

Agent-based Models and Causality

A Methodological Appraisal

Lorenzo Casini, University of Geneva

Gianluca Manzo, CNRS; University of Paris-Sorbonne

Agent-based Models and Causality

A Methodological Appraisal

Lorenzo Casini^{*}

Department of Philosophy, University of Geneva

Gianluca Manzo[†]

GEMASS – CNRS and University of Paris-Sorbonne

Draft of December 21, 2016[‡]

Computational agent-based models are entering the toolbox of quantitative sociologists. However, markedly contrasting views still exist as to its capacity to contribute to causally-oriented empirical research. Building on selected works across disciplines ranging from computer science to philosophy, we connect scholarship on causality, mechanisms, and simulation methods, and provide the first systematic discussion on how, if at all, computational agent-based models warrant causal inference. First, we argue that this method can produce causally-relevant evidence when (and only when) specific conditions are met. Then, we show that data-driven methods for causal inference face analogous challenges. Finally, upon endorsing a pragmatist view of evidence, we defend an approach to causal analysis that combines evidence from agent-based modeling and data-driven methods. This evidential variety lends credibility to causal inference in virtue of drawing on complementary, and equally important, kinds of evidence.

^{*}lorenzo.casini@unige.ch

[†]gianluca.manzo@cnrs.fr

[‡]We thank the audiences of ‘International Network of Analytical Sociologists’ (INAS), Harvard University, 12–13 June 2015, and ‘Causality and Modelling in the Sciences’ (CaMitS), UNED, Madrid, 29 June–1 July 2015, where a preliminary version of this paper was presented. One author’s work (LC) was generously supported by the Swiss National Science Foundation (grant no. CRSII 1.147685/1).

CONTENTS

INTRODUCTION	2
1 CAUSALITY, MECHANISMS, AND ABMS	6
1.1 <i>Causal inference</i>	7
1.2 <i>Mechanisms</i>	10
1.3 <i>ABMs</i>	14
2 A VARIETY OF ABMS	19
2.1 <i>Historical roots of the ABMs' diversity</i>	19
2.2 <i>From just-so stories towards more realistic models</i>	21
2.3 <i>ABMs for causal inference—a road-map</i>	23
3 ABMS AND CAUSAL INFERENCE	29
3.1 <i>The ideal case</i>	29
3.2 <i>Practical limitations</i>	32
3.3 <i>Theoretical explorations</i>	35
4 ABMS AND DATA-DRIVEN METHODS	38
4.1 <i>Data availability</i>	40
4.2 <i>Truth of the assumptions</i>	45
4.3 <i>Reliability of the method</i>	50
5 METHODOLOGICAL SYNERGY	56
DISCUSSION AND CONCLUSION	62
REFERENCES	69

INTRODUCTION

Areas of research focusing on causality, on mechanism-based explanations, and on agent-based computational modeling seem especially lively in contemporary sociology (for recent reviews, see respectively [Gangl 2010](#); [Hedström and Ylikoski 2010](#); [Bianchi and Squazzoni 2015](#)).

As to causality, there is evidence that establishing causal claims is the goal of a vast majority of empirical articles published in leading U.S. journals. This holds for both quantitative and ethnographic studies ([Abend et al., 2013](#)). One may speculate that the importance of establishing causal claims is evidenced by the regularity with which methodological discussions on causality have appeared within a variety of sociological perspectives (see, e.g., [Marini and Singer 1988](#); [Abbott 1998](#); [Doreian 1999](#); [Goldthorpe 2001](#); [Winship and Sobel 2004](#); [Mahoney 2008](#); [Mahoney and Ragin 2013](#); for a historical perspective, see also [Barringer et al. 2013](#)). One may also expect that this trend will

be reinforced by the recent and rapid diffusion in sociology of the potential outcome approach to causality ([Morgan and Winship, 2014](#)), an approach in turn fostered by older contributions in statistics (for a historical overview, see [Imbens and Rubin 2015](#), ch. 2) and economics ([Heckman, 2005](#)), and reinvigorated by recent developments in computer science ([Spirtes et al., 2000](#); [Pearl, 2009](#)) as well as philosophical discussions ([Woodward, 2003](#)).

As to mechanism-based explanations, we also observe a booming trend. It is true that the effort towards identifying generalizable fine-grained chains of small-scale events with clearly defined large-scale consequences can be traced back to the infancy of modern social sciences (see, e.g., [Elster 2009b](#)'s study of Tocqueville's œuvre), and that the notions of "social mechanism" and "generative model" were initially forged by mathematical sociologists in the Sixties (see, respectively, [Karlsson 1958](#), 16, [Fararo 1969b](#), 81, 84-5, [Fararo 1969a](#), 225; see also [Boudon 1979](#)). At the same time, it seems correct to regard [Hedström and Swedberg \(1998\)](#)'s collection on social mechanisms, coupled with philosophical studies of research practices in biology and neuroscience ([Machamer et al., 2000](#); [Bechtel and Abrahamsen, 2005](#); [Craver, 2007](#)), as the starting point of a new era of systematic investigation on the concept of mechanism-based explanation. As a by-product of this investigation, analytical sociology has progressively emerged as a distinctive style of social inquiry (see, among others, [Hedström 2005](#), ch. 6, [Hedström and Bearman 2009](#); [Demeulenaere 2011](#); [Manzo 2014a](#)). This, in turn, has triggered a considerable amount of critical reactions, which testimony how the quest for mechanism-based explanation can in fact be framed in very different ways (for some critiques, see [Abbott 2007](#); [Gross 2009](#); [Gorski 2009](#); [Boudon 2012](#); [Little 2012](#); [Lizardo 2012](#); [Sampson 2011](#); [Sawyer 2011](#); [Opp 2013](#); for a reply, see [Manzo 2010, 2014b](#)).

As to agent-based computational models, hereafter ABMs (or ABM, for the singular form and for "agent-based modeling"), the trend is even more spectacular. Although the basic principles spontaneously appeared in pioneering studies in the Sixties ([Hägerstrand, 1965](#); [Sakoda, 1971](#); [Schelling, 1971](#)), the diffusion of this type of simulation models significantly accelerated after the publication of systematic monographs such as ([Axtell and Epstein, 1996](#)), ([Axelrod, 1997](#)), and ([Epstein, 2006](#)). Nowadays, pleas for ABMs exist in a large variety of disciplines—including biology ([Thorne et al., 2007](#); [Chavali et al., 2008](#)), ecology ([Grimm et al., 2006](#)), macroeconomics ([Farmer and Foley, 2009](#); [De Grauwe, 2010](#)), quantitative finance ([Mathieu et al., 2005](#)), organization and marketing studies ([Fioretti, 2013](#)), political science ([Cederman, 2005](#)), geography ([O'Sullivan, 2008](#)), criminology ([Birks et al., 2012](#)), epidemiology ([Auchincloss and Roux, 2008](#)), social psychology ([Smith and Conrey, 2007](#)), demography ([Billari and Prskawetz, 2003](#)) and archeology ([Wurzer et al., 2015](#)). Sociology is no exception ([Macy and Flache, 2009](#); [de Marchi and Page, 2014](#)). Leading journals have started paying attention to ABMs ([Gilbert and Abbott, 2005](#); [Hedström and Manzo, 2015](#)) and the number of applications at the core of the discipline is fast increasing ([Macy and Willer, 2002](#); [Sawyer, 2003](#); [Bianchi and Squazzoni, 2015](#)).

Now, although each of these strands of literature is burgeoning, conceptual and methodological connections among them are still limited. To our mind, certain connections are particularly intuitive. Mechanisms may be counterfactually or probabilistically interpreted. Models of mechanisms are thus of direct relevance to causal inference, which is, too, concerned with probabilistic and counterfactual claims. In turn, ABMs

are a class of formal models, nowadays regarded by many social scientists as a powerful tool for studying *social* mechanisms. This, in our view, naturally prompts the question as to whether ABMs can warrant causal inference in the social domain on such a mechanistic ground. Yet, to date, there is no systematic discussion on the three-way connection between causality, mechanisms, and ABMs.

Hedström and Ylikoski (2010), for instance, reflect on both the concept of cause and mechanism, but, when they treat ABMs, the issue of how this method contributes to causal inference is not addressed. Knight and Winship (2013) criticize the way the concept of mechanism is employed within the analytical sociology literature; they propose a more precise definition of the concept, which they regard as compatible with a counterfactual view of causation, and show how this definition can be employed, using directed acyclic graphs, to *identify* causal relations; at the same time, however, ABMs are not considered. Watts (2014) reflects on the notion of causal explanation in connection with a critical analysis of a specific aspect of the mechanism-based perspective, namely action theory, but, when he addresses the methodological side of the issue, experimental and statistical methods are only quickly discussed and no attention is devoted to ABMs. Philosophical investigations exhibit a similar pattern. Several articles scrutinize the connection between the concepts of causal and mechanistic explanation (Glennan, 1996, 2002; Woodward, 2002, 2013; Menzies, 2012; Casini et al., 2011; Williamson, 2013); however, the discussion of what techniques would support the connection between methods for causal inference and strategies for mechanistic explanation is still limited (for a few exceptions, see, Steel 2004; Reiss 2009; Mouchart and Russo 2011; Hoover 2012). Analogously, among philosophical contributions discussing simulations, there is some discussion on the explanatory power of ABMs (see, e.g., Grüne-Yanoff 2009a; Casini 2014), but no systematic connection to theories of causation and methods for causal inference.

In this paper, our ambition is to fill this conceptual and methodological gap by developing a set of guidelines for connecting the sociological scholarship on methods for causal inference, social mechanisms, and ABMs. In particular, we are concerned with answering the following question: Can ABMs, in virtue of supporting mechanism-based explanation, also warrant causal inference? An answer to this question is very much needed. In fact, although some consensus exists on ABM's ability to produce mechanism-based explanations (Hummon and Fararo 1995b; Axtell et al. 2002; Cederman 2005; Epstein 1999, Epstein 2006, chs. 1-2, Sawyer 2004; Tesfatsion 2006; Manzo 2014b), strong disagreement remains on the extent to which ABMs can also provide empirical support for the existence of the postulated mechanisms (for a critique, see Grüne-Yanoff 2009a; for a defense, see Ylikoski and Aydinonat 2014). Thus, it is unclear also whether ABMs can aid causal inference on such mechanistic grounds.

Among simulation practitioners, Macy and Sato (2008) claim that “[t]he computational model can generate hypotheses for empirical testing, but it cannot ‘bear the burden of proof’”, thus implicitly proposing a division of labor according to which ABM is a tool for theoretical exploration while experimental and statistical methods for observational data are better suited, and necessary, to support causal inference. This view seems shared by quantitative scholars developing the latter methods. Morgan and Winship (2014, 341), for instance, express skepticism about “[...] the utility of many simulation-based methods of theory construction”. As clearly visible in Morgan (2013)'s overview

of the most recent developments in the field of causal inference, ABM is simply not considered as a potential player in this game. Others, however, have noted that an ABM can communicate with empirical data in different ways, which in principle makes it capable of contributing to the discovery of real-world, mechanism-mediated causal relations (Hedström, 2005; Manzo, 2007; Bruch and Atwell, 2015). Similarly contrasting positions can be found outside sociology, for instance in epidemiology. Marshall and Galea (2014), for instance, have recently argued that ABMs can be used to support causal inference by in turn supporting counterfactual reasoning. To them, Diez Roux (2014) has objected that

[...] there is a fundamental distinction between causal inference based on observations (as in traditional epidemiology) and causal inference based on simulation modeling. The traditional tools of epidemiology are used to extract (hopefully) reasonable conclusions from necessarily partial and incomplete (often messy) observations of the real world. [...] In contrast, when we use the tools of complex systems, we create a virtual world (based on prior knowledge or intuition) and then explore hypotheses about causes under the assumptions encoded in this virtual world. In the simulation model, we cannot directly determine whether X causes Y in the real world (because the world in which we are working is of our own creation); we can only explore the plausible implications of changing X on levels of Y under the conditions encoded in the model. In the real world, we have fact (what we observe) and we try to infer the counterfactual condition (what we would have observed if the treatment had been different). In the simulated world, everything is counterfactual in the sense that the world and all possible scenarios are artificially created by the scientist. (Diez Roux, 2014, 101)

This is the clearest illustration we could find of a contrast, which is implicitly (often informally) drawn between methods that aim to establish conclusions as regards “potential” outcomes on the one hand, and ABMs the other hand. On the one hand, there are traditional and widely-used methods for causal inference, which have received a unified counterfactual interpretation along the lines of the potential outcome framework, as formalized by Rubin (1974) and adopted by (among others) Heckman, Pearl and Woodward (cf. Morgan and Winship, 2014, 4-5). On the other hand, there is a novel method, viz. ABM, which produces “simulated” outcomes, and whose value for causal inference has not yet been discussed, let alone established.

These markedly different views raise two important questions: First, where does the disagreement originate? Second, is a reconciliation between these views possible? §1–§2 of our paper answer the former question, whereas §3–§4 accumulate elements to address the latter question, which is finally answered in §5. More precisely, our analysis goes through the following steps. In §1, we consider the notions of causality and mechanism, and suggest that any argument about one favorite method’s capacity to aid causal inference and model social mechanisms implicitly relies on one specific understanding of these two notions. Moreover, we explain why ABM is, on a technical level, especially compatible with one specific understanding of the concept of mechanism. This helps us diagnose why some are sympathetic towards using ABMs in causal research, while others are against it. In §2, we provide a meta-analysis that categorizes the variety of ABMs in the literature based on the kind of *phenomena* they are supposed to explain, the kind

of *information* that is used to build them, and the kind of *operations* that are performed to assess their validity. We argue that all three dimensions have a bearing on causal inference; existing opinions on the usefulness of ABMs in causal research neglect this variety, and are thus ill-founded. With such qualifications in place, in §3 we address the central issue of the paper by discussing the *in-principle* conditions for ABM to support causal inference, which will turn out related to the three dimensions uncovered by the previous section, the *in-practice* obstacles to the realization of these conditions, and the research strategies—or “theoretical explorations”—available to ABM modelers when these *in-principle* conditions cannot be satisfied. Following our discussion of the challenges faced by ABM, in §4 we illustrate how analogous issues arise in a selection of data-driven methods (in particular, randomized experiments, instrumental variables, and causal graphs). We discuss the obstacles faced by such methods, in terms of data availability, untestable assumptions, and method reliability. Finally, in §5, building on the important—but often under-estimated—fact that, similarly to ABM, data-driven methods for causal inference may generate evidence for causality only if specific conditions are fulfilled, we argue for the usefulness of shifting the focus of the debate from proposing arguments that defend the superiority of particular methods, arguments based on an (often implicit) endorsement of particular notions of causality and mechanism, to discussing how different methods may coexist and ought to cooperate. This discussion is going to be premised on the assumption that different kinds of evidence are required before the scientific community may safely accept a causal claim. According to a view that we shall call “evidential variety”, different methods may produce different kinds of evidence for a given causal claim. In consequence, since *in practice* every kind of evidence is likely to be imperfect, these methods should be combined to compensate for each other’s weaknesses. The paper is closed by a [Discussion and conclusion](#) section summarizing the major steps of the analysis and discussing possible objections to this plea for a methodological synergy for causal analysis in sociology.

To conclude this introduction, let us remark that our paper echoes similar efforts of conceptual and methodological clarification that have been produced in the past with respect to other research traditions. For instance, when small-N research reached a critical mass, some felt necessary to investigate in what sense small-N studies allow causal inference ([Mahoney, 2000](#)). As to ABMs, a similar assessment is still missing. Given the current diffusion of this method and the variety of views expressed about its potential, we believe that a systematic analysis of its contribution to causal inference is now required.

1 CAUSALITY, MECHANISMS, AND ABMS

In this section we first provide a *theoretical clarification* about the concepts of causality (§1.1) and social mechanism (§1.2). This is needed to defend our claim that the observed disagreements on the usefulness of ABM for causal inference ultimately arises from the fact that scholars in different methodological traditions endorse conflicting views on what establishing causality and identifying mechanisms mean. On this basis, we then explain why the technical infrastructure of an ABM is especially apt to implement one specific understanding of mechanism, thereby supporting a specific view of causality

(§1.3).

