden variations in the treatment (Imbens and Rubin[7]: p. 10). As noted by Sobel[27, p. 1399], "interference is the norm" in settings where behaviors are embedded in social interactions, which makes SUTVA highly implausible, and likely to be violated, in most of the contexts studied by social scientists. When violations of SUTVA are foreseeable, specific experimental designs can be conceived to prevent these violations (on cluster-based designs, *see* Hong and Raudenbush[28]; on two-stage randomization, *see* Halloran and Hudgens[29]). But, again, these designs, too, rely on further assumptions that are difficult to test empirically and, as noted by Imbens and Rubin[7, p. 11], some "more distant" versions of SUTVA are in fact formulated.

Thus, the superiority of RCTs for establishing causal claims seems unwarranted. RCTs are only apparently sufficient to generate causal knowledge through data *alone*. Any particular experimental design in fact relies on assumptions that have to be supported by external and substantive knowledge.

2.2 Instrumental Variables (IVs)

The same holds for methods for observational data that try to recreate experimental conditions in non-experimental settings. For this aim, IVs are one of the most common techniques[30]. Within this framework, to estimate the effect of a putative cause X on the outcome of interest Y, one can exploit variation on a third variable (i.e., the instrument I) that needs to be correlated with X – the so-called "*relevance*" condition (Stock and Watson[31]: p. 333) – *and* uncorrelated with the putative effect Y given X – the so-called "*exclusion*" or "*exogeneity*" restriction (Gangl[32]: p. 381). By seeing I as a sort of exogenous shock affecting X, so that it affects Y only through its effect on X, IVs promise "(…) to estimate the coefficient of interest consistently and free from asymptotic bias from omitted variables, *without actually having data* on the omitted variables or *even knowing what they are*" (Angrist and Krueger[33]: p. 73, emphasis added).

Wiley StatsRef: Statistics Reference Online, © 2014–2024 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08510

However, similarly to RCTs, this again fails to be the case in practice. It is indeed difficult or impossible to empirically confirm the validity of the relevance and the exogenous conditions.

As to the former, Bound *et al*.[34, p. 446] demonstrated that, when the instrument I is "weak," that is, only weakly correlated with the putative cause(s) X of interest, "even enormous samples do not eliminate the possibility of quantitatively important finite-sample biases" and that small violations to the exclusion restriction are amplified, thus leading to even larger biases. But how can one establish whether an instrument I is strong? The variety of existing technical solutions ultimately suggests that the problem cannot be solved with data *alone*, which led some to admit that "(…) good instruments often come from detailed knowledge of the economic mechanism and institutions determining the regressor of interest" (Angrist and Krueger[33]: p. 73).

As to the "exclusion" restriction, the problem seems even more difficult. Essentially, this condition requires that all potential pathways going from I (i.e., the instrument) to Y (i.e., the potential outcome) are controlled for. But this is a condition that, by construction, cannot be verified because it is always possible that we miss some confounders. That is why Gangl[32, p. 381] (emphasis added) admitted that "(…) it is important to realize that the exclusion restriction is an assumption that *is not testable in principle*." Again, to justify this assumption, theoretical reasoning and substantive knowledge "(…) about how and why things work" (Deaton[35]: p. 432) is required, which makes any causal inference made through IVs contingent on that knowledge (*see* also Rosenzweigh and Wolpin[36]).

Thus, the distinction between experimental and observational methods that establish robust counterfactual dependencies by relying on data, and computational methods like ABM that are incapable to contribute to causal inference, because they are based entirely on assumptions (*see* Diez Roux[20]: p. 101), seems inaccurate. Typical identification strategies of causal effects are clearly not a simple matter of data *alone*.

3 Computational Methods and Causal Inference

Given that experimental and observational methods are always likely to leave doubts about the possibility that a putative dependence relationship is confounded (and/or mediated) by unknown variables, it seems legitimate to ask whether these methods can be complemented by other approaches, in particular those that can algorithmically generate a given dependence relationship of interest, on the basis of an explicit substantive model explaining how X leads to Y.

