Causality

GIANLUCA MANZO and LUCAS SAGE
Sorbonne University, France

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, 2011). 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 intrinsically superior to another. 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. Finally, we explain how computational tools like agent-based modeling (ABM) can aid causal inference 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, 2021).

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, 2004: 806). To put some order among them, an important distinction is between *dependence* (or *difference-making*) accounts of causality, and *production* accounts of causality (Hall, 2004). 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 propose similar typologies. For instance, Goldthorpe (2001) suggests distinguishing 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 concept, "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, 2011: 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 "what causes Y." For this reason, "what if" versus "why" causal questions are also often referred to as, respectively, the "effect-of-a-cause" and the "causes-of-an-effect" approaches (see, for instance, Dawid, Faigman, and Fienberg, 2014). 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 shifting the focus of

The Blackwell Encyclopedia of Sociology. Edited by George Ritzer and Chris Rojek.

DOI: 10.1002/9781405165518.wbeos1819

quantitative scholars from backward to forward causation (for a historical overview, see Imbens and Rubin, 2015: 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, 1986: 947). To get around this, the potential outcome 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 latter to units that are not. Thus, randomization makes it possible to interpret the control group as the counterfactual: that is, what would have happened had the units not been exposed to the treatment. The effect of a cause is then conceived as the average difference between the outcomes among those that were exposed to the treatment and those that were not.

From a methodological point of view, randomized experiments are seen as the prototypical method to implement the potential outcome approach (Gelman, 2011: 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, 2020: ch. 15). That is why the potential outcome approach is now often seen as a "unified framework for the prosecution of causal questions" (Morgan and Winship, 2014: 3).

Despite its ambition of generality, the potential outcome approach is intrinsically rooted in a specific understanding of causality, namely a counterfactual perspective, which is a form of dependence accounts of causality in Hall's above-mentioned 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, 1979; Hedstrom, 2009). 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, 1992: 297). To this, followers of the potential outcome approach retort that mechanism-based explanations in fact can be rigorously tested within a counterfactual framework (Morgan and Winship, 2014: ch. 10).

The problem here is that different conceptions of mechanisms enter the picture, which complicates the dialogue between scholars animated by different understandings 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, 2013: 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)" (Makovi and Winship, 2021). 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, Darden, and Craver, 2000: 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 sequences of microlevel changes that are supposed to create the dependence connections of interest (see Andersen, 2014a, 2014b).

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, 2004: 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 (see Russo and Williamson, 2007) according to which diverse methods can produce different types of data and arguments that are equally important to support persuasive causal claims.

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 discovering causal connections, compared to computational methods like ABM that *cannot*. In a nutshell, the supposed superiority of the former for causal inference come from the fact that they rely on empirical information whereas computational models like ABM are entirely made of substantive and formal assumptions (for a particularly clear statement, see Diez Roux, 2015: 101).

Randomized controlled trials and instrumental variables – respectively, one design-based and one model-based strategy to identify causal effects from within the potential outcome approach (see Morgan and Winship, 2014: 30–33) – can be used to illustrate the point that, contrary to this widespread view, experimental and observational methods, too, heavily depend on assumptions that cannot be tested empirically, thus ultimately making the causal claim at hand contingent on a complex mix of limited empirical information and theoretical arguments.

Randomized Controlled Trials (RCTs)

From a dependence perspective on causality, in order 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.*, 2010: 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 all confounders are ruled out. Since nothing other than randomness is responsible for units being in the treatment or in the control group, if we observe a difference in the probability of Y between the two groups, then this can only be caused by the treatment. The average difference between the treatment and the control group can therefore be inferred to be caused by X.