As to the sense in which we shall employ the term *causal inference*, two remarks should suffice. First, we distinguish between statistical and causal inference. The former concept usually indicates the estimation (and generalization) of the value of a (set of) parameter(s), with its associated uncertainty, from a limited set of observations (see [Cox 2006](#), 7, and [Snijders and Steglich 2015](#)). By causal inference, we refer in contrast to the more general operation of using limited information to establish the existence of a non-spurious connection between two properties of the world. There are obvious connections between statistical and causal inference, as for instance evidenced by the use of the statistical significance of an estimated parameter to infer the presence of a causal connection between two variables (more on the challenges faced by this practice in §4). However, as stressed by [Pearl \(2010, 2\)](#), these operations do not overlap. Second, we do not tie the concept of causal inference to a specific kind of evidence or inquiry (unlike, for instance, [Hedström 2009](#), who interprets the concept of causal inference in terms of that of social mechanism). As we shall argue later (§5), there are indeed good reasons to believe that different kinds of evidence can contribute to convince an external observer of the causal nature of an association. Hence, a more neutral definition of the concept of causal inference is the most fruitful to frame our discussion.

1.1 Causal inference

According to [Cartwright \(2004\)](#),

[t]he term ‘cause’ is highly unspecific. It commits us to nothing about the kind of causality involved nor about how the causes operate. Recognizing this should make us more cautious about investing in the quest for universal methods for causal inference. (*ibid.*, 806)

The quote emphasizes the difficulty to provide a single, unitary, and uncontroversial account of causality. For us, this philosophical observation has the following implication. To the extent that the concept of causality can be understood in different ways, it is likely that the views of those favoring/opposing ABM as a relevant method for causal inference in fact depend on their (implicit) intuitions on the nature of causality. Thus, making explicit such intuitions is necessary for a balanced assessment of the usefulness of ABM for causal inference compared to other methods.

Scholarship on causality in philosophy is helpful in this respect. It provides theoretical coordinates to map the different ways we can conceive of a causal relationship. For our purposes, the most relevant distinction is between *dependence* (or *difference-making*) accounts of causality, and *production* accounts of causality (on the distinction between the two notions, see [Hall, 2004](#)). Roughly, among dependence accounts, one finds regularity, probabilistic and counterfactual views of causality whereas, among production accounts, one finds process, entities-and-activities and dispositionalist accounts (for similar categorizations, see [Kistler 2002](#), [Psillos 2007](#) and [Reiss 2013](#), ch. 5). Two basic intuitions inspire these two groups of theories of causation. The idea behind dependence accounts is that causes are such that their obtaining makes a difference to the

obtaining of their effects. In contrast, the idea behind production accounts is that causes as such that they help generate, or bring about, their effects.¹

Sociologists have developed similar categorizations. A notable example is Goldthorpe (2001), who has remarked that causation can be interpreted as: (1) “robust dependence”—in this case, the causal claim depends on showing that X continues to affect Y when a set Z of other variables, also possibly related to Y , are introduced in the analysis (*ibid.*, 2); (2) “consequential manipulation”—in this case, “genuine causation is that if a causal factor, X , is manipulated, then, given appropriate controls, a systematic effect is produced on the response variable, Y ” (*ibid.*, 4-5); or, (3) “generative process”—in this case, “[...] what is important is the nature and the validity of the account given of the process that underlies the association appealed to [...]” (*ibid.*, 9). In the terms of the aforementioned philosophical categories, causation as “robust dependence” and “consequential manipulation” clearly exemplify dependence accounts of causality, whereas what Goldthorpe labels causality as “generative process” naturally falls within production accounts.

In addition, Goldthorpe observes that these views on causality combine in practice with distinctive methods of social inquiry. The view that causality essentially depends on controlling for confounders has informed time-series analysis, the early generation of causal path analysis, structural equations models, and more generally the large panoply of multivariate quantitative, regression-like techniques for survey data analysis. The “consequential manipulation” view squares with the methodology of randomized experiments. Interestingly, Goldthorpe hesitates to identify a specific method that illustrates the view of causality as “generative process”. Although he sees the potential of simulation methods as a possible option to test the validity of a proposed account of the underlying process (*ibid.*, 14), he still prioritizes statistical methods with respect to the goal of testing the hypothesized direct and indirect consequences of a postulated underlying process (*ibid.*, 12-3)—a view that he restates in his latest book (2016, ch. 9).

The association between specific methods for causal inference and views on causation is especially visible in the methodological literature on the so-called “potential outcomes”. This is driven by the ambition to introduce the perspective of randomized experiments into the analysis of data generated outside an experimental setting (for a historical overview, see Imbens and Rubin, 2015, ch. 2). Accordingly, the main task of the analysis becomes to show that individuals (or, other units of analysis) that are exposed to different treatment states are likely to exhibit different responses, or outcomes. The causal effect of a given treatment state is then conceived as the difference between the outcome of those who were exposed to it and that of those who were not. In this way, the potential outcome approach is in essence tied to a counterfactual understanding of causation. Establishing causal claims indeed amounts to quantify *what-if* outcomes, viz. how a given group of units of analysis would have responded, had their treatment value been different. As noted by Morgan and Winship (2014, 4), *what-if*, or potential, outcomes, are counterfactual in the sense that they “exist in theory but are not observed”.

¹For Hall, dependence and production accounts are irreducible to one another, so we have distinct concepts of cause. In the present investigation, we remain agnostic on what causality essentially is. In §5, we shall discuss however the undesirable consequences, on a methodological level, of maintaining that distinct and irreducible notions of cause exist.

For our purposes, the important point here is that the potential outcome approach, with its counterfactual understanding of causation, is now regarded by many as a “unified framework for the prosecution of causal question” (Morgan and Winship, 2014, 3). As such, it is seen as a tool that allows one to recast traditional multivariate statistical instruments in the terms of this particular view of causality. In this regard, Morgan and Winship’s discussion of matching and regression estimators (2014, chs. 5-7) is especially illuminating. They elegantly show how the classic method of controlling for confounders can be reinterpreted as 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.

Whilst we agree that the language of potential outcomes provides a powerful interpretive key to more rigorously think of statistical methods for observational data, we should point out that this interpretive key is still highly specific. In terms of the aforementioned philosophical distinctions, the counterfactual view is a type of dependence, or difference-making, account of causality. From within a “production” perspective, it may be regarded as limited, in the sense that, in David Cox’ words, it lacks “an explicit notion of an underlying process or understanding at an observational level that is deeper than that involved in the data under immediate analysis” (1992, 297). To this, Cox adds: “my preference, however, is to restrict the term [causality] to situations where some explanation in terms of a not totally hypothetical underlying process or mechanism is available”.

Thus, for one thing, it is clear that statisticians and sociologists subscribe to different accounts of causality. For another thing, these different accounts tend to come with different judgments as regards their explanatory depth. Goldthorpe (2001, 8-9), for instance, clearly states that the view of causation as “generative process” should be seen as an improvement on the “robust dependence” and “consequential manipulation” accounts because “it would appear to derive, rather, from an attempt to spell out what must be added to any statistical criteria before an argument for causation can convincingly be made”. Hedström (2009), similarly, remarks that only the presence of a fully-fledged mechanism authorizes causal inference and allows one to reach explanatory depth. To be sure, scholars within the potential outcome tradition would find this priority judgement unjustified because, so they would claim, mechanism-based explanations can be easily formulated within a counterfactual view and tested by an appropriate use of statistical methods (cf. Morgan and Winship, 2014, ch. 10). As we shall see next, however, different understandings of the concept of mechanism are at work here.

As for now, the point we want to make is that, since there are different intuitions on causality, if one evaluates the usefulness of ABM for causal inference on the basis of one’s intuition, one is likely to reach contrasting conclusions. On the one hand, dependence accounts rely on quantitative tools that prioritize finding non-spurious relationships (control for confounders), establishing counterfactual claims (*what-if* outcomes), and when possible, estimating unbiased parameters (correct standard errors) that quantify such relations, in a way that allows for the extrapolation from a sample, or test population, to an unobserved target population. From this perspective, ABM may seem unnecessary to establish causation: what matters is data quality and how creatively one is able to describe these data. On the other hand, production accounts of causality prioritize finding a credible narrative that accounts for the observed patterns. In this

sense, production accounts value more the identification of the mechanism responsible for the phenomena, hence explanation, than prediction and intervention. The idea is that dependence relations are not constitutive of causality but rather the manifestation of it. Within this perspective, ABM appears as a crucial tool to establish causation: it provides a formal device to prove that the dependence relationship under scrutiny are deducible by unfolding the postulated (formalized) narrative.²

Now, no matter how clashing these views may appear at first, we believe that they in fact can, and should, be reconciled. Later on, especially in §3.2 and §4, we shall accumulate elements that suggest that proper causal inference requires a combination of dependence and production accounts of causation, thus a synergy between experimental and statistical methods for observational data on the one hand, and ABM on the other. We shall fully develop this argument when defending our “methodological synergy” thesis (§5). For now, let us take at face value the disagreement observed in the literature, and turn to a second source of heterogeneity of intuitions, namely the concept of mechanism, which contributes to generate this disagreement.

1.2 Mechanisms

As we have seen, causality accounts that fall within the “production”—as opposed to the “dependence”—camp requires the identification of an underlying mechanism for inferring causality from data. Thus, a second issue that then arises when assessing whether ABM can aid causal inference is whether this method is necessary to model mechanisms. In answering this question, a second source of potential disagreement becomes visible.

Similarly to the concept of causality, the concept of (social) mechanism, too, has received a variety of interpretations, both in philosophy (Reiss, 2013, 104-5) and in the social sciences (in sociology, see Mahoney 2001, 579-80, Hedström 2005, 25, and Gross 2009, 360-2; in political science, see Gerring 2008). As recently observed by Kalter and Kroneberg (2014), the term *mechanism* has clearly penetrated much empirical research in sociology but it is still employed with a variety of meanings. Mapping their variety is important because judgments on the appropriateness of ABM for studying social mechanisms, and thus for helping causal inference on mechanistic grounds, are likely to be sensitive to the understanding of this concept.

Philosophical scholarship on mechanisms provides useful theoretical coordinates. For our purposes, the most relevant distinction is between what we shall label “horizon-

²The different reactions to Lucas (1976)’ critique to causal inference in macroeconomics (e.g., to the claim that inflation causes employment) provide a nice historical illustration of the difference between the two camps. In the “dependence” camp, there was a data-driven reaction, which emphasized the centrality of intervention-like methods and led to a more sophisticated use of statistics, the diffusion of time-series econometric models, and the development of vector autoregression (VAR) methods (Sims, 1980), in the tradition of Granger (1969). In contrast, in the “production” camp, there was a theory-driven reaction, which demanded that macroeconomic models be enriched with “micro-foundations”. This led initially to the intense use of rational choice models calculating economic aggregates based on individual preferences and expectations—a development encouraged by Lucas (1976) himself—and, consequently to the critique of representative-agent assumptions (see Sargent 1993 and Kirman 1992; cf. Hoover 2008a,b), to introducing agent-based computational models for solving the aggregation problem in the presence of heterogeneity (Tsfatsion, 2002, 2006; Arthur, 2006; Kirman, 2010).

tal” and “vertical” views of mechanisms (the rationale behind this terminological choice will become evident soon). According to the former view, a mechanism is interpreted as a network of variables that stand in particularly robust relations.³ In contrast, according to the vertical view, a mechanism is envisaged as a “complex system” (Glennan, 2002, S344) comprising a set of unities—entities and activities (Machamer et al., 2000), or component parts and operations (Bechtel and Abrahamsen, 2005), or parts and interactions (Glennan, 2002)—that, by interacting over time, generate some behavior of the system.⁴

The vertical view is not incompatible with paying attention to the robustness of the relationships between the interacting parts that compose the mechanism. Yet, on this view the distinctive feature of a mechanism are the dynamics of the changes a mechanism brings about. From their activity-centered perspective, Machamer et al. (2000) put this point by saying that “[e]ntities often must be appropriately located, structured, and oriented, and the activities in which they engage must have a temporal order, rate, and duration” (*ibid.*, 3) and that “[a] description of 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” (*ibid.*, 12). From his interaction-centered perspective, Glennan (2002, S344) makes the same point when he remarks that “[a] mechanism operates by the interaction of parts. An interaction is an occasion on which a change in a property of one part brings about a change in a property of another part”. Thus, what matters to the vertical view is the sequence of micro-level changes that dynamically generate a given behavior or connection of interest. Our use of the term *vertical*, as opposed to *horizontal*, is primarily meant to capture this productive relationship between the *explanans* and the explanandum, and, more particularly, the granularity of the details provided to account for this relationship. (More on granularity in the next subsection.)

Without using the philosophical terminology, sociologists have engaged in a dispute about the merits and limitations of the horizontal and vertical accounts of mechanisms since the Seventies. Among quantitatively-oriented scholars, the confrontation between the two views already appeared clearly in the muscular critique Hauser (1976) moved against Boudon (1974)’s study of the temporal connection between inequality of educational and social opportunity in western countries. Hauser’s most general point was that Boudon did not make use of the best-developed framework for multivariate causal modeling at the time, namely path analysis. This, according to Hauser, meant that Boudon’s results were based on fragile assumptions, and that Boudon’s tests of validity of his model were weak. Although Boudon (1976) acknowledged that some of

³Woodward (2002) exemplifies this view when he defines a mechanism as “an organized or structured set of parts or components, where (ii) the behavior of each component is described by a generalization that is invariant under interventions, and where (iii) the generalizations governing each component are also independently changeable” (*ibid.*, S375). “Invariance” (which is the technical name for this sort of robustness) is verified by “ideal” interventions. An ideal intervention *I* on a putative cause X_i with respect to a putative effect X_j is such as to set the value of X_i , so that any change in X_j following *I* is to be ascribed to X_i —that is, *I* does not directly cause X_j , or cause or is statistically correlated with any X_k , which causes X_j and does not lie on $I \rightarrow X_i \rightarrow X_j$ (Woodward, 2003, 98).

⁴Machamer et al. (2000, 3) exemplify this view when they define a mechanism as “[...] composed of both entities (with their properties) and activities. The organization of these entities and activities determines the ways in which they produce the phenomenon”.

Hauser’s methodological points were appropriate, he essentially replied by claiming the right to explore alternative research avenues—“to go beyond the statistical relationships to explore the generative mechanisms responsible for them” (*ibid.*, 1187). Boudon’s alternative consisted in designing “ideal-typical models” detailing how the aggregate patterns of interest—which are only summarized, but not explained, by statistical estimates (*ibid.*, 1176, 1178-9, 1183)—can emerge from the dynamic and likely nonlinear relation among the actors’ choices and their reactions to other actors’ choices, as well as structural constraints (*ibid.*, 1180, 1185-6). Interestingly, as made more evident by a later article (1979), Boudon regarded numerical simulations as a necessary tool for this alternative mechanism-based research strategy, although the type of simulations he employed were not, technically speaking, ABMs.

More recent sociological scholarship shows that the bone of contention still is the opposition between the vertical and horizontal view of mechanisms that was behind the Hauser-Boudon debate. In one of the first meta-theoretical discussions on how the concept of mechanism may re-orientate empirical research in sociology, Pawson (1989, 130-1) noted that, although a mechanistic representation may have the cognitive function of making a connection between quantitative variables intelligible, it should not be conceptually equated with, nor methodologically operationalized as, a statistical control and/or a set of intervening variables. This view animated the well-known volume on social mechanisms edited by Hedström and Swedberg (1998), which, as we recalled in this paper’s Introduction, launched a new wave of discussions on mechanism-based explanations in sociology. As correctly noted by Mahoney, this new wave was explicitly motivated by the ambition to go beyond correlation analysis and by the rejection of the view that a mechanism can be simply understood “as an *intervening variable* or set of *intervening variables* that explain why a correlation exists between an independent and dependent variable” (Mahoney, 2001, 578; emphasis added). Hedström and Swedberg (1998) went indeed back to the Hauser-Boudon debate, attacked the path-analytic tradition in sociology, and ultimately subscribed to the claim that “sociologists in the multivariate modeling tradition still make only rhetorical use of the language of mechanisms” (*ibid.*, 17). Hedström (2005, ch. 5)’s more recent writings on analytical sociology endorse a similar view. This is confirmed by the recent review on causal mechanisms in the social sciences by Hedström and Ylikoski (2010), who whilst taking pain to stress that a variety of definitions exists (*ibid.*, 51), as a matter of fact downplay the view that mechanisms consist of networks of intervening variables (*ibid.*, 51-2). This is clear when they explain why Woodward (2002)’s counterfactual account of mechanisms—a typical example of the horizontal view—is insufficient: “[a] mechanism tells us why the counterfactual dependency [between cause and effect] holds and ties the *relata* of the counterfactual to the knowledge about entities and relations underlying it” (*ibid.*, 54), which shows their endorsement of the vertical view. Along similar lines, Kalter and Kroneberg (2014) note that, in much empirical research, “mechanisms as *intervening variables*” are “*mistakenly* seen to ‘explain’ the presumed causal effect of an independent variable on a dependent one” (*ibid.*, 101; emphasis added).