ABM is a flexible computational method to accomplish this task (for an introduction, *see* Wilenski and Rand[37]). An ABM is a computer program written in an object-oriented language. Objects are "computational entities that encapsulate some state, are able to perform actions, or methods, on this state, and communicate by message passing" (Wooldridge[38]: p. 28). When the program is executed, the objects dynamically evolve according to the rule they encapsulate. The important point is that the objects do not have any specific substantive content, they can represent atoms, cells, individual actors, or any other entities at higher level of abstraction (like firms or states). Objects can be as heterogeneous as desired, and they can have all sorts of attributes. All sorts of structures of interactions between objects can be represented. The method is intrinsically dynamic which allows one to trigger fine-grained sequences of time-stamped events. Finally, an ABM can contain units at different levels of analysis and allows to observe how dynamic micro-level changes progressively lead to macro-level changes. For these reasons, ABM is intrinsically rooted within a view of mechanisms as multilevel dynamic systems of interacting units[39], and thus well suited to implement a production view of causation[40].

ABMs are typically considered as disconnected from causal inference because, unlike experimental and observational methods, they are seen as based entirely on assumptions. Thus they could only speak of the

model's artificial world but they could not say anything about the real world to which the model is assumed to refer (*see* Diez Roux[20]: p. 101). However, this view only covers one specific type of ABM. In fact, there is nothing intrinsic to ABM that limits this algorithmic technique to pure abstractions.

From the early applications of ABMs in the social sciences in the 1960s, two tendencies have co-existed. On the one hand, one found highly abstract models like Schelling's[41] model of residential segregation whose main goal was to explore the macro-consequences of a supposedly general preference for not being in a minority, no matter what specific groups were at stake. On the other hand, one found data-driven models like Hägerstrand's[42] study of the diffusion of agricultural innovations in Sweden where the postulated microscopic mechanism was grounded in empirical observations and simulated outcomes were (at least qualitatively) compared with data on the actual diffusion patterns. In Schelling's case, the ABM followed the logic of the "Keep It Simple, Stupid" (KISS) principle according to which assumptions should be as simple and abstract as possible because the ABM is intended to help understanding general processes present in a variety of specific applications (*see* Axelrod[43]: p. 5). In Hägerstrand's case, the ABM followed instead the "Keep It Descriptive, Stupid" (KIDS) principle according to which one should start with a model that is as descriptively rich as the available data allows. Then, simplifications can be introduced as long as they are not inconsistent with what is known about the empirical functioning of the phenomenon under scrutiny, and as long as the simplification does not reduce the model's performance in reproducing the to-be-explained outcome (*see* Edmonds and Moss[44]). When the KISS principle is prioritized, ABMs are envisaged as "tools to thinks with" (O'Sullivan and Perry[45]: pp. 14–15) whereas when one follows the KIDS principle, the goal is to design "high fidelity models"[46].

Recent reviews of the ABM field suggest that highly abstract ABMs inspired by the KISS principle are still very frequent but that ABMs seeking stronger connections with empirical data are becoming more common (*see* Bianchi and Squazzoni[47]: pp. 299–300). While the value of highly simplified ABMs is indisputable for theory development, when ABM wants to help causal inference, ABMs with "high-dimension realism"[48] should be given priority.

In fact, there are at least three dimensions along which an ABM can gain realism. First, the micro-level assumptions of an ABM can be anchored to sociological and/or psychological theories, possibly in turn supported by empirical and/or experimental evidence, rather than relying on mere intuitions on how agents behave and influence each other. This dimension can be labeled "theoretical realism." Second, the socio-demographic features of an ABM as well as the parameters and functions adopted to implement its micro-level assumptions and interaction structures can be estimated through empirical and/or experimental data exogenous to the model – a form of "empirical calibration" sometimes referred to as "*input validation*" (Delli Gatti *et al*.[49]: pp. 169–172) – rather than on arbitrarily chosen statistical distributions, functional forms or abstract models of network topologies. This dimension can be labeled "input realism." Finally, the simulated outputs of an ABM can be systematically confronted with well-specified datasets describing the target of interest – a form of empirical validation sometimes called "*output validation*" (*see* Delli Gatti *et al*.[49]: p. 165) – rather than general qualitative patterns abstractly defined. This dimension can be labeled "output realism."