As noted by Deaton and Cartwright (2018: 2), the appeal of RCTs comes from the fact that randomization to the treatment seems to make the method require "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, 2007: 16–17) as well as the possible temporal heterogeneity of these effects within subjects (Sampson, Winship, and Knight, 2013: 13, 18–19). Turnarounds for tackling treatment response heterogeneity obviously exist (see Manski, 2013: 63–76), but these solutions require additional assumptions (see Morgan and Winship, 2014: 425–427), which, as admitted by the method's proponents themselves, are "credible to the degree that someone thinks it so" (Manski, 2003: 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 potential interferences among the units' potential outcomes as well as the absence of hidden variations in the treatment (Imbens and Rubin, 2015: 10). As noted by Sobel (2006: 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, 2013; on two-stage randomization, see Halloran and Hudgens, 2016). But, again, these designs, too, rely on further assumptions that are difficult to test empirically, and, as noted by Imbens and Rubin (2015: 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 require external and substantive knowledge to be defended.

Instrumental Variables (IVs)

The same holds for methods for observational data that try to recreate experimental conditions in nonexperimental settings. For this aim, IVs are one of the most common techniques (see Bollen, 2012). 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 "relevance" condition (Stock and Watson, 2010: 333) – *and* uncorrelated with the putative effect Y given X – the "exclusion" or "exogeneity" restriction (Gangl, 2013: 381). By seeing I as a sort of exogenous shock affecting X, so that the effect of I on Y only goes through 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, 2001: 73; emphasis added). However, similarly to RCTs, this again fails to be the case in practice. It is indeed difficult or impossible to determine empirically the validity of the relevance and the exogenous conditions.

As to the former, Bound, Jaeger, and Baker (1995: 446) demonstrated that, when the instrument I is "weak," that is, only marginally 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 does one establish whether an instrument I is weak? 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, 2001: 73).

As to the "exclusion" restriction, the problem seems even more severe. 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 to miss some confounders. That is why Gangl (2013: 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, 2010: 432) is required, which makes any causal inference made through IVs contingent on that knowledge (see also Rosenzweigh and Wolpin, 2000).

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 of contributing to causal inference, because entirely made of assumptions (see Diez Roux, 2015: 101), seems inaccurate. Typical identification strategies of causal effects are clearly not a simple matter of data *alone*.

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 Wilensky and Rand, 2015). 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, 2009: 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 a 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 different levels of analysis, and allows one to observe how dynamic microlevel changes progressively lead to macrolevel changes. For these reasons, ABM is intrinsically rooted within a view of mechanisms as multilevel dynamic systems of interacting units (see Vu *et al*., 2020), and thus well suited to implement a production view of causation (see Anzola, 2020).

ABMs are typically considered as disconnected from causal inference because, differently from experimental and observational methods, they rely entirely on assumptions. Thus, they cannot tell us anything about the world outside the world of the model (see Diez Roux, 2015: 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 coexisted. On the one hand, there were highly abstract models like Schelling's (1971) 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, there were data-driven models like Hägerstrand's (1965) 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, 1997: 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 allow. 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, 2005). When the KISS principle is prioritized, ABMs are envisaged as "tools to think with" (O'Sullivan and Perry, 2013: 14–15), whereas when one follows the KIDS principle, the goal is to design "high fidelity models" (de Marchi and Page, 2014).

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 multiplying (see Bianchi and Squazzoni, 2015: 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" (Bruch and Atwell, 2015) should be given priority.