However, in order to show how the appraisal of ABM for warranting causal inference is sensitive to different intuitions on what a mechanism is, it is important to accurately account for the horizontal intuition, too. In fact, as Morgan and Winship remind us, “[f]or decades, social scientists have considered the explication of mechanisms through

the introduction of intervening and mediating variables to be essential to sound explanatory practice in causal analysis” (2014, 330). The counterfactual approach to causality—a typical case of dependence account of causality, in last section’s terminology—is now reshaping this methodological tradition. As a consequence, the concept of mechanism as a network of intervening variables is also being reshaped in these terms. Knight and Winship (2013) are particularly clear on this point. They regard definitory attempts within the vertical perspective as “unsatisfactorily vague” (*ibid.*, 278) and propose a definition that, in their view, better clarifies in what sense a mechanism has a structure and is causal. According to this view, mechanisms are “modular sets of entities connected by relations of counterfactual dependence” (*ibid.*, 283).⁵ In sum, Knight and Winship’s definition amounts to viewing a mechanism as “[...] a causal relationship involving one or more *intervening variables* between a treatment and an outcome” (*ibid.*, 282; emphasis added). They ultimately propose directed acyclic graphs (DAGs) as a framework for discussing under what conditions a net of “mechanistic variables” (Morgan and Winship, 2014, 335) allows one to identify causal effects.

From the horizontal viewpoint, dissatisfaction with the vertical view of mechanisms is not only conceptual but also methodological. Two slightly different, albeit related, methodological objections may be found in the literature. The first objection, in a nutshell, is that it is unclear what it means to empirically evaluate alternative hypotheses on mechanisms when mechanisms are regarded as dynamic complex systems. Morgan and Winship (2014, 345) are explicit on this point when they object that the mechanism movement—they refer here to Goldthorpe (2001)’s and Hedström (2005)’s proposals—runs the risk of falling prey of a “mechanism anarchy”, that is, a proliferation of mechanistic models, with no clear-cut proofs of their empirical significance, or, alternatively, a “mechanism warlordism”, that is, of a proliferation of mechanistic models mainly supported by the scientific reputation of their proposers. As a remedy, they suggest a division of labor according to which the “generative mechanism movement”, in their own words, contributes to causal inference by developing “how-possible” and “how-plausible” models, while “causal analysis”, meaning quantitative techniques for observational data from within a potential outcome approach, provides the tools for assessing the claims implied by models, which have the pretension to describe actual mechanisms (*ibid.*, 346-7).

Morgan and Winship’s proposal is motivated by a second objection, which those who regard mechanisms as chains of variables, raise against those who regard them as complex dynamic systems. ABMs are, so the objection goes, not reliable methods for providing evidence for the existence of the postulated mechanisms. Morgan (2005) formulates this objection explicitly when, in relation of Hedström and Swedberg (1998)’s early volume, he claims that “Sorensen and others got it only partly right. Without a doubt, they correctly identified a major problem with quantitatively oriented sociology. But, they did not offer a sufficiently complete remedy” (*ibid.*, 26). For this reason, they do not take seriously the proposal of using simulation methods, in particular ABMs, as a tool for studying mechanisms in the vertical sense—in sociology (Hedström, 2005, ch. 6), economics (Epstein, 2006, chs. 1 and 2) or political science (Cederman, 2005).

⁵The modularity requirement refers to the fact that the component parts of the system are independently manipulable in the sense of Woodward (2002) (see fn. 3 above).

To those who view mechanisms as chains of intervening variables, simulation seems only a tool for “theory construction”, and, even to this task, a tool of limited utility because of its alleged lack of transparency (Morgan and Winship, 2014, 341, fn. 15).

It appears now clear that the different intuitions on both causality and mechanisms connect in a systematic way. If one has a dependence intuition about causality, one will tend to see mechanisms as—horizontal—chains of intervening variables. In contrast, if one has a production intuition about causality, one will tend to understand mechanisms as complex systems of interacting lower-level units that—vertically—trigger higher-level outcomes. The two understandings of mechanisms correspond to different ways to open a “black box” underpinning a cause-effect connection. From the horizontal perspective, one opens a black box by uncovering intermediate variables between a treatment and an outcome. As shown by the philosopher Peter Menzies (2012), the horizontal view is typical of the literature on structural equation models and causal graphs (cf. Pearl, 2009). As we have just seen, from the horizontal perspective, ABMs would seem unnecessary to study social mechanisms and establish causation on mechanistic grounds. In contrast, from the vertical perspective, one opens a black box by breaking the system down into parts and showing that the dynamic of the interactions among them can generate the aggregate behavior under scrutiny. From the vertical perspective, simulation methods, and ABMs in particular, would thus seem powerful tools for studying the details of these complex dynamics.

We suspect that part of the skepticism against using ABMs for the study of social mechanisms lies in that it is still unclear in what sense this simulation technique represents a mechanism differently from a statistical model. For Morgan (2005, 31), for instance, “[t]he appeal for mechanisms is a useful rallying cry, but the originality of a mechanism-based sociology has been oversold. [...] Arguing that mechanisms are concatenations of nonlinear functions is not an argument against the use of variables, since the primitive elements of functions – defined as inputs and outputs – can be redefined as variables”. This statement deserves special attention because it could be used to argue that, since a theoretical representation of a mechanism requires variables and functions (something we entirely agree with), a structured set of intervening/mediating variables can be considered a “mechanism sketch” (cf. Morgan and Winship, 2014, 346-52), and multivariate statistics a tool for directly testing mechanism-based explanations (on this point, see also Opp, 2007, 121). From the vertical view of mechanism, however, this implication would be incorrect because it fails to appreciate that the role performed by the (numerical and logical) variables and the functions relating and operating on these variables within a formal model of a mechanism is different from that of variables within a statistical model. In the latter, variables and functions are used to detect a pattern of *average* effects which may reflect the *aggregate* statistical signature of the postulated underlying mechanism. In the former, in contrast, variables and functions are used to represent the entities’ *micro-level* properties, activities and relations, such that the postulated mechanism triggers dynamics that bring about the aggregate connection under scrutiny. In the next section, we shall defend this “granularity” argument on a more technical basis.

1.3 ABMs

In the [Introduction](#), we noted that the vertical account of mechanisms motivated the diffusion of ABMs in several disciplines. What features, if any, of the technique’s deep infrastructure can justify such association? Here we address this question by bringing to light the technique’s intrinsic potentialities. The practical difficulties one must handle for realizing such potentialities will be discussed later on (§3).

At bottom, an ABM is a computer program designed to formally represent a set of hypotheses and executed to deduce, in a numerical form, the logical implications of such hypotheses. The computer program is of a particular kind, however. An ABM (in its purest form) is made up of “objects”, which from a computer science viewpoint 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). That is why [de Marchi and Page \(2014, 1\)](#) define ABMs as consisting of “autonomous, interacting computational objects, called agents, often situated in space and time” and ([Macy and Flache, 2009](#), 248) note that an ABM “replaces a single integrated model of the population with a population of models, each corresponding to an autonomous decision maker”. Any single object can indeed be seen as a computer program (see [Wooldridge, 2009](#), 5). As a consequence, within object-oriented programming, the modeling process amounts to decomposing the explanans into a set of “classes” of objects, namely groups of objects that share the same properties and functions, and arranging them in such a way that the behavior of objects in one class constitutes the input for the behavior of objects in another class. Studying the explanans, that is, simulating its computational model, means updating the attributes attached to the objects that make up the ABM, iterating the rules that define the objects, and letting the objects communicate and thus influence each other over (the simulated) time.

Thus, when ABM is viewed in terms of its fundamental computational components, namely objects, the affinity with the vertical view of mechanisms becomes manifest. Like social (or biological) mechanisms, which are made of entities (at several levels of organization) with their properties and activities, and are mobilized when these entities act and communicate with each others, ABMs are made of objects with their attributes and procedures (or methods, or functions) and are turned into dynamic processes when the objects are invoked and asked to execute the procedures attached to them. In sum, ABM inherits from its object-oriented basis an internal structure that is homologous to the structure and the functioning of what one wants to study within a vertical view of mechanisms.⁶

The detour through the deep, object-oriented infrastructure of ABM is not a purely technical digression. It also helps to better see the source of ABM’s flexibility for designing models of mechanisms that aim to directly represent aspects of social life that

⁶The deep connection between object-oriented languages and ABM lead some author to use the label “agent-based object modeling” ([Miller and Page, 2007](#), 78) or “object-oriented simulation methodology” (see [Hummon and Fararo 1995a](#), 8; see also [Hummon and Fararo 1995b](#)) instead of that of ABM itout court. To be sure, it can be argued that, in practice and in theory (see, respectively, [Nikolai and Madey 2009](#), and [Izquierdo et al. 2009](#)), any language containing minimal requirements can be used to program an ABM. However, it is widely agreed that it is practically impossible to build complex ABMs without using object-oriented programming (see, e.g., [Shalizi, 2006](#), §5), which proves the existence of an intimate link between this programming style and ABM’s flexibility and generality.

statistical methods, as well as other simulation-based modeling strategies, are not able to handle with similar easiness. In this respect, we shall draw the readers' attention to the following elements: *heterogeneity*; *microfoundations*; *interdependence structures*; *time*; and *multi-level settings*.

HETEROGENEITY Since ABM is about the design and manipulation of single computational entities, namely objects, averaging is never a required simplification and unrealistic modeling shortcuts, such as a representative agent, can be avoided (Gallegati and Kirman, 1999). The objects' heterogeneity can take several forms within an ABM (Epstein, 2006, xvi, 7). First, objects in the same class, whilst by definition sharing the same properties (and activities), are such that these properties can get different values. Second, objects in different classes by construction possess different types of behavior. Third, by playing with the objects' scheduling, objects can be represented as being heterogeneous in terms of behavioral sequences, that is, the time at which a given behavior is realized. Finally, the objects are conceptually "empty", meaning that by playing with variables, vectors or other data structures, the objects' states can model any attributes of the entities of interest, and by creating logical and numerical functions over these states, the objects' "methods" can be used to model every activities of these entities. In consequence, heterogeneity can take the form of multiple classes of objects representing different types of entities at different levels of abstractions, such as organizations and actors. As stressed by Miller and Page (2007, 84-5), homogeneity may be a convenient and theoretically legitimate assumption. The point is that ABM does not constrain us a priori to impose homogeneity because of tractability issues that are unrelated to substantive considerations. Within ABM, the right amount/type of heterogeneity becomes a modeling problem itself that can be directly addressed by studying the high-level consequences of this or that amount/type of heterogeneity.

MICROFOUNDATIONS Objects are defined by the attributes and functions one attaches to them. Similarly to attributes, the objects' functions can also be of all sorts. Since the model is solved by simulation, there is no a priori constraint on the type (logical or numerical) and form (continuous or discrete) assumed by these functions. This allows a great deal of flexibility in designing the entities' behaviors. When objects are used to model individuals, a large spectrum of options are available to represent the actors' reasoning and choices, for instance, simple heuristics (Miller and Page 2004, 10; Todd et al. 2005), heuristic-based game-theoretic strategies (Alexander 2007, 38-42; Gintis 2009, 72-3), sophisticated maximizing behaviors (Shoham and Leyton-Brown, 2009), complex cognitive reasoning (Wooldridge, 2000), or argument-based decisions (Gabbriellini and Torroni, 2014). Thus, contrary to the frequent association established between ABM and rational-choice theorizing (see, e.g., Elster, 2009a, §2), the tool is agnostic on the kind of micro-foundations a modeler should subscribe to. In fact, it can accommodate a large variety of cognitive mechanisms driving the actors' behavior (Miller and Page, 2007, 81-3) and formally support a deep critique of rational choice theory, insofar as ABMs show that, contrary to sophisticated rational behaviors, simple rules are enough to derive stable macro-equilibria on realistic time scales (Epstein 2006, ch. 1; Manzo and Baldassarri 2015).

INTERDEPENDENCY According to Wooldridge's aforementioned definition of objects, one of the features of computational objects is that they communicate with each other. This is the aspect that helps us see why ABM is so flexible in embedding entities' behaviors within local structures (Epstein, 2006, 6, 52). Since the attributes' values can travel from an object to another, it is easy to make one object's behavior depend on another object's behavior. This dependence can assume three general forms. First, object interdependence can be mediated by a global aggregate, namely an outcome derived from the behavior of all objects present in the artificial population that feeds back onto the subsequent behaviors of each object. Second, object interdependence can be mediated by a local aggregate, namely an outcome derived from the behavior of all objects to which the focal object is connected that feeds back onto the subsequent behaviors of the focal object. Third, object interdependence can be purely dyadic, that is, the relevant input for a given object comes from a single other object. In the latter two cases, the relevant object's neighborhood can be defined on a spatial and/or relational basis. By exploiting the information exchange at the deep level of computer's memory addresses (Hummon and Fararo, 1995a), ABM allows one to specify in a large variety of ways space dependencies (Miller and Page, 2004) and/or network dependencies (Rolfe, 2014). The important point here for sociological theorizing is that, by playing with the objects' attributes, functions, and communication, ABM allows one not only to design spatial and relational structures but also to design mechanisms, which clarify how such structures affect lower-level entities' components such as beliefs, opportunities, perceptions or desires.

TIME Since ABM is a simulation-based method, it is intrinsically dynamic (Miller and Page, 2007, 80-1, 83-4). When the first set of objects is invoked to execute the procedures defining their behavior, a chain of activities, reactions, and updating is triggered, such that the final higher-level outcome is generated step-by-step by a concatenated cascade of local upward aggregations and downward effects. Thus, when an ABM is simulated, the mechanisms defining it are transformed into the unknown process potentially contained in these mechanisms. The point that needs to be stressed here is that time itself can be modeled within an ABM. As shown by Axtell (2000)'s seminal article (but see also Miller and Page, 2004), the order in which objects are invoked and updated, as well as the order in which procedures are executed by a given object, are themselves dependent on modeling choices. This offers to sociologists the possibility to systematically explore the higher-level consequences of different hypotheses on action and interaction scheduling (for arguments on the necessity of taking time into account in sociology, see Abbott 2001, ch. 7, and Winship 2009).

MULTI-LEVEL SETTINGS ABM's capability of integrating different levels of analysis is both static and dynamic. From a static point of view, as testified by so-called "agent/role/group" architectures in computer science (Ferber et al., 2005), work on cancer growth in biology (Zhang et al., 2009; Wang et al., 2013), and research in computational organization theory (Carley, 2002; Harrington and Chang, 2005; Fioretti, 2013), the conceptual emptiness of the objects implies that entities as diverse as particles, molecules, cells, beliefs, actors, groups, organizations, and states can be modeled and co-exist within a single ABM, thereby allowing the co-habitation of several levels

of analysis. From a dynamic point of view, by exploiting the objects' communication, the simulation of an ABM allows one to establish dynamic relations between these levels. In this regard, it is important to appreciate that, within an ABM, three different types relations can be established. First, it is possible to create “lateral” connections, that is, relations that create interdependencies among objects representing entities at the same level of analysis independently from the activities of objects representing entities at lower levels of abstraction. Second, “downward” relations are possible, that is, relations that establish interdependencies between the behaviors of objects at a given level of analysis and those of objects at a lower level of analysis, or relationships involving local/global aggregates generated at time t that feed back onto the objects' behavior at time $t + 1$ (we saw this case while discussing interdependence structures). Third, “upward” relations can be triggered, that is, relations that create interdependencies between the behaviors of objects at a given level of analysis and those of objects at a higher level of analysis or objects that collect the behavior of lower-level objects to compute the resulting outcome at a higher level of abstraction. This last form of transition between levels is especially important. As explicitly noted by Coleman (1986, 1316), statistical techniques for observational data are traditionally good at assessing the effect of group- and individual-level factors on individual-level outcomes (and, today, we may add, the effect of network- and individual-level features on network-level outcomes), but they are not equally efficient at developing “methods for characterizing systemic action resulting from the interdependent actions of members of the system”. By iterating the objects' behavior, by making the objects communicate, and by collecting the local products of these behaviors over time, the simulation of an ABM is able to produce the macro level step-by-step. In this sense, Epstein (2006, 21) claims that “agent-based models allow us to study the micro-to-macro mapping”.