Examples of ABMs characterized by the explicit attempt to combine simultaneously theoretical, input, and output realism are still rare but they do exist. For instance, Manzo *et al*.[50] wanted to explain why technological innovations spread at difference paces among Indian and Kenyan potters with different religious backgrounds. To understand whether the structure of family ties within these religious sub-groups may have impacted on adoption probabilities, Manzo and colleagues designed an ABM whose micro-level assumptions were grounded within the theory of "complex contagions" (theoretical realism); then, crucial parameters of the ABM were empirically calibrated, in particular the kinship networks through which the diffusion process was supposed to flow (input realism); finally, simulated diffusion curves were generated as a function of different hypotheses on imitative and learning behaviors, and systematic confrontation

Wiley StatsRef: Statistics Reference Online, © 2014–2024 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08510

with the actual diffusion curves was performed (output realism). By manipulating the imitation mechanisms driving agents' choices, given the empirically calibrated network structure, the authors showed that the effect of kinship networks on the rate of diffusion was mediated by the behaviors (documented through field observations) of central potters within the network so that the same network property could in fact lead to fast or slow diffusion in different sub-groups depending on the specific features of those behaviors.

Obviously, no ABM will reach full realism on the three dimensions. The degree of theoretical, input, and output realism of an ABM should be seen as a continuum, and different observers may provide different assessments of where an ABM stands on that continuum (on the inescapably subjective character of judgment on a model's realism, *see* Sudgen[51]). The point rather is that, contrary to the common view seeing ABMs as pure abstractions, the method is structurally able to embed various combinations of substantive knowledge *and* data and that it is precisely this combination that makes an ABM more or less "credible" to perform causal inference. As noted by Sugden[52, p. 23], "if we are to make inductive inferences from the world of a model to the real world, we must recognize some significant similarity between those two worlds." Theoretical, input, and output realism of an ABM are three ways to increase the degree of "parallelism" between the mechanism(s) the ABM wants to describe and the real-world mechanism. Thus, the higher an ABM scores on theoretical realism, and the best it is calibrated on the input side and validated with respect to its outputs, the higher the likelihood the ABM can serve as a "mimicking" device, and, on this ground, work as an inferential device. As stated by Morgan[53, p. 337] with reference to earlier micro-simulations in macro-economics, "(…) it is this mimicking at two levels that enabled Orcutt's simulation to offer both accounts of the world in the model, and a credible basis for inferences to the real world that the model represents."

In particular, when theoretical realism is high and exogenous empirical data are introduced within the ABM, so that the model is constrained on the input side, the ABM becomes an empirically-constrained device with *its own* behavior. The novel knowledge the ABM is able to produce *on its own* concerns the connection between the low-level empirically-grounded mechanisms and the larger-scale patterns associated with these mechanisms. This knowledge is novel because it was absent from the data that were used to calibrate the model. It is in this sense that when the ABM is empirically calibrated (theoretically realistic and empirically validated), it generates knowledge that it is relevant for causal inference from a *production* perspective on causality.