In fact, there are at least three dimensions along which an ABM can gain realism. First, the microlevel 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 about how agents behave and influence each other. This dimension can be labeled "theoretical realism." Second, the sociodemographic features of an ABM as well as the parameters and functions adopted to implement its microlevel 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*., 2018: 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 data sets describing the target of interest – a form of empirical validation sometimes called "output validation" (see Delli Gatti *et al*., 2018: 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 exist. For instance, Manzo *et al.* (2018) wanted to explain why technological innovations spread faster and more widely among Indian and Kenyan potters with different religious backgrounds. To understand whether the structure of family ties within these religious subgroups may have impacted on adoption probabilities, Manzo and colleagues designed an ABM whose microlevel 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 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 subgroups 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 inescapable subjective character of judgment on a model's realism, see Sugden, 2013). The point rather is that, contrary to the common view that sees 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 (2000: 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 better 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 (2012: 337) with reference to earlier microsimulations in macroeconomics, "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 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.* (2016) built an ABM with the explicit intent to assess the extent to which Christakis and colleagues' regression modeling approach to the estimation of the causal effect of social ties on various health-related outcomes were able to do so when homophily may confound genuine social influences (for an overview of this debate, see Christakis and Fowler, 2013). To this aim, Zachrison *et al.* (2016: 3–4) designed an empirically calibrated ABM to create longitudinal network data of exactly the same kind as those studied by Christakis and colleagues 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 associates on the network data generated by their ABM, Zachrison *et al.* (2016: tables 1–5) found that, while Christakis and colleagues' 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 a mechanistic ground are very demanding because *in practice* full empirical calibration is never possible (see, for instance, Grüne-Yanoff, 2009). 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 defend their realism. Various sensitivity (see Delli Gatti *et al.*, 2018: 151–162) and robustness (see Railsback and Grimm, 2019: 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 (Flache and Matos Fernandes, 2021). And, after all, ABM does not seem special in this respect. As now popularized by the debates on the "researcher degree of freedom" problem, experimental designs and statistical methods for observational data involve a number of hidden choices and assumptions that require to be checked (see Wicherts *et al.*, 2016). To do so, one solution precisely is to turn to various forms of sensitivity analysis to assess the extent to which the empirical estimates produced by those methods are robust against different models' specifications (Young and Holsteen, 2017) and/or potential confounders (see Gangl, 2013: 385–390). And, these robustness checks within experimental and observation methods are as challenging as within an ABM. In Gangl's (2013: 399) words, these tools "would be degraded to little more than a computational exercise in absence of background empirical and theoretical knowledge suggesting the likelihood and extent of confounding" (Gangl, 2013: 399).

Conclusion

Statisticians and social and computer scientists have differently conservative attitudes regarding the conditions under which causal inference is possible, but the vast majority of them tend to associate causal inference with specific methods (see Gelman, 2011: 959). Our main message is that causal inference would gain from being understood at the intersection of different views on causation and mechanisms as well as by combining different methods incorporating these different views. Experiments, 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, at best, may be 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 argue that the uncertainty about the causal nature of a given relation between two happenings can in fact be reduced in (at least) two ways (see Illari, 2011). 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. However, on the other hand, data and arguments on how given entities, activities, and interactions combine to produce some behavior actually do 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 (Williamson, 2019), this is the main reason why data and arguments along dependence and difference-making lines should be recursively combined with data and arguments on mechanisms understood as multilevel dynamic systems of interacting entities and activities. We strongly advocate this perspective because it pushes one to combine experiments, statistical methods for observational data, and computational methods like ABM rather than establishing dubious hierarchies between them. Experimentalists, statistically oriented scholars, and simulationists all struggle with analogous 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 sensitivity and robustness procedures that are necessary to justify assumptions that are empirically unverifiable. Thus, method synergies seem more fruitful than method "warlordisms."

SEE ALSO: Causal Inference; Computational Sociology; Experimental Methods; General Linear Model; Mathematical Sociology; Micro-Macro Links; Multivariate Analysis; Path Analysis; Regression and Regression Analysis; Social Network Analysis

References

Andersen, O. (2014a) A field guide to mechanisms: part I. *Philosophy Compass*, 9 (4), 274–283. doi: 10.1111/phc3.12119.

Andersen, O. (2014b) A field guide to mechanisms: part II. *Philosophy Compass*, 9 (4), 284–293. doi: 10.1111/phc3.12118.

Angrist, J.D. and Krueger, A.B. (2001) Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic Perspectives*, 15 (4), 69–85.