Thus, with respect to crucial elements of sociological analysis, the deep infrastructure of ABM allows a great deal of flexibility, granularity, and generality for the implementation of the vertical view of mechanisms.⁷ By *flexibility*, we mean that an ABM is not restricted to model any specific kind of entities, properties, activities, interdependence structure, level of analysis, sequence of activation or behavioral rule (Axtell, 2000). By *granularity*, we mean that an ABM does not restrict a priori the level of detail at which one can describe each of these elements. By *generality*, we mean that an ABM can include several formalisms, each of which can be used to model a specific aspect of the mechanisms under scrutiny—this feature of an ABM has been called “pluriformalization” (Varenne, 2009, 14).⁸

⁷Some statistical techniques, such as multilevel statistical models (Courgeau, 2003), or mathematical formalisms such as recursive Bayesian networks (Casini et al., 2011; Clarke et al., 2014), too, presuppose a vertical understanding of mechanisms in that they deal with connections among levels of analysis. The point here is that ABM is not only capable to describe the sequences of events responsible for the connections among the levels and to summarize it through a (set of) statistic coefficient(s) but also to recreate it, dynamically.

⁸Other computational techniques, too, for instance cellular automata, artificial neural networks, or genetic algorithms, are “bottom-up” and focus on the behavior of single entities (Gilbert and Troitzsch, 2005, chs. 7, 10); as such, they, too, may be used to operationalize the vertical view of mechanisms. However, by looking at some applications in computational biology (Zhang et al., 2009; Wang et al., 2013) and economics (Hayward, 2006), it seems fair to claim that ABM can incorporate all of the modeling

2 A VARIETY OF ABMS

In addition to the diversity of views, which scholars endorse on the concepts of causality and mechanisms, a third element contributes to explain the diversity of opinions on using ABMs for causal inference. This is the large variety of ABMs on offer. In particular, ABMs are very diverse with respect to the kind of phenomena they are supposed to explain, the kind of theoretical and empirical information that is used to build them, and the kind of operations that are performed to assess their validity. Thus, depending on the specific type of ABM considered, one can reach different conclusions on the potential of this method for establishing causal claims—even if one evaluates this method in its own terms, viz. a vertical view of mechanisms and a production view of causality.

In this section, we first reconstruct the historical roots of the ABM’s diversity with respect to the link the modeler aims to establish between model and reality (§2.1). Then, by considering applications from different disciplines, we document a slow trend in the literature towards more realistic ABMs (§2.2). Finally, we develop a typology of ABMs, which clarifies how different forms of realism may co-exist within a single model (§2.3). It is on this basis that in §3 we shall then proceed to a general discussion of potentialities and limitations of ABM for causal inference.

2.1 *Historical roots of the ABMs’ diversity*

ABM has been used in two different ways since the beginning of its history in the social sciences. In this respect, let us first consider Thomas Schelling (1971)’s acclaimed model of ethnic segregation. Schelling postulated an ideal uni-dimensional or—in the most famous model’s variant—bi-dimensional space in which “stars” and “zeros” asynchronously decide to change their location as a function of their closest neighbors’ features. Schelling’s goal was to see whether, starting from a random distribution of stars’ and zeros’ locations, (more or less) weak homophilous tastes were sufficient to generate clusters when location choices were repeated over time. To answer this question, Schelling varied several aspects of the model, like the intensity of preference for like-neighbors and the size of groups and neighborhoods, and studied how spatial patterns changed as a function of these modifications. For our discussion, what matters is that Schelling did not use on the input side any specific sociological or psychological theory to justify theoretically his micro-level assumptions; nor did he use any empirical data to set the stars’ and zeros’ ethnic preferences or their relevant neighbors; instead, he drew on a (weak) structural analogy, based on intuitions and common sense, between his fictional mechanism and the mechanism for segregation in reality. On the output side, simulated patterns were not confronted with empirical data on ethnic segregation in specific geographical area; as Schelling himself admitted, his analysis concerns any phenomena in which two groups have some tendency to stay apart from each other (*ibid.*, 144, 158).

In a far less-known research, the Swedish geographer Torsten Hägerstrand (1970) followed a radically different strategy. His ABM *ante litteram* was designed to account for patterns of temporal and spatial concentration of farm innovations in two Swedish

advantages of such techniques, and thus provide a more general and powerful tool.

regions. Hägerstrand's hypothesis was that adopters can contaminate potential adopters as an inverse function of the physical distance between them. To study this hypothesis, he designed an ideal bi-dimensional space in which "robots" (in his own words) meet and spread information as the inverse to the squared physical distance separating them. The crucial point here is that Hägerstrand distributed robots on the grid in a way that reproduces the distribution of farms in the Swedish regions of interest and set the matrix of dyadic contact probabilities on the basis of independent statistical sources concerning local telephone traffic and migration fluxes in the regions of interest. Downstream, he systematically compared the simulated temporal and spatial patterns of adoption with the Swedish actual data.

Thus, a simple mechanism lies at the heart of both Schelling's and Hägerstrand's agent-based simulations, viz. some distaste for dissimilar others and spatial proximity in interpersonal exchanges respectively. In both case, the mechanism was intuitively believed to be general and realistic. Ultimately, this belief relied on common sense. Given the same starting point, however, Schelling and Hägerstrand operated differently with the posited lower-level mechanism. Schelling's goal was to explore the space of its logical implications at the population level. The simulation was used to discover counter-intuitive consequences. The model's parameter and structure were partly modified to assess the robustness of the surprising outcomes discovered. After all, the realism of the model was of secondary importance. It was the model's heuristic value that really mattered to him. Hägerstrand's goal was different. He wanted to reproduce a specific portion of reality at a specific time and place. To increase the confidence in the posited lower-level mechanism, he anchored it to specific empirical values and used it to generate simulated patterns under realistic input constraints. Hägerstrand was not interested in exploring the space of the model's logically possible outcomes. Instead, he aimed to incrementally refine the model until the simulated spatial patterns of adoption acceptably matched the actual Swedish data.

Schelling and Hägerstrand implicitly outlined two orientations that still deeply inform contemporary studies using ABM. In their purest form, these orientations are illustrated by the so-called "KISS" and "KIDS" principles. Supporters of the *Keep It Simple, Stupid* (KISS) approach share [Axelrod \(1997\)](#)'s conviction that "if the goal [of ABM] is to enrich our understanding of fundamental processes that may appear in a variety of applications [...], then simplicity of the assumptions is important, and realistic representation of all the details of a particular setting is not" (*ibid.*, 5). Within this perspective, the model's target is highly abstract: at best, it consists of qualitative properties shared by a large set of phenomena (see [Boero and Squazzoni 2005](#)'s distinction between "theoretical abstractions" and "typifications"). ABMs are envisaged as "tools to think with" ([O'Sullivan and Perry, 2013](#), 14-5) or "intuition engines" ([de Marchi and Page, 2014](#)). On the other hand, supporters of the *Keep It Descriptive, Stupid* (KIDS) principle ([Edmonds and Moss, 2005](#)) believe that the ABM flexibility for mechanism design is so high that there is no reason for starting with the simplest model. Instead, one should start with a model that is as complex as the available evidence allows and only later come up with simplifications, if understanding of the model and new evidence justify them. Within this perspective, the goal is to design "high fidelity models" ([de Marchi and Page, 2014](#)) characterized by "high-dimension realism" with respect to both micro-level assumptions and higher-level targets of interest ([Bruch and Atwell, 2015](#)).

2.2 From just-so stories towards more realistic models

Extensive literature reviews show that abstract ABMs are by far more frequent than empirically-oriented ones (see, e.g., [Macy and Willer 2002](#); [Sawyer 2004](#); and, more recently, [Squazzoni 2012](#), chs. 2-3). Arguably, this imbalance is related to how ABM was framed by the first programmatic work aiming to introduce this method into the social sciences. In [Epstein and Axtell \(1996\)](#)’s foundational *Growing Artificial Societies*, for instance, ABMs are seen as “laboratories” in which as simple as possible micro-level rules are shown to be sufficient to generate a macro-level outcome of interest (*ibid.*, 4, 18, 20, 22, 177). With the exception of the *Sugarscape*’s variant in which agents are supposed to act as neoclassical rational consumers, however, Epstein and Axtell never use existing sociological or psychological theories to set up these micro-level rules. Similarly, it is only in the book’s conclusion that empirical data are considered as a possible source for the design of “physically realistic environmental model” (*ibid.*, 164). On the explananda side, although Axtell and Epstein put forward the principle of ‘qualitative similarity’ between the simulated outcomes and the macroscopic target, their targets are never precisely defined on the basis of quantitative data or historical cases. Hence, *Sugarscape*’s capacity to generate realistic income distributions, evolving network friendship, migration dynamics, or a “proto-story”, just to name a few collective outcomes the authors are interested in, is difficult to ascertain. Between their simulated phenomena and their real-world counterparts there is only a “phenomenological” analogy. The *Sugarscape* style still dominates many research areas, including cooperation ([Axelrod, 1997](#)), trust and reputation (for an overview, see [Pinyol and Sabater-Mir, 2013](#)), the emergence of norms ([Axtell et al., 2006](#)), and cultural and opinions dynamics (for an overview, see [Xia et al., 2011](#)). [Deffuant et al. \(2003\)](#)’s response to some critics of this orientation provides an especially clear illustration of the motivation animating these sub-fields. In particular, they overtly argue in favor of micro-level assumptions relying on “common sense psychological observations” and reject sociological and psychological knowledge because of its supposedly inconsistency and empirical fragility.

However, over the last ten years or so, a moving-away trend from common-sense based ABMs has appeared. Epstein’s follow-up book to *Growing Artificial Societies*, namely *Generative Social Science*, overtly devotes an entire section (2006, 12-6) to data-driven ABMs, which focus on clearly defined empirical collective phenomena, and provides a detailed account of research in archeology based on high-fidelity agent-based simulations (*ibid.*, chs. 4-6; for a critique, see [Grüne-Yanoff 2009a](#)). Modelers that used to praise the *Keep It Simple, Stupid* principle are also more and more advocating the use of calibrated and more firmly validated ABMs ([Boero and Squazzoni, 2005](#)). Recently, [Hassan et al. \(2010\)](#) explicitly argue for taking inspiration from more traditional micro-simulation models and using empirical distributions instead of arbitrary (typically, uniform) probability distributions to initialize the agents’ core variables.

Signs of dissatisfaction with abstract ABMs multiply even in research sub-fields in which they have traditionally dominated research practices. For instance, [Sobkowicz \(2009\)](#) reviewed a large set of opinion dynamics models and criticized socio-physicists for ignoring the existing sociological/psychological literature in the model-building stage and for virtually never confronting the models’ outcomes with clearly delimited macroscopic quantitative data (see also [Castellano et al. 2009](#) and [Chattoe-Brown 2014](#)). [Mäs](#)

and Flache (2013) are clearly sensitive to these objections and propose a new model of opinion dynamics whose micro-level assumptions on individuals and interactions are systematically based on empirical research in social psychology on attitude changes and memory processes. These assumptions as well as the opinion trends predicted by the model are then experimentally tested. ABMs in which experimental data are used to establish the agent’s behavioral rules are starting to complement abstract ABMs also in other research areas, like game-theoretic ABMs of cooperative behavior (Wunder et al., 2013), reputation dynamics (Boero et al., 2010), or diffusion of innovations (Cointet and Roth, 2007). As noted by Heckbert et al. (2010, 45-6), who reviewed ABMs at the intersection of ecology and economics, although “the calibration and the validation of models” remains one research frontiers for the field, it seems descriptively correct to conclude that “ABM has increasingly moved from exploratory models with ad hoc representations of underlying processes to face the rigor of empirical validation”. Even in areas such as economics, which have traditionally been more theory- and data-driven, and where ABMs were initially used only as toy models that made no or little use of theory and data, it is now sometimes maintained that ABMs may accomplish more than just-so stories, but that to this end empirical validation (Fagiolo et al., 2007) or more interaction with experimental research on economic decision making (Duffy, 2006) are necessary.

Within this general trend, a careful meta-analysis of 21 recent ABMs in different research areas (see Table) indicates the existence of three distinct sub-trends, which bear in different ways on ABM’s capacity to warrant causal inference, namely:

- (a) a trend towards “theoretical realism”—i.e., from common-sense models to models informed by sociological/psychological theories;
- (b) a trend towards “empirical calibration”—i.e., from input-data-free models to calibrated models;
- (c) a trend towards “empirical validation”—i.e., from models with abstract macroscopic targets to models with well-specified macroscopic targets.

Surprisingly, the three sub-trends have been relatively independent. The review below serves to illustrate their possible combinations, thereby showing how the lack of one or more between (a)-(c) may provide an additional reason why the use of ABMs for causal inference is often viewed with skepticism.

We should emphasize that our goal is not exhaustiveness. Our Table is framed in terms of the presence/absence of (a)-(c), although each dimension should obviously be considered as a continuum. Our point here simply is to document this variability. Thus, the summary in our Table is best seen as a map providing a benchmark to locate a given ABM within the complex space defined by the three dimensions. Indeed, our ultimate goal is to propose a categorization of ABMs that helps judge to what extent different ABMs can produce causally-relevant evidence.⁹

⁹Other categorizations of ABMs are available in the literature (see, e.g., Boero and Squazzoni 2005; Brenner and Werker 2007). Ours differs from existing ones in that it specifically focuses on the ABMs’ potential causal significance and aims to identify categories that are sufficiently fine-grain as to make it possible to compare ABM to concurrent methods for causal inference.

2.3 ABMs for causal inference—a road-map

A first case to be considered is the absence of theoretical realism, empirical calibration, and empirical validation (see Table, [category 1](#)). The aforementioned *Sugarspace* models ([Epstein and Axtell, 1996](#)) are a typical illustration of this category of models, which one may call “toy-models”. To a large extent, the recent history of ABM research is an attempt to take distance from this type of models.

One way to move away from toy-models is by achieving theoretical realism—without empirical calibration or validation (see Table, [category 2](#)). [Janssen and Jager \(2001\)](#), for instance, ask for more realistic ABMs of the diffusion of market products. They argue that, to achieve this goal, it is necessary to use specific psychological theories of preference changes. Consistently, they build on a variety of approaches in social psychology to design the agents’ behavioral rules. However, the theory is not used to specify the functional forms of these rules. All model’s parameters and agents’ variables in the end come from theoretical probability distributions. [Janssen and Jager \(2003\)](#) attempt to further increase the realism of the model with respect to its network components. Reference is made to the literature on complex networks, which describes what real networks look like, but only abstract topologies are then implemented. As to the target side, the plea for theoretical and empirical realism at the micro-level is not accompanied by a request for empirical validation at the aggregate level. To defend the empirical relevance of the simulated market dynamics, the authors rely only on phenomenological analogies between the simulated outcomes and highly abstract and loosely defined empirical dynamics. This strategy is well exemplified by sentences like “[s]uch market resembles the daily shopping of most people, and refers to products such as coffee, toothpaste and milk” ([Janssen and Jager, 2001](#), 760). [Delre et al. \(2010\)](#) continue Janssen and Jager’s original search for psychological and relational realism but the gap between this theoretical constrain and the concrete way the ABM is implemented and validated remains.

ABMs from transport economics illustrate another way to take distance from toy-models, which consists in maximizing empirical calibration—without any particular attention to specific sociological/psychological theories or to empirical validation (see Table, [category 3](#)). [Dugundji and Gulyás \(2008\)](#), for instance, study the individuals’ transportation mode in the municipality of Amsterdam by relying on a combination of econometric techniques and agent-based simulations. The ABM is entirely empirically calibrated in that the agents’ behavior is based on a series of (a-theoretic) discrete choice models estimated on survey data predicting the individual’s choice for the transportation mode as a function of individuals’ socio-demographic features (e.g., gender, income, age, education, and residential location) and influence variables reporting the proportion of people leaving in the same district with similar socio-economic profile choosing this or that transportation mode. While the statistical model provides a cross-sectional picture of the marginal effects of these variables, Dugundji and Gulyás employ the ABM to embed the statistical models into a dynamic framework so that they can explore the aggregate consequences of the statistically estimated behavior when actors are assumed to be exposed to changes in locally aggregate behaviors. Thus, we have here an ABM in which the agents’ behavior is descriptively accurate but no reference to existing micro-level sociological or psychological theories is made. Moreover, while the model is entirely calibrated on the input side, no empirical validation on the output

side is performed.