Actually, this is precisely the way in which ABMs are exploited by the rare studies that confront statistical methods for causal inference with ABM on the same causal issue. For instance, Zachrison *et al*.[54, pp. 3–4] built an ABM to assess to what extent homophily may have confounded dyad-based regression's results produced by Christakis and Fowler[55] as to the causal effect of social ties on various health-related outcomes (for an overview of this debate, *see* Christakis and Fowler[56]). To this aim, Zachrison *et al*.[54] designed an empirically calibrated ABM to generate longitudinal network data of exactly the same kind as those studied by Christakis and Fowler[55] but where the role of network influence and homophily were perfectly known because the two mechanisms were explicitly coded within the ABM. By estimating the regression models specified by Christakis and Fowler[55] on the network data generated by their ABM, Zachrison *et al*.[54] found that while Christakis and Fowler's[55] statistical model was able to detect correctly the presence/absence of the network effect, regardless of the presence/absence of the homophily mechanism, the statistical model was not able to detect correctly the presence/absence of homophily, no matter whether the network influence mechanism was present or absent from the ABM. This is an illustration of a potential major contribution of ABM to causal inference: by explicitly modeling mechanisms that may confound the dependence relationship of interest, an ABM increases our confidence in the robustness of the putative causal effect.

To this argument, one may rightly retort that the conditions under which an ABM can be exploited as an inferential device on mechanistic ground are very demanding because *in practice* full empirical

DOI: 10.1002/9781118445112.stat08510

calibration is never possible (*see*, for instance, Grüne-Yanoff[57]). While this obstacle is real, ABM has internal tools to assess the consequences on a model's results of its pieces for which empirical data and/or well-developed theories may be insufficient to support their realism. Various sensitivity (*see* Delli Gatti *et al*.[49]: pp. 151–162) and robustness (*see* Railsback and Grimm[58]: pp. 300, 312–313) techniques are increasingly adopted by ABM modelers to assess the extent to which an ABM's simulated outcomes are contingent on unverified or unverifiable assumptions. Similarly, several heuristics exist to understand the internal dynamic of an ABM so that one can reduce the probability of making errors of interpretation as to what generates the model's outcomes[59]. And, after all, ABM does not seem special in this respect. As now popularized by the debates on the so-called "*researcher-degree-of-freedom*" problem, experimental designs and statistical methods for observational data involve a number of hidden choices and assumptions that have to be checked[60]. To do so, one solution is to resort to various forms of sensitivity analysis to assess the extent to which the empirical estimates produced by those methods are robust against different model specifications[61] and/or potential confounders (*see* Gangl[32], pp. 385–390). These robustness checks within experimental and observational methods are as challenging as within an ABM. In Gangl's[32, p. 399] words, these tools "would be degraded to little more than a computational exercise in the absence of background empirical and theoretical knowledge suggesting the likelihood and extent of confounding."

4 Conclusion

Statisticians and social and computer scientists have differently conservative attitudes as to the conditions under which causal inference is possible, but the vast majority of them tend to associate causal inference with specific methods (*see* Gelman[1]: p. 959). Our main message is that causal inference should instead be understood at the intersection of different views on causation and mechanisms as well as by combining different methods incorporating these different views. Experiments and statistical methods for observational data and computational models like ABM are exposed to similar problems of insufficient data and untestable assumptions for which only theoretical arguments may be at best available. Given these common limitations, these methods produce qualitatively diverse knowledge that can be fruitfully integrated to reach persuasive causal claims.

Philosophers of science convincingly argued that the uncertainty about the causal nature of a given relation between two events can in fact be reduced in (at least) two ways[62]. On the one hand, data and arguments on credible entities, activities, and interactions – that is, a mechanism seen as a multilevel dynamic system – help to reduce the doubt that the relation of interest could disappear, or be weakened, because of unobserved confounders, effect modifiers, and/or intervening variables. A theoretically realistic and well-calibrated ABM provides this type of mechanistic knowledge. On the other hand, data and arguments on how given entities, activities, and interactions combine to produce some behavior actually does not guarantee that their organized operation is not masked in the broader context of further (possibly unknown) modes of interaction and outer influences so that the overall outcome of the postulated mechanisms *is not* the expected putative causal relation of interest. That is why knowledge produced by experimental and observational methods on robust networks of variables also influences one's belief that the relation may be causal.

According to the "evidential pluralism" perspective on causality[63], this is the main reason why data and arguments along dependence and making-difference lines should be recursively combined with data and arguments on mechanisms understood as multilevel dynamic system of interacting entities and activities. We strongly advocate this perspective because it pushes to combine experiments and statistical methods for observational data and computational methods like ABM rather than establishing dubious hierarchies

Wiley StatsRef: Statistics Reference Online, © 2014–2024 John Wiley & Sons, Ltd.