Antonakis, J., Bendahan, S., Jacquart, P., and Lalive, R. (2010) On making causal claims: a review and recommendations. *The Leadership Quarterly*, 21, 1086–1120.

Anzola, D. (2020) Causation in agent-based computational social science, in *Advances in Social Simulation: Looking in the Mirror* (ed. H. Verhagen, M. Borit, G. Bravo, and N. Wijermans), Springer Proceedings in Complexity, Springer Nature Switzerland AG, Cham, pp. 47–62.

Axelrod, R. (1997) *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton University Press, Princeton, NJ.

Bianchi, F. and Squazzoni, F. (2015) Agent-based models in sociology. *Computational Statistics*, 7 (4), 284–306.

Bollen, K.A. (2012) Instrumental variables in sociology and the social sciences. *Annual Review of Sociology*, 38, 37–72.

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), Free Press, New York, pp. 51–64.

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. *Journal of the American Statistical Association*, 90 (430), 443–450.

Bruch, E. and Atwell, J. (2015) Agent-based models in empirical social research. *Sociological Methods and Research*, 44 (2), 186–221.

Cartwright, N. (2004) Causation: one word, many things. *Philosophy of Science*, 71, 805–819.

Cartwright, N. (2007) Are RCTs the gold standard? Contingency and dissent in science. Technical Report 01/07. CPNSS, LSE.

Christakis, N.A. and Fowler, J.H. (2013) Social contagion theory: examining dynamic social networks and human behavior. *Statistics in Medicine*, 32, 556–577.

Cox, D.R. (1992) Causality: some statistical aspects. *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 155 (2), 291–301.

Dawid, A.P., Faigman, D.L., and Fienberg, S.E. (2014) Fitting science into legal contexts: assessing effects of causes or causes of effects? *Sociological Methods & Research*, 43 (3), 359–390.

Deaton, A. (2010) Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48, 424–455.

Deaton, A. and Cartwright, N. (2018) Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210, 2–21.

Delli Gatti, D., Fagiolo, G., Gallegati, M., *et al.* (2018) *Agent-Based Models in Economics: A Toolkit*, Cambridge University Press, Cambridge.

de Marchi, S. and Page, S.E. (2014) Agent-based models. *Annual Review of Political Science*, 17, 1–20.

Diez Roux, A.V. (2015) The virtual epidemiologist: promise and peril. *American Journal of Epidemiology*, 181 (2), 100–102.

Edmonds, B. and Moss, S.J. (2005) From KISS to KIDS: an "antisimplistic" modelling approach, in *Multi Agent Based Simulation 2004* (ed. P. Davidson *et al.*), vol. 3415 of Lecture Notes in Artificial Intelligence, Springer, Berlin, pp. 130–144.

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.

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.

Gelman, A. (2011) Causality and statistical learning. *American Journal of Sociology*, 117 (3), 955–966.

Goldthorpe, J.H. (2001) Causation, statistics and sociology. *European Sociological Review*, 17 (1), 1–20.

Grüne-Yanoff, T. (2009) The explanatory potential of artificial societies. *Synthese*, 169 (3), 539–555.

Hägerstrand, T. (1965) A Monte Carlo approach to diffusion. *European Journal of Sociology*, 6 (1), 43–67.

Hall, N. (2004) Two concepts of causation, in *Causation and Counterfactuals* (ed. J. Collins, N. Hall, and L.A. Paul), MIT Press, Cambridge, MA, pp. 225–276.

Halloran, M.E. and Hudgens, M.G. (2016) Dependent happenings: a recent methodological review. *Current Epidemiology Reports*, 3 (4), 297–305. doi: 10.1007/s40471-016-0086-4.

Hedstrom, P. (2009) Studying mechanisms to strengthen causal inferences in quantitative research, in *The Oxford Handbook of Political Methodology* (ed. J.M. Box-Steffensmeier, H.E. Brady, and D. Collier), Oxford University Press, Oxford, pp. 319–335.