The same combination of theoretical agnosticism at the micro-level, full empirical lower-level calibration, and no systematic output empirical validation can be found in research at the intersection of computer science and behavioral economics. [Wunder et al. \(2013\)](#) study cooperation behavior through a variant of public good games in which the final reward is not computed on the entire population of players but only over the player's neighbors. After collecting data from a large-scale web-based experiment, they design a series of formal models predicting individual contributions and fit each of them to experimental data. Then, they use an ABM in which agents behave according to the formal model best fitting the experimental data in order to explore the agents' cooperative behavior under theoretical conditions that were not covered by the original experiment. Wunder et al. describe clearly the logic behind their analysis when they note "our approach preserves the 'ABM as thought experiment' tradition of ABM, but attempts to ground it in agent rules that are calibrated to real human behavior within at least some domain". They also honestly acknowledge that this form of empirical calibration gives priority to predictive accuracy over individual-level cognitive plausibility (based on theoretical knowledge).

ABMs in computational finance illustrates the last way of taking distance from toy-models, viz. by introducing some form of empirical validation—with no use of specific sociological/psychological theories or attention to empirical calibration (see Table, [category 4](#)). In these models, the system-level dynamics of interest typically are "stylized facts" such as the fat tails of the unconditional distributions of returns, and the volatility clustering and persistence of prices. To generate these statistical features, [Lux and Marchesi \(1999\)](#) model the stock market as a fluid undergoing phase transition, where traders (fundamentalists, pessimist chartists and optimist chartists) switch from one group to another depending on the comparison of the respective profits, an opinion index and the price trend. In the so-called Santa Fe artificial stock market, [Arthur et al. \(1997\)](#) endeavor to account for the stylized facts by an evolutionary mechanism, where traders learn from the observation of prices by modifying their trading strategies through random mutation and cross-over of their best performing trading rules. The general hypothesis that both models purport to test is whether the stylized facts depend on agents' behavioral heterogeneity, a feature that traditional rational-choice macroeconomic models do not postulate. In both cases the explanandum is very general, not a particular succession of events leading to a bubble and subsequent crash, but robust statistical features of the time series of prices, as observed in real stock markets over very long time intervals, including both "catastrophic" events and (seemingly) tranquil periods. The generality of the explanandum that the simulations are able to reproduce, together with the presence of behavioral heterogeneity, is taken to compensate the unrealisticness of the micro-level assumptions and the scarce attention to input-level calibration.

Foundational papers in computational demography illustrate a different way to take distance from toy models, namely a combination of theoretical realism and empirical validation, but no empirical calibration (see Table, [category 5](#)). For instance, [Todd et al. \(2005\)](#) want to explain the similarity of the age distributions at the first marriage across countries. To this end, they design an ABM of sequential mate search. Experimental research on cognitive heuristics, which, they argue, provides a more realistic portrait of the actors' choice under uncertainty than traditional rational-choice models, is used

to design the agents' behavioral rules. Similarly to the aforementioned work on market dynamics, however, this connection with the existing theoretical and experimental literature does not translate into an empirical calibration of the model's micro-level assumptions. In spite of this, and differently from the two previous sets of examples, Todd et al. assess the relative explicative power of each of their model variants by (qualitatively) comparing the simulated distributions with empirical distributions (namely, for Norway and Romania). The lack of empirical calibration of the model's theoretically motivated lower-level assumptions is acknowledged by [Billari et al. \(2007\)](#) who, despite this limitation, introduce an additional heuristic in the model, viz. the imitation of the neighbors' marriage choices. In order to apply this model to a specific phenomenon, viz. population growth, in a specific context, viz. UK, [Silverman et al. \(2013\)](#) inject detailed contextual demographic data into it but they do not calibrate the core micro-level assumptions. They programmatically defend this combination as a means to a more theoretically-oriented demography (for a similar effort in historical demography, see [Gonzales-Bailon and Murphy, 2013](#)).

In contrast, advanced forms of empirical calibration coupled with empirical validation can be found in ABMs of disease diffusion in epidemiology (see Table, [category 6](#)).¹⁰ [Ajelli et al. \(2011\)](#), for instance, simulate a pandemic at the scale of a single country (namely, Italy) through an ABM in which agents are assigned socio-demographic characteristics, geographical locations and movement probabilities on the basis of a variety of survey, administrative and census data. Spatial structures at the level of municipalities are also represented as well as physical locations of schools and workplaces. Agents are assumed to meet at random within households, schools and workplaces whereas dyadic interactions within the general population depend on geographical distances among actors. The model is used to predict the evolution of influence-like disease at several scales (namely, country, census areas, and municipalities) and within different groups of the population (for an empirical validation of the model's predictions against specific data, see [Ajelli et al., 2010](#)). [Frias-Martinez et al. \(2011\)](#) show how it is possible to achieve even more fine-grained empirical calibration by using cell phone records instead of census and survey data. In this way, they argue, interaction probabilities can be based on the real networks of contacts and mobility fluxes among actors. They calibrate and test the model on Mexican data concerning the H1N1 flu outbreak in 2009 in order to assess the impact of government preventive interventions. Thus, ABMs of disease diffusion prove able to combine contextual, spatial, individual- and interaction-level empirical calibration with direct forms of outcome validation. However, since diseases are supposed to flow from an individual to another without the intervention of any cognitive mechanisms, these models typically do not make any reference to sociological/psychological actor-level theory.¹¹

In sociology, a similar variety of combinations of theoretical realism, empirical cal-

¹⁰For a programmatic statement from a software engineering viewpoint, see [Parker and Epstein \(2011\)](#).

¹¹We note that, although logically possible, we did not find examples of ABMs falling in [category 7](#) (see Table), where specific sociological/psychological theories inform the model's micro-level assumptions (theoretical realism) and full empirical calibration is present but empirical validation is absent. We suspect that this is explainable by the fact that when one takes pain to build a very realistic model on the input side, this is because one intends to demonstrate how the model fully accounts for some phenomenon, which requires empirical validation as well.

ibration, and empirical validation can be found among ABMs that attempt to take distance from toy-models. [Manzo and Baldassarri \(2015\)](#), for instance, provides an example falling in [category 2](#) (see Table). They aim to explain inequality growth in status distributions, and to this end build an ABM of deference exchanges. Programmatically, they ask for more realistic micro-level specifications on individual behaviors and interactions. To fulfill this requirement, the agents' behavioral rules are designed to incorporate existing theoretical and experimental evidence in social psychology and sociology on imitation, homophily, and reciprocity. However, this quest for realism at the micro-level is not accompanied by a direct empirical calibration of the model's lower-level assumptions. Moreover, although specific sets of regularities are described at the outset as explananda, the simulation outcomes in terms of status inequality, thought quantified, are not directly confronted with macroscopic empirical patterns of status inequality.

[DiMaggio and Garip \(2011\)](#) provides an example falling in [category 3](#) (see Table). They want to explain the persistence of inequality in adoption rates of a new technology (namely, Internet) among actors with different educational background. To this end, they build an ABM in which the agents' choices depend on the economic costs of the new technology and the choices of the agent's social contacts. On the input side, the size of agent population as well as agents' attributes like income, education, race, and network size come from US survey data. On the output side, DiMaggio and Garip simulate the model by manipulating the internal composition of the agents' network in terms of homophily along income, race, and education lines. Thus, the empirical calibration of the agents' core attributes (though not of behavioral rules) co-exists with an exploratory strategy with respect to the model's macroscopic outcomes. [Fountain and Stovel \(2014\)](#) follow the same research strategy (in that only the distribution of the worker's skills and the firms' size is grounded in US empirical data) in their study of the role of network structure and referrals in job career instability. Full calibration of the agents' behavioral rules, but no macroscopic empirical validation, is provided by [Bruch and Mare \(2006, 690-4\)](#), who use US vignette-based survey data for estimating the function that better describe the actors' preferences for the racial mix characterizing the neighborhood in which they would wish to leave. Then, they inject the estimated function within the bi-dimensional version of Schelling's original segregation model and explore the levels of racial segregation that emerge. (A similar strategy is at the heart of [Bruch 2014, 1245-7.](#))

[Manzo \(2013\)](#) provides an example falling in [category 5](#) (see Table). He focuses on specific cross-sectional patterns of inter-generational educational mobility (in France) and builds an ABM combining actors' cost-benefit rational evaluations with network-based mimetic behaviors. Here, too, although the model's micro-specification systematically builds on existing theoretical and empirical literature on educational inequalities (theoretical realism), empirical calibration at the input levels is absent (except for the size of agent population and background education subgroups). The simulated patterns of educational inequalities, however, are systematically compared to the empirical ones, and the observed difference between the two series of data is quantified.

[Hedström \(2005, ch. 6\)](#) provides an example falling in [category 6](#) (see Table). He studies unemployment in the Stockholm metropolitan area among 20-24 year-old youngsters. In particular, he first estimates a statistical model predicting the likelihood of leaving unemployment depending on actor-level attributes and unemployment among peers, and then makes his virtual agents choose to leave unemployment according to the esti-

mated logistic equations. He employs the model with a “counterfactual purpose” (*ibid.*, 138) in the sense that he wants to assess the extent to which the evolution of the unemployment level in the population of interest changes as compared to the actual level (thus, empirical confrontation between simulated and empirical aggregate data is also present) when the value of the coefficient expressing the marginal effect of unemployment among someone’s peers is manipulated.

Finally, let us mention [Mäs and Flache \(2013\)](#)’s analysis of group bi-polarization, which provides a case study approaching the simultaneous co-existence of theoretical realism, empirical calibration, and empirical validation (see Table, [category 8](#)). Here, theoretical and empirical research in social psychology on attitude changes and memory is mobilized to justify the model’s micro-level assumptions; an experiment is carefully designed to test the core behavior postulated for the artificial agents; the ABM is partly modified with respect to its micro-level assumptions to make it consistent with the experiment design; finally, the simulated aggregate predictions are compared with the experimental results.

In sum, ABMs across many disciplines show a high degree of diversity. The examined literature shows that, contrary to a widespread conviction, models seeking to escape the *Keep It Simple, Stupid* principle do exist. As documented by recent and more extended reviews of the literature (see [Bianchi and Squazzoni, 2015](#), 299-300, table 2), these models are still a minority, but their frequency is increasing. Although limited, our own meta-analysis is meant to make a more specific point, viz. among the subset of models with the ambition of being more than toy-models, this quest for closer links to reality is, in fact, pursued through different combinations of “theoretical realism” (meaning: the systematic use of the existing theoretical and empirical literature available in sociology and psychology to design the ABM), empirical calibration (meaning: the direct injection of empirical/experimental information into the model’s lower-level infrastructure), and empirical validation (meaning: the use of empirical information to assess the realism of the model’s macroscopic consequences). We maintain that this observation is key to assess ABM’s usefulness for causal inference. Given the diversity we have documented, an indiscriminated claim against this method’s capacity to produce causally-relevant evidence appears more difficult to defend. Depending on how an ABM is built, studied, and validated, it will get more or less close to this goal. This is the claim we shall defend in the next section, on a more systematic basis.

category	examples	realism	calibration	validation
1	Epstein and Axtell (1996)	×	×	×
2	Janssen and Jager (2001, 2003); Delre et al. (2010); Manzo and Baldassarri (2015)	✓	×	×
3	Bruch and Mare (2006); DiMaggio and Garip (2011); Fountain and Stovel (2014); Dugundji and Gulyás (2008); Wunder et al. (2013)	×	✓	×
4	Arthur et al. (1997); Lux and Marchesi (1999)	×	×	✓
5	Todd et al. (2005); Billari et al. (2007); Gonzales-Bailon and Murphy (2013); Silverman et al. (2013) ; Manzo (2013)	✓	×	✓
6	Ajelli et al. (2010, 2011); Frias-Martinez et al. (2011); Hedström (2005)	×	✓	✓
7	—	✓	✓	×
8	Mäs and Flache (2013)	✓	✓	✓

Table. Meta-analysis of 21 ABMs from the literature. Each ABM is categorized depending on whether it meets certain standards of theoretical realism and empirical calibration on the input side, and empirical validation on the output side.

3 ABMS AND CAUSAL INFERENCE

We have now all the elements to provide a principled answer to our main question: Can ABM provide causally-relevant evidence? On the one hand, we have shown indeed that this question could not be properly answered without acknowledging the variety of accounts of causality and mechanisms on offer (§1). ABM squares with a production accounts of causality and a vertical view of mechanism, and thus should be judged in its own terms as to its capacity to provide causally-relevant evidence along these lines. On the other hand, we have argued that judgments on the value of ABM for causal inference should account for the large observed diversity among ABMs as regards the relations that the modeler establishes between the ABM under scrutiny and the target the model aims at explaining (§2). In this regard, the point is that, while every type of ABM (see the Table in §2.3) can detail the source of a given connection among (a set of) variables *within* the model, the strength of the evidence varies from case to case in the sense that different ABMs will reduce in different ways the uncertainty on whether the source of the observed connection as reproduced by the model does in fact capture the most likely and/or relevant social mechanism responsible for the connection *outside* the model, that is, in the real world. Building on these theoretical and methodological clarifications, we now explain what it takes for an ABM to be sufficient for causal inference in ideal cases (§3.1), what kind of practical limitations hamper the ideal use of ABMs for causal inference (§3.2), and how to circumvent such limitations (§3.3).

3.1 *The ideal case*

Let us first consider “toy” ABMs (category 1 of the Table), in which the model’s low-level infrastructure is not based on (a) pre-existing sociological and/or psychological theories and/or indirect use of empirical evidence (theoretical realism); data are not used to (b) initialize parameters’ values and/or behavioral rules (empirical calibration), or (c) quantify the model’s capacity to replicate its target (empirical validation). Typically, these models are so simple that it is (relatively) easy to fully inspect and understand their internal functioning. Toy models thus produce generative knowledge in the sense that they teach us what dynamic chain of events, or process, bring from the model’s micro-level assumptions to a given set of macro-level consequences. In some sense, this knowledge may be causally interpreted. Through a systematic manipulation of the model’s parameters (sensitivity analysis) and micro-level infrastructure (robustness analysis), it is indeed perfectly possible to establish (probabilistically) that a given parameter/aspect of the model is responsible for a change in macro-level simulated patterns, all other model’s parameter/aspect being held constant. Given the complete closed nature of the system, it is also possible to ascertain by which modifications of the simulated process this change in the macro-level results was obtained. However, this knowledge is purely internal to the numerical system instantiating the ABM under scrutiny. Since no specific theory or data supporting the theoretical assumptions are present on the input side, it is impossible to argue that the ABM’s low-level infrastructure mimics any aspect of social reality. Similarly, since the macro-level numerical consequences are not confronted with any specific empirical regularities, it is difficult to see what the ABM is actually replicating. Thus, the model does not allow one to claim that the mechanisms depicted by it

identify the mechanisms at work in the real world. As a consequence, one does not get any causal knowledge on any target micro-level mechanisms, or the dynamic process generated by it.

A step toward the production of more convincing causally-relevant evidence is accomplished by ABMs in which specific micro-level theories and/or indirect empirical/experimental evidence are used to design the model's micro-level infrastructure (category 2 of the Table). The reason is that the theory and indirect empirical evidence mobilized give a precise meaning to the model's components and parameters. In toy ABMs, what reason is there to believe that, say, an agent's variable called "opinion" represents an actor's opinion rather than another individual-level attribute? Similarly, what reason is there to believe that the functional form expressing the behavior of a consumer is (more) suited to represent a consumer rather than the behavior of a different type of individual or collective entities? ABMs are intrinsically dynamic numeric systems, thus the labels one puts on this or that agent's attribute or behavior is partly arbitrary. The presence of a specific theory or indirect empirical/experimental evidence reduces the arbitrariness by imposing a precise meaning on the numerical and logical symbols on which the ABM operates. This, in turn, makes the results more easily interpretable with respect to the target. Then, if it can be proven through robustness analysis that different functional forms by which the same theoretical elements may be represented do not alter the high-level consequences generated by the manipulation of this or that low-level component of the ABM, an observer should consider the generative knowledge produced by such an ABM as more causally-relevant than that generated by some toy ABM designed using pure common-sense intuitions.¹²

To this, it may obviously be objected that the gain from using specific theories as ABMs' inputs still has the limitation that theory's trustworthiness is itself contingent on specific audiences, historical contexts, and the empirical support it has received at the time it is used to input the ABM under scrutiny. Moreover, sociological and psychological theories are themselves open to interpretation and often portray actors in very different ways. In this sense, theory alone cannot help an ABM to exclude that different mechanisms postulated by competing theories are equally good at reproducing the connection under scrutiny.