DOI: 10.1002/9781118445112.stat08510

between them. Experimentalists, statistically oriented scholars, and computational modelers all struggle with similar issues of precision, accuracy, and calibration. In each methodology, knowledge that is equally but diversely relevant for defending a given causal claim is produced through a complex combination of background substantive knowledge, external empirical evidence, and sensitive and robustness procedures that are necessary to justify assumptions that are empirically unverifiable. Thus method synergies seem more fruitful than method "warlordism."

References

[1] Gelman, A. (2011) Causality and statistical learning. *Am. J. Sociol.*, **117** (3), 955–966.
[2] Manzo, G. (2022) *Agent-based Models and Causal Inference*, John Wiley & Sons, Ltd, Chichester, UK.
[3] Cartwright, N. (2004) Causation: one word, many things. *Philos. Sci.*, **71**, 805–819.
[4] Hall, N. (2004) Two concepts of causation, in *Causation and Counterfactuals* (eds J. Collins, N. Hall, and L.A. Paul), MIT Press, Cambridge, MA, pp. 225–276.
[5] Goldthorpe, J.H. (2001) Causation, statistics and sociology. *Eur. Sociol. Rev.*, **17** (1), 1–20.
[6] Dawid, A.P., Faigman, D.L., and Fienberg, S.E. (2014) Fitting science into legal contexts: assessing effects of causes or causes of effects? *Sociol. Methods Res.*, **43** (3), 359–390.
[7] Imbens, G.W. and Rubin, D.B. (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press, Cambridge.
[8] Holland, P.W. (1986) Statistics and causal inference. *J. Am. Stat. Assoc.*, **81** (396), 945–960.
[9] Hernán, M.A. and Robins, J.M. (2020) *Causal Inference: What If*, Chapman & Hall/CRC, Boca Raton, FL *https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/*.
[10] Morgan, S.L. and Winship, C. (2014) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd edn, Cambridge University Press, Cambridge.
[11] Boudon, R. (1979) Generating models as a research strategy, in *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld* (ed. P.H. Rossi), The Free Press, New York, pp. 51–64.
[12] Hedstrom, P. (2009) Studying mechanisms to strengthen causal inferences in quantitative research, in *The Oxford Handbook of Political Methodology* (eds J.M. Box-Steffensmeier, H.E. Brady, and D. Collier), Oxford University Press, Oxford, pp. 319–335.
[13] Cox, D.R. (1992) Causality: some statistical aspects. *J. R. Stat. Soc. A*, **155** (2), 291–301.
[14] Knight, C.R. and Winship, C. (2013) The causal implications of mechanistic thinking: identification using directed acyclic graphs (DAGs), in *Handbook of Causal Analysis for Social Research* (ed. S.L. Morgan), Springer, Dordrecht, pp. 275–299.
[15] Makovi, K. and Winship, C. (2021) Advances in mediation analysis, in *Research Handbook on Analytical Sociology* (ed. G. Manzo), Edward Elgar, Cheltenham, UK ch. 20.
[16] Machamer, P., Darden, L., and Craver, C. (2000) Thinking about mechanisms. *Philos. Sci.*, **67**, 1–25.
[17] Andersen, O. (2014a) A field guide to mechanisms: part I. *Philos Compass*, **9** (4), 274–283. doi: 10.1111/phc3.12119.
[18] Andersen, O. (2014b) A field guide to mechanisms: part II. *Philos Compass*, **9** (4), 284–293. doi: 10.1111/phc3.12118.
[19] Russo, F. and Williamson, J. (2007) Interpreting causality in the health sciences. *Int. Stud. Philos. Sci.*, **21** (2), 157–170.
[20] Diez Roux, A.V. (2015) The virtual epidemiologist–promise and peril. *Am. J. Epidemiol.*, **181** (2), 100–102.
[21] Antonakis, J., Bendahan, S., Jacquart, P., and Lalive, R. (2010) On making causal claims: a review and recommendations. *Leadersh. Q.*, **21**, 1086–1120.
[22] Deaton, A. and Cartwright, N. (2018) Understanding and misunderstanding randomized controlled trials. *Soc. Sci. Med.*, **210**, 2–21.
[23] Cartwright, N. (2007) Are RCTs the gold standard? Contingency and dissent in science, CPNSS, LSE. Technical report 01/07.
[24] Sampson, R.J., Winship, C., and Knight, C. (2013) Translating causal claims. Principles and strategies for policy-relevant criminology. *Criminol. Public Policy*, **12** (4), 587–616.
[25] Manski, C.F. (2013) Identification of treatment response with social interactions. *Econ. J.*, **16** (1), S1–S23.
[26] Manski, C.F. (2003) *Partial Identification of Probability Distributions*, Springer, New York.
[27] Sobel, M.A. (2006) What do randomized studies of housing mobility demonstrate?: Causal inference in the face of interference. *J. Am. Stat. Assoc.*, **101** (476), 1398–1407.
[28] Hong, G. and Raudenbush, S.W. (2013) Heterogeneous agents, social interactions, and causal inference, in *Handbook of Causal Analysis for Social Research* (ed. S.L. Morgan), Springer, Dordrecht, pp. 331–352.