Hernán, M.A. and Robins, J.M. (2020) *Causal Inference: What If*, Chapman & Hall/CRC, Boca Raton, FL.

Holland, P.W. (1986) Statistics and causal inference. *Journal of the American Statistical Association*, 81 (396), 945–960.

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.

Illari, P. (2011) Mechanistic evidence: disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science*, 25 (2), 1–19.

Imbens, G.W. and Rubin, D.B. (2015) *Causal Inference for Statistics, Social, and Biomedical Sciences*, Cambridge University Press, Cambridge.

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.

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

Makovi, K. and Winship, C. (2021) Advances in mediation analysis, in *Research Handbook on Analytical Sociology* (ed. G. Manzo), Edward Elgar, Cheltenham.

Manski, C.F. (2003) *Partial Identification of Probability Distributions*, Springer, New York.

Manski, C.F. (2013) Identification of treatment response with social interactions. *Econometrics Journal*, 16 (1), S1–S23.

Manzo, G. (2021) *Agent-Based Models and Causal Inference*, Wiley Blackwell, Chichester.

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, *Journal of Archaeological Method and Theory*, 25 (4), 1109–1154.

Morgan, M.S. (2012) *The World in the Model: How Economists Work and Think*, Cambridge University Press, Cambridge.

Morgan, S.L. and Winship C. (2014) *Counterfactuals and Causal Inference: Methods and Principles for Social Research*, 2nd edn, Cambridge University Press, Cambridge.

O'Sullivan, D. and Perry, G.L.W. (2013) *Spatial Simulation: Exploring Pattern and Process*, Wiley Blackwell, Chichester.

Railsback, S.F. and Grimm, V. (2019) *Agent-Based and Individual-Based Modeling: A Practical Introduction*, 2nd edn, Princeton University Press, Princeton, NJ.

Rosenzweigh, M.R. and Wolpin, K.I. (2000) Natural "natural experiments" in economics. *Journal of Economic Literature*, 38, 827–874.

Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21 (2), 157–170.

Sampson, R.J., Winship, C., and Knight, C. (2013) Translating causal claims, principles and strategies for policy-relevant criminology. *Criminology & Public Policy*, 12 (4), 587–616.

Schelling, T.C. (1971) Dynamic models of segregation. *Journal of Mathematical Sociology*, 1, 143–186.

Sobel, M.A. (2006) What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. *Journal of the American Statistical Association*, 101 (476), 1398–1407.

Stock, J.H. and Watson, M.W. (2010) *Introduction to Econometrics*, 3rd edn, Addison-Wesley, Boston, MA.

Sugden, R. (2000) Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, 7, 1–31.

Sugden, R. (2013) How fictional accounts can explain. *Journal of Economic Methodology*, 20 (3), 237–243.

Vu, T.M., Probst, C., Nielsen, A., *et al*. (2020) A software architecture for mechanism-based social systems modelling in agent-based simulation model. *Journal of Artificial Societies and Social Simulation*, 23 (3), 1. Available at http://jasss.soc.surrey.ac.uk/23/3/1.html (accessed May 19, 2021).

Wicherts Jelte, M., Veldkamp Coosje, 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. *Frontiers in Psychology*, 7, 1832.

Wilensky, U. and Rand, W. (2015) *An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo*, MIT Press, Cambridge, MA.

Williamson, J. (2019) Establishing causal claims in medicine. *International Studies in the Philosophy of Science*, 32 (1), 33–61.

Wooldridge, M. (2009) *An Introduction to MultiAgent Systems*, Wiley Blackwell, Chichester.

Young, C. and Holsteen, K. (2017) Model uncertainty and robustness: a computational framework for multimodel analysis. *Sociological Methods & Research*, 46 (1), 3–40.

Zachrison, S.K., 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 Medical Research Methodology*, 16, 174. doi: 10.1186/s12874-016-0274-4.