To reduce this kind of uncertainty, a first step is the use of specific datasets precisely defining the target to be replicated ("empirical validation"). Several ABMs briefly presented in §2.3 systematically compare macro-level simulated data with well-defined and quantified cross-sectional patterns or historical time series, which allows them to assess how the regularities of interest follow the postulated micro-level specification and how manipulating a certain specification gets us away from the observed regularities (categories 4, 5, 6, and 8 of the Table). Indeed, in order to claim that a given low-level

¹²It is worth noting that a very similar argument was put forward by Willer and Walker (2007) with respect to lab experiments. In particular they distinguish "empirically driven" and "theory-driven" experiments, the former being designed on the basis of the method of difference in order to discover new correlations while the latter being designed on the basis of a specific theory in order to test this theory (*ibid.*, 12, chs. 3-4). Willer and Walker's main argument is that "theory-driven" experiments are not exposed to the limitation of generalizability typically attributed to lab experiments. The reason, they argue, is that "theory is the bridge that connects observations made in the controlled laboratory environments to the world outside the lab" (*ibid.*, 58; the argument is developed at length in ch. 6).

specification is sufficient to generate a given outcome and that a given intervention on a certain specification changes the probability of observing an outcome, the outcome must be clearly specified. If there is no well-defined and quantified macro-level pattern, the causal relevance of the micro-level foundation cannot simply be determined. To use a metaphor, it would be like performing a regression analysis without observing a dependent variable.

Although not explicitly referred to ABM, the following statement by [Knight and Winship \(2013\)](#) contains the typical objection that one may address to (possibly theory-driven) empirically-validated ABMs for causal analysis:

Standard practice in mechanism-based analysis typically involves observing an association between a possible cause and effect and the positing mechanisms that could potentially link them. Yet, if explanation involves the identification of causal mechanisms, this approach is insufficient. The possible correspondance between a mechanism and an observed association does not imply causality unless it can be demonstrated that the association could *only* be due to the hypothesized mechanism. (*ibid.*, 284)

To address this concern, “empirical calibration”, namely the direct use of empirical data that fix the models’ parameters and micro-level infrastructure, is extremely helpful. In the previous section, we commented on several ABMs in which agents’ attributes, functional forms expressing behavioral rules, relational and geographic locations were directly derived from survey, census or digital data (see categories 3, 6, 8 of the [Table](#); for additional examples, see, respectively, [Brown and Robinson 2006](#), and [Magliocca et al. 2014](#)). Using the best data available, the uncertainty surrounding the right amount/type of actors’ heterogeneity and/or the potential disagreement on the most realistic theories of action or of the correct representation of the actors’ local environments are fixed from the beginning. Input data thus help to exclude competing mechanisms (or to adjudicate among several specifications of the same mechanism). In addition, when such empirical calibration is in place, it seems correct to regard macro-level consequences, which follow an intervention on an empirically-calibrated component of the model’s micro-level specification, as indicators of counterfactual connections in the real world, because the virtual basis on which this intervention operates, viz. the agents’ attribute distributions, behaviors, and local context, is a replica of its real-world counterpart. Thus, in response to [Diez Roux \(2014, 101\)](#)’s critique of ABMs from a potential outcome perspective (see [Introduction](#)), we deny that in ABMs “everything is counterfactual”, or just speculative, as long as the models’ parameters and micro-level infrastructure are directly based on empirical information. When this condition is met, competitive mechanisms can be assessed and non-merely-virtual counterfactual claims can be established.

Thus, in principle, ABMs can produce strong causally-relevant evidence. The conditions under which this goal can be achieved are demanding, however. First, specific micro-level theories (and indirect empirical/experimental evidence) must be used to design the model’s low-level infrastructure (theoretical realism). Second, the model’s macro-level consequences must be systematically confronted with well-defined and quantified empirical patterns (empirical calibration). Third, empirical information must be directly injected into the model’s low-level infrastructure (empirical calibration). The crucial point is that an ABM with causal ambitions should satisfy theoretical realism, empirical validation, and empirical calibration *at the same time* ([category 8](#) of the [Table](#)). The first ingredient provides the model’s micro-level infrastructure with a specific

set of meanings so that the agents' attributes and behaviors are not pure verbal labels opened to a variety of interpretations. The second condition is necessary to prove that the postulated micro-specification brings about the macro-level explanandum: without a clearly-defined explanandum, no precise cause-effect connection can indeed be established. The last ingredient is necessary to eliminate competing theoretical accounts of a given macroscopic patterns (see [Epstein, 2006](#), 9-10) and to determine which among several parameter combinations leading to the same result is the putative cause. This problem is well illustrated by [Gonzales-Bailon and Murphy \(2013\)](#)'s model discussed in §2.3.

In a nutshell, the combination of theoretical realism, empirical validation, and empirical calibration allows ABMs to produce generative knowledge that is causally-relevant in that the relations established within the numerical realm of the model can be mapped, both in terms of inputs and outputs, onto their real-world counterparts. The ABM becomes an inferential device because it is a “mimicking” device. As stressed by [Mary Morgan \(2012\)](#) with reference to earlier simulation models in macro-economics, namely Orcutt's simulation of the business cycle, “[i]t 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 represent” (*ibid.*, 337).

3.2 Practical limitations

The conditions we have identified for an ABM to become a mimicking device, thus supporting causal inference on mechanistic grounds, are very strict. One may object that, in practice, these conditions can never be *fully* satisfied. The major obstacle seems to come from data availability. Indeed the flexibility of ABM for mechanism design comes at a cost: more granularity demands more fine-grained information to set up the model's micro-level infrastructure ([Ajelli et al., 2011](#)). As to the criterion of empirical calibration, the consequence is that this operation is likely to remain incomplete. In fact if one carefully looks into the examples of empirically-calibrated ABMs presented in §2.3, two facts are evident. On the one hand, some aspects of the models' micro-level specification are more frequently calibrated than others. In particular, population-level parameters (e.g., the size of the population and agent subgroups) as well as the agents' attributes always come from empirical data; the agents' behavioral rules are less often empirically calibrated and even less are interaction structures (on the difficulty of finding appropriate network data to be directly injected into ABMs, see [Rolfe, 2014](#)); agents' scheduling (what an agent does what and when, basically time) is always speculative. On the other hand, when behavioral rules are empirically calibrated (see, e.g., [Bruch and Mare 2006](#); [Dugundji and Gulyás 2008](#); [Hedström 2005](#)), they always follow functional forms (often regression-like) that are descriptively accurate but behaviorally unrealistic, which is simply the case because the calibration is based on survey data that do not allow fine-grained description of the mechanisms behind the actors' decisions (the same holds for experimental data used by [Wunder et al. 2013](#), who overtly admit to give priority to prediction over cognitive plausibility in the way they calibrate the agents' behavior). The obstacles to empirical calibration loom even larger when one take the intrinsic dynamic nature of ABM seriously. From this perspective, it appears that, unless rich longitudinal data are available, proper empirical-calibration is impossible. As noted by [Hansen and](#)

Heckman (1996, 100) with respect to simulation models more generally, it is indeed unclear what the value of using cross-sectional estimates only at the beginning of a simulated dynamic process really is.

Data availability also poses problems to empirical validation. Here the basic problem is to agree on what “replicating the observed patterns” means. When are simulated data and empirical data close enough? Obviously a variety of standard approaches exist to assess the fit between two datasets (for an overview, see O’Sullivan and Perry, 2013, 211-22). The problem is that it is still unclear what the best strategy is for agent-based simulations (see, e.g., Thorngate and Edmonds 2013; Thiele et al. 2014). This is crucial not only to ascertain the generative capacity of a given low-level specifications but also to assess the extent to which a given manipulation of a certain specification gets us away from the observed outcome, which is key to the counterfactual reasoning that causal knowledge is typically meant to support. One may even wonder whether population-level cross-sectional patterns and/or time series are the most relevant types of data to be used for assessing the causal relevance of the mechanistic evidence produced by an ABM. As noted by Macy and Flache (2009, 262), a mere input-output mapping does not tell us anything on whether the internal functioning of the ABM represents the real-world mechanism. After all, the generative capacity of an ABM ultimately resides in the set of micro-level processes, triggered by the execution of the computer program describing the postulated mechanism. Thus, should not a proper validation of an ABM go through a comparison between simulated and real-world micro-level processes? Obviously, if one endorses this view of model validation, according to which the target to replicate is the actual set of micro-level processes and not just their aggregate outcome, the limitations imposed by data availability becomes exorbitant.

As noted by Oreskes et al. (1994, 643) with respect to numerical simulations more generally, data limitations are such that confirmation “is always inherently partial”. As a consequence, even theoretically-guided, empirically-calibrated and validated ABMs, in which one attempts to empirically calibrate the model’s micro-level infrastructure and to confront the simulated outputs against actual data, may not completely convince an observer about ABM’s capacity to produce causally-relevant mechanistic evidence. For instance, in spite of their huge effort to achieve the best possible empirical calibration, Ajelli et al. (2011), comparing results from an epidemic meta-population model and an ABM, note: “It is however difficult to state which of the two predictions is the most accurate. On one hand the high level of realism of the ABM should make the prediction reliable. On the other hand this high realism is not free of modeling assumptions [...]”.¹³ Similarly, despite the presence of empirical information at several levels of analysis and systematic empirical validation, Gonzales-Bailon and Murphy (2013, 136) admit that “the simulations were not capable of providing causal explanations of fertility behaviour”.

Should one then conclude from these *in-practice* limitations—which, to reiterate, are not *in-principle* limitations—that ABMs are of *no* value for causal inference if they are theoretically-guided but only partially empirically-calibrated and validated? To answer this question, it is preliminarily imperative to understand whether full empirical calibra-

¹³The authors refer in particular to how they model movements among municipalities and the probability of getting infected through contact in the general population.

tion and validation are even appropriate requirements for an ABM and, if not, how one can compensate their lack with resources available *from within* the ABM methodology.

As to the first point (we will address the latter in the next subsection), let us remark that historically, indeed, ABMs have been naturally employed precisely to model aspects of social phenomena that existing data and techniques were unable to capture. At the beginning of his pioneering study, [Schelling \(1971, 147\)](#) noted that “the simple mathematics of ratios and mixtures tells us something about what outcomes are logically possible, but tells us little about the behavior that leads to, or that leads away from, particular outcomes”. Since then, it is a leitmotif that the strength of ABM is its capacity to deal with heterogeneity, complex interplay between behaviors and networks, and loops among several levels, all elements that empirical data (and statistical methods to describe them) are unable to grasp. This is the main reason why ABM is traveling across disciplines. (For a statement on this point in, for instance, epidemiology and social psychology see respectively [Auchincloss and Roux 2008](#) and [Smith and Conrey 2007](#).) In an influential article on numerical simulations more generally, [Oreskes et al. \(1994, 664\)](#) make this point very explicitly: “Fundamentally, the reason for modeling is a lack of full access, either in time or space, to the phenomena of interest”. Thus, is it not paradoxical to ask full calibration and validation from an ABM? What is indeed the point of demanding of a technique devised because of its power to model aspects of social mechanisms, which available data (and existing mathematical and statistical models) were not able to track, that it be constrained by those very same data?

Ultimately, in order to understand the value of ABMs for causal inference, it is crucial to appreciate that there is a fundamental trade-off for ABM to be genuinely useful, a trade-off between flexibly designing mechanisms that connect scarce input data to reality, and designing realistic mechanisms by drawing on exhaustive inputs. Since data are *de facto* limited, the more one wants empirically to constrain the simulation, the more one is obliged to reduce the granularity of the mechanism(s) that can be postulated. As noted by [Fagiolo et al. \(2007, 211-12\)](#), the quest for empirically calibrated and validated ABMs contains a potentially conservative stance: it can lead to adapt the type of mechanisms we design to available data, which would lead to under-exploiting the method’s potential for mechanism design. The capacity to produce causally-relevant evidence in virtue of full empirical calibration and validation is inversely correlated to the granularity of what the ABM can model to aid causal inference. The more weight is given to full empirical calibration and validation, the less ABMs can be used to gain insight about those phenomena for which data are missing. ABMs’ scope and specificity push in opposite directions. At the same time, we do appreciate that empirical calibration and validation are legitimate requirements when it comes to empirical research. Given this trade-off, is there any way to find, in presence of data limitation, a balance between the very motivation for using ABM, viz. the granularity and flexibility it allows for mechanism design, and the legitimate requirement of empirical calibration and validation for causal analysis? In the next subsection, we discuss methodological resources available *from within* ABM to achieve this goal.

3.3 Theoretical explorations

Within the context of a detailed analysis of spatial simulation models, O’Sullivan and Perry (2013, ch. 7) note that, although there still is a tendency to give confrontation with empirical data priority in assessing the value of a simulation model, “alternative to data confrontation based on model evaluation and statistical validation have come to the fore over recent years” (*ibid.*, 226). Given the data limitations we have just highlighted, we share this concern with diversifying the operations needed to prove that ABMs can produce causal knowledge. In particular, we interpret this diversification as a systematic combination of theoretical explorations and empirical calibration/validations to which the model should be submitted to gain credit as a trustworthy device for supporting causal claims. These operations play indeed different, but complementary, roles. By “theoretical explorations” of an ABM, we refer to the following four operations: (1) *sensitivity analysis*, namely the assessment of the variability of the (macro-level) simulated outcomes as a function of the model’s parameter values; (2) *robustness analysis*, namely the assessment of the variability of the simulated outcomes as a function of the model’s internal details, meaning the *formal* and *substantive* assumptions on which it is based;¹⁴ (3) *dispersion analysis*, namely the analysis of the simulated outcomes’ variability across simulation runs; and (4) *model analysis*, namely the analysis of the simulation’s internal functioning. Each of these operations should be seen as a complementary resource to empirical calibration and validation in that they contribute to reduce the uncertainty generated by data limitation. Let us explain more systematically in which sense this is so.

SENSITIVITY ANALYSIS Sensitivity analysis can be performed in different ways (for an overview, see Saltelli 2000; more specifically on ABMs, see Thiele et al. 2014 and Stonedahl and Wilensky 2010). “Global” sensitivity analysis, whose aim is to fully describe the model’s behavior over its entire parameter space, is especially important for two reasons. First, once the complete simulated input-output mapping is produced, it is easier to discover errors and/or logically inconsistencies in the model (the model’s original justification and the expected results being the benchmark to assess such inconsistencies). Thus, a systematic, preliminary theoretical exploration of the model prevents us from requiring empirical calibration/validation for models that are internally inconsistent (for technical or theoretical errors). Second, as noted by Helbing (2012, 42), since empirical/experimental estimates that can be used as a model’s input are themselves uncertain, knowledge of the model’s behavior within the model’s larger parameter space is important to put in perspective the empirical data used (on this point, for numerical simulations more generally, see also Oreskes et al., 1994, 641-2). Properly conducted, sensitivity analysis also helps counter the criticism that ABMs lead to results that are unreliable because based on specific numerical inputs—which is a common crit-

¹⁴By *formal* assumptions, we refer here to generic properties of the model such as the type of probability distributions postulated or the functional forms used to related some objects’ attributes. By *substantive* assumptions, we refer instead to assumptions more directly related to the objects’ behaviors and interactions, typically as derived from domain-specific theories and knowledge. The distinction is obviously a matter of degree. We shall go back to this distinction in §4.3, in relation to data-driven methods for causal inference.

icism to simulation-based methods more generally (see [Fararo and Kosaka 1976](#), 431-3, [Sørensen 1976](#), 85, 89, and more recently, [Gould 2002](#), 1169-70). Indeed, when global sensitivity analysis is used to design the model's response surface (see, in general, [Law 2007](#), 643-55, and for an example, [Fararo and Butts 1999](#), 51-2), as stressed by [Leombruni and Richiardi \(2005](#), 106) and [Epstein \(2006](#), 29-30), it is difficult to criticize ABM for not being able to generate fully specified and generalizable results. One counter-objection would be that, as the dimensionality of the parameter space increases, global sensitive analysis is difficult, even impossible. In this case, however, sampling-based sensitivity techniques can be applied (see [Saltelli et al., 2000](#), chs. 2 and 6). [Van de Rijdt et al. \(2009\)](#) and [Manzo and Baldassarri \(2015\)](#), for instance, stress that proper sensitivity analysis is still often not systematically performed in many ABMs in sociology and show, through specific case studies, how this tool can help increase one's confidence in the model's capacity to identify mechanisms that generate consequences in line with theoretical expectations and background empirical knowledge.