DOI: 10.1002/9781118445112.stat08510

[29] Halloran, M.E. and Hudgens, M.G. (2016) Dependent happenings: a recent methodological review. *Curr. Epidemiol. Rep.*, **3** (4), 297–305. doi: 10.1007/s40471-016-0086-4.
[30] Bollen, K.A. (2012) Instrumental variables in sociology and the social sciences. *Annu. Rev. Sociol.*, **38**, 37–72.
[31] Stock, J.H. and Watson, M.W. (2010) *Introduction to Econometrics*, 3rd edn, Addison-Wesley, Boston, MA.
[32] Gangl, M. (2013) Partial identification and sensitivity analysis, in *Handbook of Causal Analysis for Social Research* (ed. S.L. Morgan), Springer, Dordrecht, pp. 377–402.
[33] Angrist, J.D. and Krueger, A.B. (2001) Instrumental variables and the search for identification: from supply and demand to natural experiments. *J. Econ. Perspect.*, **15** (4), 69–85.
[34] Bound, J., Jaeger, D.A., and Baker, R.M. (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *J. Am. Stat. Assoc.*, **90** (430), 443–450.
[35] Deaton, A. (2010) Instruments, randomization, and learning about development. *J. Econ. Lit.*, **48**, 424–455.
[36] Rosenzweigh, M.R. and Wolpin, K.I. (2000) Natural "natural experiments" in economics. *J. Econ. Lit.*, **38**, 827–874.
[37] Wilenski, U. and Rand, W. (2015) *An Introduction to Agent-based Modelling. Modelling Natural, Social, and Engineered Complex Systems with NetLogo*, MIT Press, Cambridge, MA.
[38] Wooldridge, M. (2009) *An Introduction to Multi Agent Systems*, John Wiley & Sons, Ltd, Chichester.
[39] Vu, T.M., Probst, C., Nielsen, A. *et al.* (2020) A software architecture for mechanism-based social systems modelling in agent-based simulation model. *J. Artif. Soc. Soc. Simul.*, **23** (3), 1 *http://jasss.soc.surrey.ac.uk/23/3/1.html*.
[40] Anzola, D. (2020) Causation in agent-based computational social science, in *Advances in Social Simulation: Looking in the Mirror*, Springer Proceedings in Complexity (eds H. Verhagen, M. Borit, G. Bravo, and N. Wijermans), Springer Nature Switzerland AG, Cham, ch. 5, pp. 47–62.
[41] Schelling, T.C. (1971) Dynamic models of segregation. *J. Math. Sociol.*, **1**, 143–186.
[42] Hägerstrand, T. (1965) A Monte Carlo approach to diffusion. *Eur. J. Sociol.*, **6** (1), 43–67.
[43] Axelrod, R. (1997) *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton University Press, Princeton, NJ.
[44] Edmonds, B. and Moss, S.J. (2005) From KISS to KIDS – an 'antisimplistic' modelling approach, in *Multi Agent Based Simulation 2004*, Lecture Notes in Artificial Intelligence, vol. 3415 (eds P. Davidson, B. Logan, and K. Takadama), Springer, Berlin, pp. 130–144.
[45] O'Sullivan, D. and Perry, G.L.W. (2013) *Spatial Simulation: Exploring Pattern and Process*, John Wiley & Sons, Ltd, Chichester.