ROBUSTNESS ANALYSIS Robustness analysis aims to assess the extent to which the model's results depend on the model's internal structure ([Railsback and Grimm, 2011](#), 302-6). In particular, the following aspects are especially relevant: (a) the probability distributions used to input the simulation's core parameters; (b) the functional forms used to relate objects' attributes; (c) the spatial/relational structure through which objects interact; (d) the objects' invoking order and scheduling (on the last two aspects, see [Axtell 2001](#) and [Miller and Page 2004](#)). As noted above, empirical data are usually missing to fully calibrate all of these aspects, which leads critics to argue that ABMs' results are unreliable because too many internal details of the simulation contribute to produce them (see, e.g., [Grüne-Yanoff, 2009a](#), 547). In this respect, robustness analysis is a crucial resource in that it allows one to assess how the macro-level consequences of the ABM under scrutiny vary when the components of its low-level infrastructure for which there is no data receive different implementation. The more stable results are against these modifications, the larger our confidence should be in the ABM's capacity to produce causally-relevant mechanistic evidence. For this reason, some even claimed that robustness analysis can confirm hypotheses (see, e.g., [Weisberg, 2006](#)). Among the models presented in §2.3, [Bruch and Mare \(2006\)](#) perform an especially systematic robustness analysis showing how the intensity of spatial segregation generated by Schelling's model in fact crucially depend on the function one employs to represent agents' residential choices (cf. [Bruch and Mare, 2009](#)).

DISPERSION ANALYSIS Dispersion analysis concerns the quantification of the simulated outcomes' variability when the same simulation (meaning, with the same parameter values) is repeated several times. ABMs typically include several stochastic components. (For instance, the objects' attributes are often initialized from probability distributions; the objects' behavior is often probabilistic; the objects' invoking order is usually randomized.) For this reason, one simulation can be seen as a specific realization of an unknown stochastic process. As a consequence, each run leads to different results. As stressed by [Miller and Page \(2007](#), 74-5), critics consider this across-run variability as a lack of predictability in the results generated by an ABM. In fact, when global sensitivity analysis is combined with repeated model executions, a multi-dimensional

space can be produced, where for each combination of parameters' values a distribution of outcomes is generated. Each distribution is then described by appropriate dispersion measures and the overlap between distributions, hence the potential lack of significance between two manipulations, is evaluated (see [Helbing, 2012](#), 47). In this way, an ABM can generate predictions under the form of outcomes with a certain probability. This procedure is relevant for empirical calibration. For a model's components for which empirical data are not available, that changes in (a subset of) parameters do not lead to appreciable differences in the simulated outcomes increases our confidence in results' stability. As to empirical validation, when appropriate data are available, thinking in terms of distributions of simulated outcomes pushes us to put empirical data in a similar form, for instance by means of resampling techniques. This asks for an even more rigorous validation of the model's micro-level specification, insofar as not only a single observed patterns/trend must be replicated but also the observed (or estimated) variability around it. Among the models commented on in §2.3, the one by [Manzo \(2013\)](#) provides this type of dispersion analysis.

MODEL ANALYSIS Model analysis refers here to the set of strategies that can be used to understand (and describe) the set of events, behaviors and feedbacks—in other words, the process—triggered by the mechanisms coded in the computer program once this program is executed. That ABMs are black-boxes is indeed a common criticism among social scientists (see [Morgan and Winship, 2014](#), fn. 15, 341) and philosophers of science (see [Humphreys \[2009, 618-9\]](#)'s argument on ABMs' alleged “epistemic opacity”). But in fact, although the process may be time-consuming, an ABM is inspectable in practice as deeply as one wishes. Apart from sensitivity and robustness analyses, which themselves help develop intuitions on which pieces of the model are responsible for the outcomes (see, e.g., [Railsback and Grimm, 2011](#), 280, 282), mathematical techniques based on the theory of Markov processes are available precisely to describe the sequence of states through which an ABM evolves ([Young, 2006](#); [Izquierdo et al., 2009](#); [Gintis, 2013](#)). (Among the models presented in §2.3, [Bruch and Mare 2006](#) apply a variant of this approach to their re-analysis of Schelling's model.) The model's main mechanisms can also be introduced sequentially so that it is easier to assess their role in the model's dynamic and their relative weight in the determination of the outcome. (Among the models presented in §2.3, [Manzo \(2013\)](#) follow this strategy when analyzing his ABM of educational stratification.) Moreover, since an ABM can be simulated again and again, it is possible to collect data on it at various levels of analysis and specific measures can be computed to inspect this or that aspect of the model's dynamic ([Railsback and Grimm, 2011](#), 284). (Among the models presented in §2.3, [Manzo and Baldassarri 2015](#) perform this operation to illuminate the link between the model's dynamics at the agent- and system-level; for a similar strategy relying on multi-level regression applied to data collected during the simulation, see [Fountain and Stovel 2014](#).) Thus, it is always possible to shed light on the internal functioning of an ABM. This is important to assess the value of an ABM for causal inference: knowing how an ABM replicates a given (set of) macro-level observed regularities, rather than only knowing whether it is able to do so, provides us with more elements to evaluate the plausibility of the story the model tells us, thus its extrapolability to the real-world.

In the light of the existence of the above forms of theoretical exploration, we maintain that the ABM methodology has *internal* resources to cope with the practical problems related to empirical calibration and validation that we have highlighted in the previous section. Comparing the simulated outcomes to actual data (empirical validation) is without a doubt necessary to establish whether or not the model’s micro-level infrastructure is able to generate the observed outcomes of interest. Feeding empirical information within an ABM (empirical calibration) is without a doubt necessary to increase the realism of initial conditions and to reduce the probability that several micro-level specifications equally well replicate the patterns/trends of interest. Empirical data, however, are likely to be incomplete. In different ways, sensitivity, robustness, dispersion and model analysis help to assess the uncertainty surrounding data themselves as well as the model’s components for which no data are available at all. Thus, in the end we suggest that an ABM’s capacity to produce causally-relevant evidence relies not only on empirical calibration/validation but also on the model’s systematic theoretical exploration. Theoretical exploration and empirical validation/calibration of an ABM are complementary tasks: the strength of one set of operations help to counterbalance the limitations of the other. As observed by Muldoon (2007) with respect to simulation more generally, it is only the combination of these operations that can lead to the construction of “robust simulations”, i.e. “simulation that can teach us about the world” (*ibid.*, 883). In the next section, we show that, contrary to common views, experimental and statistical methods for causal analysis must confront similar *practical* challenges.

4 ABMS AND DATA-DRIVEN METHODS

In this section, we move from ABM to data-driven methods, that is, methods for causal inference that rely on information collected by letting individuals act in real systems. From within the point of view on causality (viz. a dependence, or difference-making view) and mechanisms (viz. a horizontal view) that animates these methods (see §1), we ask under which conditions they produce strong evidence of causality by their own lights, namely evidence that certain changes in an independent variable make a systematic difference to a dependent variable. The claim we want to make is that, similarly to ABM, data-driven methods have to face equally demanding challenges when employed to establish “horizontal” causal claims, that is, claims about robust probabilistic or counterfactual dependencies.

Several statisticians and sociologists have already formulated cautious statements about the causal interpretability of the empirical estimates produced by statistical methods for observational data. These warnings have concerned a variety of specific techniques among which path analysis (Freedman, 2005), structural equations models (Wang and Sobel, 2013), regression for categorical variables (Breen and Karlson, 2013), statistical models of network data (VanderWeele and An, 2013), and social interaction econometric models (Durlauf and Ioannides, 2010). As Gelman and Hill (2007, 167) observe with regard to regression methods, the common concern here is that “[c]ausal interpretations of regression coefficients can only be justified by relying on much stricter assumptions than are needed for predictive inference”. A sizable literature also exists on the practical conditions (like sample size) that are required to obtain reliable statistical estimates of causal effects (for a recent statement on the use and misuse of

hierarchical regression, see [Bryan and Jenkins 2015](#)). Thus, the unrealistic, and often untestable, character of the assumptions/conditions allowing causal inference by means of regression-like methods for observational data led some even to propose to limit these methods to descriptive purposes (for an especially radical statement, see [Berk 2010](#); [Berk et al. 2013](#)). However, these cautionary tales have been relatively inconsequential, in terms of both research practices and perceptions.

Concerning research practices, it is certainly possible that, as suggested by [Morgan and Winship \(2014, 13\)](#), the “naive usage of regression modelling”, also thanks to influential reflective critiques such as those of Andrew Abbott ([1988](#); [1997](#); [1998](#)), is coming to an end. At the same time, we suspect that this is only true of a small élite of researchers (in selected departments). Review papers in other discipline like management or psychology that attempt to assess how statistical methods for causal inference are applied in empirical research suggest that the overwhelming majority of papers do not satisfy, or simply ignore, the assumptions/conditions required for drawing correct causal inferences (see, e.g., [Antonakis et al. 2010](#)). Review papers for sociology exhibits similar patterns (see Bollen [[2012, 62-6](#)]’s analysis of how instrumental variables are applied in sociology). The very motivation for pushing the potential outcome approach in sociology testifies this state of affairs. As [Morgan and Winship \(2014, 7, 79, 121\)](#) see it: on the one hand, the counterfactual component of the potential outcome perspective can help frame causal questions more rigorously, by importing the experimental protocol within survey-based research; and on the other hand, the use of acyclic directed graphs can help regiment the way data are selected so as to correctly identify causal effects (cf. [Elwert 2013, 261-2](#)).

As regards perceptions of data-driven methods, consider once again the instructive excerpt by Diez Roux on the alleged difference between potential and simulated outcomes (see [Introduction](#)). That statement clearly illustrates the perceived intrinsic superiority of data-driven methods for causal inference. That they rely on “partial and incomplete (often messy) observations” is considered obvious but (hopefully) not too problematic; that they rely on assumptions that are unrealistic and often untestable is not even mentioned. On the other hand, ABM is seen as entirely relying on “prior knowledge or intuition”, which is taken to be “one of the most vexing problems” of the approach; that an ABM can be based on empirical information is not considered. Thus, a “fundamental distinction” is perceived to exist between data-driven methods, which allow one to establish causal connections in the real world, and simulation methods, which allegedly do not.

In the following three subsections, we shall argue that, in the light not only of current research practices but also of more fundamental methodological arguments, these perceptions are ill-grounded. Building on previous cautionary notes on data-driven methods, we add an explicit comparison with ABM, and develop three lines of reasoning of increasing generality. First, we argue that data-driven methods, too, have to deal with insufficiently fine-grained data that make some of their assumptions untestable (§4.1). Second, we explain how they, too, rely on assumptions that are, from a methodological point of view, *in principle* untestable (§4.2). Finally, as a consequence of the two first steps, we argue that both ABM and data-driven methods provide a reliable methodology for causal inference *only if* certain overarching principles do in fact hold (§4.3). To defend our argument, we focus in particular on randomized experiments, instrumental

variables, and causal graphs. This choice is meant to illustrate what challenges data-driven methods have to face both when confounding can be eliminated by design and when it can not, in which case other assumptions have to be introduced. Thus, albeit limited, this selection of methods allows us to tackle issues arising within the main identification strategies of causal effects (cf. [Morgan and Winship 2014](#), 30-3).

Let us finally note that the following discussion has no intention to dismiss data-driven methods. On the contrary, our ultimate goal is to motivate our plea for evidential variety (developed in §5), according to which neither ABM nor data-driven methods *alone* are sufficient to establish causal claims, because the kind of evidence they produce is, from different points of view, equally incomplete.

4.1 Data availability

As we have seen, data-driven methods are regarded as intrinsically superior for causal inference to ABMs on the ground that they rely on real data, and do not play around with unsupported assumptions and fictional parameters. Contrary to this view, we argue in this subsection that data-driven methods, too, require data that are in fact often missing, which prevents them from producing convincing evidence of causality. Below we describe a number of limitations involving data availability, as typically acknowledged by (some of) the (finest) proponents of data-driven methods themselves.

RANDOMIZED EXPERIMENTS Since randomized experiments are regarded as “the failsafe way to generate causal evidence” in many discipline ([Antonakis et al., 2010](#), 1086), let us first consider this method. The idea is simple: probabilistic relevance under intervention on a population of interest is strong evidence for causality. Suppose in a test population all causes of an *outcome* O (e.g., economic growth in a poor country) are held fixed, with the exception of the putative cause, a *treatment* T (e.g., financial aid from a richer country), and it is observed that $\Pr(O|T) > \Pr(O)$. Then, T is regarded as the cause of O in that population. Obviously, often we have no knowledge of all possible causes that might confound the result. Randomized experiments sidestep the problem. Units of analysis are randomly assigned to the treatment so that any possible systematic association between unobserved factors and the effect of interest is broken down (see [Gangl, 2010](#), 26). Thus, randomization allows one to treat a given population as a black box and estimate reliable mean responses in the presence of heterogeneity.

As penetratingly noted by [Cartwright \(2007a\)](#), however, randomized experiments are only apparently a self-sufficient causal machinery. Even when one is only interested in an average treatment effect, the generalizability of its statistical estimate to some target population ultimately rests on “auxiliary assumptions” that “are very demanding, demanding of information that is not supplied by the RCT and that is hard to come by” (*ibid.*, 19). Among these assumptions, those that concern the way the effect of the putative cause T on O varies across subgroups of the target population are crucially important (*ibid.*, 16-7).¹⁵ Not surprisingly, turn-arounds for tackling individual-level heterogeneity have been devised. For instance, the “set identification” approach, which

¹⁵To scrutinize directly the population’s causal structure, one should be in position to assess individual-level effects, but this is impossible because the treatment and its absence cannot be observed on the same individuals at the same time (more on this in §4.3).

Manski (2003) originally formulated with observational data in mind but lately extended to the counterfactual analysis of treatment responses within an experimental setting (see Manski 2007, chs. 7, 10 and Manski 2013, 63-76), renounces to point estimates of average effects and only identifies (non-parametric) regions in which these estimates may be reasonably found. This result can be reached by formulating weak assumptions on how different subgroups of a given population react to the treatments (for a clear description of these assumptions, see Morgan and Winship 2014, 425-7). At the same time, Manski admits that the “credibility” of these assumptions is a “subjective matter” (Manski, 2003, 1).¹⁶

Thus, the widespread view that randomized experiments provide the “gold standard” for causal inference, to be approximated in non-experimental contexts, too (Angrist and Pischke, 2010), essentially because randomized experiments establish conclusions in virtue of assumptions that do not require any domain-specific knowledge (Duflo et al., 2008), seems inaccurate. Similarly to ABM, causal inference in randomized experiments requires empirical information that is often missing, and ultimately requires assumptions that can only be justified by theoretical reasoning, substantive knowledge, or expert judgments.

INSTRUMENTAL VARIABLES The same problem reappears with data-driven methods that try to recreate experimental conditions in non-experimental settings. In observational studies, the way the units of analysis are assigned to the putative causal variable of interest (the treatment) is not controlled by the researcher, which implies that unconfoundedness cannot be claimed by construction (Imbens and Rubin 2015, ch. 3). When typical conditioning strategies, such as matching or regression (see Morgan and Winship 2014, chs. 5-7) cannot be used, because either the assignment mechanism is unknown or, if it is known, data are missing on relevant controlling covariates, one possibility is to rely on “instrumental variables” (for a detailed overview, see Bollen, 2012).

As claimed by Angrist and Krueger (2001, 73), “[t]he instrumental variables methods allow us 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”. Thus, in principle, by exploiting the existence of variables, whose properties resemble the properties of intervention variables, the technique promises to help draw causal inferences without relying on extensive background knowledge about the causal structure. More precisely, given a putative causal relation between T (e.g., aids) and O (e.g., growth), the instrumental variable approach purports to identify the causal effect of T on O , based on the existence of an instrumental variable I (e.g., an earthquake disrupting the aids as a function of its distance from its epicenter) that is (i) highly correlated with the putative cause T , and (ii) uncorrelated with the putative effect O given T . (i) is typically called “relevance” condition, whereas (ii) is referred to as “exogeneity” or “exclusion” restriction (see, respectively, Stock and Watson 2010, 333, and Gangl 2013, 381). To obtain reliable estimates, both conditions must be satisfied. The question we are interested in here is the extent to which the data under scrutiny allow the researcher to judge whether or not this is the case. We first consider the relevance condition. (We shall discuss the exclusion restriction in §4.2.)

¹⁶He claims: “An assumption is credible to the degree that someone thinks it so” (Manski, 2003, 48).