[46] de Marchi, S. and Page, S.E. (2014) Agent-based models. *Annu. Rev. Polit. Sci.*, **17**, 1–20.
[47] Bianchi, F. and Squazzoni, F. (2015) Agent-based models in sociology. *Comput. Stat.*, **7** (4), 284–306.
[48] Bruch, E. and Atwell, J. (2015) Agent-based models in empirical social research. *Sociol. Methods Res.*, **44** (2), 186–221.
[49] Delli Gatti, D., Fagiolo, G., Gallegati, M. *et al.* (2018) *Agent-Based Models in Economics. A Toolkit*, Cambridge University Press, Cambridge.
[50] Manzo, G., Gabbriellini, S., Roux, V., and M'Mbogori, F.N. (2018) Complex contagions and the diffusion of innovations: evidence from a small-N study. *J. Archaeol. Method Theory*, **25** (4), 1109–1154.
[51] Sudgen, R. (2013) How fictional accounts can explain. *J. Econ. Methodol.*, **20** (3), 237–243.
[52] Sugden, R. (2000) Credible worlds: the status of theoretical models in economics. *J. Econ. Methodol.*, **7**, 1–31.
[53] Morgan, M.S. (2012) *The World in the Model: How Economists Work and Think*, Cambridge University Press, Cambridge.
[54] Zachrison, K.S., Iwashyna, T.J., Gebremariam, A. *et al.* (2016) Can longitudinal generalized estimating equation models distinguish network influence and homophily? An agent-based modeling approach to measurement characteristics. *BMC Med. Res. Methodol.*, **16**, 174. doi: 10.1186/s12874-016-0274-4.
[55] Christakis, N.A. and Fowler, J.H. (2007) The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.*, **357**, 370–379.
[56] Christakis, N.A. and Fowler, J.H. (2013) Social contagion theory: examining dynamic social networks and human behavior. *Stat. Med.*, **32**, 556–577.
[57] Grüne-Yanoff, T. (2009) The explanatory potential of artificial societies. *Synthese*, **169** (3), 539–555.
[58] Railsback, S.F. and Grimm, V. (2019) *Agent-Based and Individual-Based Modeling: A Practical Introduction*, 2nd edn, Princeton University Press, Princeton, NJ.
[59] Flache, A. and de Matos Fernandes, C.A. (2021) Agent-based computational models, in *Research Handbook on Analytical Sociology* (ed. G. Manzo), Edward Elgar, Cheltenham, UK ch. 24.
[60] Wicherts, J.M., Veldkamp, C.L.S., Augusteijn Hilde, E.M. *et al.* (2016) Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid p-hacking. *Front. Psychol.*, **7**, 1832.
[61] Young, C. and Holsteen, K. (2017) Model uncertainty and robustness: a computational framework for multimodel analysis. *Sociol. Methods Res.*, **46** (1), 3–40.
[62] Illari, P. (2011) Mechanistic evidence: disambiguating the Russo–Williamson thesis. *Int. Stud. Philos. Sci.*, **25** (2), 1–19.
[63] Williamson, J. (2019) Establishing causal claims in medicine. *Int. Stud. Philos. Sci.*, **32** (1), 33–61.

DOI: 10.1002/9781118445112.stat08510