As noted by [Bollen \(2012, 59\)](#), the relevance condition has started to receive serious attention only later than the exclusion restriction. Although apparently innocuous, this is however a crucial assumption. In an important article, indeed, [Bound et al. \(1995\)](#), re-examine one of the seminal applications of instrumental variables in economics, namely ([Angrist and Krueger, 1991](#)), and show that, with a finite sample, instrumental variable estimates are “[...] biased in the direction of the expectation of the OLS estimator” (*ibid.*, 445). The point is that the size of this bias increases as a function of the strength of the correlation between the instrumental variable(s) and the endogenous variable(s). When the instrument is “weak”, that is, only marginally correlated with the putative cause(s) of interest, “even enormous samples do not eliminate the possibility of quantitatively important finite-sample biases” (*ibid.*, 446). This point would be later easily admitted by [Angrist and Krueger \(2001, 79\)](#) themselves. In addition, [Bound et al. \(1995, 444\)](#) show that, when an instrument is weak, small violations to the exclusion restriction are amplified, thus leading to even larger biases. Given how consequential a violation of the relevance condition can be, it is not surprising that a large literature has emerged on strategies to empirically assess the *strength* of an instrument (for an overview, see [Stock and Yogo 2002, §4](#) and [Stock and Yogo 2005](#)) as well as on methods that are robust against weak instruments, at least in large samples (for an overview, see [Stock and Yogo 2002, §5-§6](#)).

But how do we know if an instrument is *weak*? [Steiger and Stock \(1997\)](#) played important role in diffusing the following “rule-of-thumb”, which seems now ubiquitous in econometric textbooks: a first-stage partial-F test of less than 10 indicates the presence of weak instruments (cf. [Stock and Watson 2010, 350](#)). However, careful inspection of ([Stock and Yogo, 2002, 521-2](#)) suggests that weak instruments can be conceptualized in two different ways—one based on relative bias and the other on size—such that the statistical test leads to different threshold values, which in turn are a function of the number of instruments considered. Moreover, when [Stock and Yogo \(2005\)](#) discuss their rule of thumb in comparison with other possible test procedures, each providing its own rejection thresholds (*ibid.*, tables 5.1-5.4), they qualify their justification for the 10-value threshold as “not unreasonable”, and acknowledge that “[...] when the number of instruments is moderate or large, the critical value is much larger and the rule of thumb does not provide substantial assurance that the size distortion is controlled” (*ibid.*, 103).

To conclude, it seems fair to say that, although the relevance condition is in principle empirically testable, available procedures cannot have the last word. [Bollen \(2012, 59\)](#) notes that “diagnostics and tests for weak [instrumental variables] continue to evolve”; [Stock \(2001, 7581\)](#) remarks that “[...] no single preferred way to handle weak instruments has yet emerged [...]”; and [Bound et al. \(1995, 449\)](#) explicitly recommend to use them only as “rough guides”. In our view, the variety of existing procedures and threshold values indicates that data *alone* cannot provide a conclusive argument in favor of the strength of an instrument. In this respect, it is instructive to go back to [Angrist and Krueger \(1991\)](#)’s pioneering use of the quarter of birth as an instrument for estimating the causal effect of schooling on earnings. The empirical correlation between the instruments and the endogenous variables, viz. schooling, was very low (R^2 between 0.001 and 0.002). Actually, the justification for the instrument’s choice essentially came from the argument and the independent empirical evidence that the authors were able to offer to the reader. And it is arguably thanks to this reasoning that the article has become a

classic in the instrumental variable literature, even if the instrument was proven to be sub-optimal on a pure statistical ground (Bound et al., 1995).

Thus, similarly to ABM, causal inference in instrumental variable methods relies on assumptions whose justification cannot be *fully* derived from empirical data but requires independent theoretical reasoning, substantive knowledge, or expert judgments. As noted by Angrist and Krueger (2001, 73), “[...] good instruments often come from detailed knowledge of the economic mechanism and institutions determining the regressor of interest”.¹⁷

CAUSAL GRAPHS That causal inference based on data-driven methods, too, cannot exclusively rely on empirical data is made especially evident by the research in the causal graphs tradition, which aims for computerized causal discovery algorithms (Spirtes et al., 2000; Pearl, 2009; Korb and Nicholson, 2011). In general, a causal graph is a (non-parametric) mathematical object made of vertices and (missing) edges, where the former represent variables and the latter represent (the absence of) connections among these variables. When a directed edge from vertex *A* to vertex *B* exists, *A* is said to be a parent of *B*, and *B* is said to be a child (or descendant) of *A*. So-called “directed acyclic graphs” (DAGs) are a special class of graphs, such that all edges are directed, and there are no directed cyclic paths (*ibid.*, ch. 2). Minimally, two fundamental conditions must hold for the edges of a DAG to be causally interpretable (Spirtes et al. 2000, 11, 13; cf. Spirtes 2010, 1651, 1654): (i) each variable in the graph must be independent from its non-descendants given its parents (“Markov condition”)¹⁸ and (ii) all the (conditional) independencies among variables are implied by the Markov condition applied to the graph under study (“faithfulness”).

In the Nineties, a research program at the intersection of philosophy and computer science proposed causal graphs as a tool for automating the discovery of causal structures when background knowledge is scarce and/or uncertain.¹⁹ Despite the variety of algorithms proposed (for a detailed description, see Spirtes et al. 2000, ch. 5; for a software perspective overview, see Kalisch et al. 2012), the common logic of the approach is the following: given the distribution over observed variables, the algorithm iteratively deletes and orients edges as a result of statistical tests for conditional independence (constraint-based search algorithms) or as a result of changes in a given model selection

¹⁷Notice that Angrist and Krueger (2001, 76) explicitly reject “one of the most mechanical and naïve, yet common, approaches to the choice of instruments”, viz. one that “uses atheoretical and hard-to-assess assumptions about dynamic relationships to construct instruments from lagged variables in time series or panel data”.

¹⁸This is essentially is a statement of conditional independence, and a fancier version of Reichenbach (1956)’s principle of the common cause, which states that common causes (e.g., drops in pressure) screen off the spurious dependence among their effects (e.g., rain events and changes in barometer measure).

¹⁹Spirtes (2010) provides a clear statement of the motivation for this research program: “in new domains such as climate research (where satellite data now provide daily quantities of data unthinkable a few decades ago), fMRI brain imaging, and microarray measurements of gene expression, the number of variables can range into the tens of thousands, and there is often limited background knowledge to reduce the space of alternative causal hypotheses. [...] In such domains, non-automated causal discovery techniques from sample data, or sample data together with a limited number of experiments, appear to be hopeless, while the availability of computers with increased processing power and storage capacity allow for the practical implementation of computationally intensive automated search algorithms over large search spaces” (*ibid.*, 1648).

statistics, typically the Bayesian Information Criterion (score-based search algorithms). The search does not necessarily output a single graph. As noted by [Kalisch et al. \(2012, 2\)](#), “[...] in general one cannot estimate a unique DAG from observational data, not even with an infinite amount of data, since several DAGs can describe the same conditional independence information” (for different forms of graph indistinguishability, see [Spirtes et al. 2000](#), ch. 4).

The proponents of algorithmic search methods from observational data contend: “given our assumption, with an oracle that can correctly answer questions about conditional independencies and dependencies in a population, the outputs of our algorithms are correct” ([Spirtes and Glymour, 1997](#), 561). However, the methods have been harshly attacked on both conceptual and technical grounds (see, e.g., the exchange between [Freedman and Humphreys 1996, 1999](#) and [Spirtes and Glymour 1997](#)). Two main kinds of limitation have emerged: (i) *in practice*, data alone are not sufficient to generate reliable causal models; (ii) substantive, domain-specific, and expert judgments enter the search process at different stages.

As to the first point, [Freedman and Humphreys \(1996, 117\)](#) highlight that the core of causal discovery algorithms, that is, testing for conditional independence, is fragile when working with real data: “exact conditional independence cannot be determined from any finite sample. [...] correlations of 0.000 and 0.001 – at the population level – play very different roles in [Spirtes, Glymour, and Scheines’] theory. A sample of realistic size cannot distinguish between such correlations” (cf. [Freedman and Humphreys, 1999, 37](#)). As a consequence, they argue, causal discovery algorithms tend to output different causal models as a function of the significance level and assumptions behind the chosen statistical test ([Freedman and Humphreys, 1999, 42](#)). Proponents of computerized causal discovery algorithms recognize this limitation. [Spirtes and Glymour \(1997, 561\)](#) concede that “given the problem of sampling error, no algorithm whose output is a function of the sample could guarantee success” (cf. [Spirtes et al. 2000, 87-90, 296, 350-1](#), and [Spirtes 2010, 1656](#)). Interestingly, they suggest that simulation can be used to assess the robustness of the algorithms, and report on sensitivity tests that indicate that for samples larger than 2000 observations, when variables have no more than two or three parents, errors concerning presence/absence of connections among variables are rare whereas those concerning causal direction are more frequent ([Spirtes and Glymour, 1997, 562](#)).

As to the second point, namely the role of background knowledge, critics of causal discovery algorithms note that the method’s followers tend to underestimate the role of theoretical inputs (see [Freedman and Humphreys, 1999, 29, 41](#)), although they are in fact necessary to overcome data limitations and guarantee the method’s reliability: “Causal discovery algorithms succeed when they are prevented from making mistakes” (*ibid.*, 40). *De facto*, [Spirtes et al. \(2000, 93\)](#) allow the algorithms to incorporate prior knowledge about the existence or nonexistence of certain edges in the graph, the orientation of some of the edges, or the time order of the variables. As [Spirtes \(2010, 1654, 1655\)](#) overtly admits, background knowledge is ultimately necessary to select the most plausible causal graph among the set of equivalent graphs the algorithm generates.

In the end, although popularized as a fully inductive, data-driven approach to observational data, causal discovery algorithms, not unlike ABM, do in fact require independent theoretical reasoning, substantive knowledge, and expert judgments to single out

correct causal models.

4.2 Truth of the assumptions

As suggested by the quote by Diez Roux (see [Introduction](#)) we commented on in this section’s introduction, advocates of data-driven methods tend to believe that, differently from the assumptions used in ABM, those used in data-driven methods are always testable. Contrary to this view, in §4.1 we have argued that some crucial assumptions that are necessary to causally interpret the results produced by data-driven methods cannot in fact be verified empirically because of *practical* data limitation. Thus, as is the case with ABM, theoretical and/or empirical knowledge external to the method is ultimately required to make the horizontal evidence produced by the method conclusive. In this subsection, we want to make an even stronger point. In particular, we shall argue that some of the assumptions on which data-driven methods rely to establish causal claims are in fact *in principle* untestable. By this we mean that the only condition that would make some assumptions verifiable is a perfect and complete knowledge of the myriad of events affecting a given phenomenon. Given that this condition is obviously never met, the untestable character of the assumption does not derive here from the amount of data available but from a methodological impossibility. In other words, data will never tell us, by construction, if the assumption is tenable or not. Thus, we shall argue, irrespective of data availability, causal inference by data-driven methods will always have *leaps*, just as causal inference from ABMs, due to the unavoidable uncertainty as regards the truth of some of the assumptions on which the methods rely.

RANDOMIZED EXPERIMENTS Again, let us first consider randomized experiments. As mentioned, this method is considered the gold standard for causal inference because unit randomization is believed to solve the problem of confounding by design. In §4.1, with respect to the potential heterogeneity of causal effects within the population under scrutiny, building on [Cartwright \(2007a\)](#)’s critique we have highlighted that randomized experiments are in fact only apparently self-sufficient causal machinery. The reason is that in order to generalize the results to a target population, assumptions on the causal structure of the population under exam, which are difficult to test empirically, are required (for a similar statement in development economics, see [Deaton 2010](#)). Here we consider a more general and deeper assumption of the experimental approach, the so-called “stable unit treatment value assumption” (SUTVA) (for other labels, see [Morgan and Winship, 2014](#), 48), which, as noted by [Gangl \(2010, 38\)](#), is “much more restrictive and problematic than is commonly recognized in the discipline”.

According to [Imbens and Rubin \(2015, 10\)](#)’s recent formulation of an idea originally labeled and made explicit by [Rubin \(1980, 591\)](#), SUTVA requires that “the potential outcomes for any unit do not vary with the treatments assigned to other units, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes”. Imbens and Rubin insist on the fact that the assumption contains two components, the first clearly referring to potential interferences among units and the second concerning potentially unnoticed “hidden variations of treatments”. When experimenting with humans, both requirements can easily be violated. As to the former, violations typically obtain when direct or indirect social interactions are at work;

as to the latter, violations can arise from variations in the quality of the treatment, in the quality of those who administrate it, or in subtle behaviors endorsed by the units themselves (*ibid.*, 10-2).

SUTVA violations are consequential because, in the presence of interferences and/or hidden treatment heterogeneity, one may attribute to the treatment a causal impact that in fact arises from other unobserved and subtle processes. To make an example, in his critical analysis of studies assessing the causal impact of residential relocations within the “Movement to Opportunity” (MTO) in-the-field experimental program, [Sobel \(2006, 1401\)](#) notes that “[b]ecause the ITT [intent to treat] and TOT [treatment on the treated] are defined assuming no interference, and this assumption is not reasonable, it is not clear what parameters MTO researchers are estimating or what policies their analyses might support”. Sobel formally shows that, when social interactions are taken into account, the average treatment effect in fact amounts to the difference between the average effect of the treatment and “spillover effects on the untreated” (*ibid.*, 1403-5). According to him, this situation must be very common because “interference is the norm” (1399) in the social world (cf. [Gangl, 2010, 38](#)). That is why [Hong and Raudenbush \(2013, 337\)](#) claims that “[...] SUTVA, while arguably plausible in the classic clinical trial, appears highly implausible and often consequential in social settings”.

Our point is not that randomized experiments cannot be devised in such a way as to prevent foreseeable violations of SUTVA. On the contrary, a common strategy to achieve this goal is to design the experiment in such a way that the units of analysis are assigned to clusters within which the assumption of absence of social interactions and treatment homogeneity can be more easily defended (for an overview of such complex designs, see [Hong and Raudenbush 2013](#)). The point is rather that these designs, too, rely on assumptions that may be difficult to test empirically and, as noted by [Imbens and Rubin \(2015, 11\)](#), to some extent some “more distant” versions of SUTVA must be posited. Thus, on what grounds is the plausibility of this assumption ultimately established? Given the nature of the potential biasing processes at issue, fueled by diffuse social interdependencies and individual-level reactions that are difficult to trace, empirical data can hardly have the last word. There will always be some uncertainty about whether or not violations of SUTVA occur. That is why [Imbens and Rubin \(2015, 10\)](#) themselves claim that

[SUTVA exemplifies assumptions] that rely on external, substantive, information to rule out the existence of a causal effect of a particular treatment relative to an alternative. [...] these assumptions, and other restrictions discussed later, are not informed by observations – they are assumptions. That is, they rely on previously acquired knowledge of the subject matter for their justification. Causal inference is generally impossible without such assumptions, and thus it is critical to be explicit about their content and their justifications.

Thus, similarly to ABM, even the method that is regarded as the gold standard for causal inference, namely randomized trials, in fact relies on assumptions like SUTVA, which cannot be empirically tested and for which theoretical reasoning, substantive knowledge and expert judgments are necessary to reduce the uncertainty on the occurrence of violations.

INSTRUMENTAL VARIABLES Instrumental variables provide a further clear illustration of this challenge, which data-driven methods, too, have to face. As we have seen in

§4.1, this method promises to sidestep the problem of the practical impossibility to control for all potential confounders by exploiting a variation that is “external” to the system (of variables) under scrutiny. Causally interpretable estimates are conditional, however, on two assumptions, which are, to recall, the relevance condition (the instrument I must be highly correlated with the putative cause T) and the exogeneity, or exclusion, condition (I must be uncorrelated with the putative effect O given T). In §4.1, we have discussed the practical difficulties of empirically ascertaining whether or not the former assumption is met. Here we want to stress that the obstacles in verifying the latter are even more severe.

In fact, as noted by Deaton (2010, 431), the problem with this assumption is that “exogeneity is an identifying assumption that must be made prior to analysis of the data, empirical tests cannot settle the question”. Essentially, instrument exogeneity requires that all potential pathways going from I to O are controlled for. This is a condition that, by construction, cannot be verified relying on the data under scrutiny, and that, since it is always possible that a confounder is not taken into account, is *in principle* unverifiable. That is why Gangl (2013, 381), more overly than others, admits that “[. . .] it is important to realize that the exclusion restriction is an assumption that is not testable in principle”. Morgan and Winship (2014, 301-2) provide a clear explanation of why this is the case. They show that the still common strategy of conditioning on T and check whether I and O are independent cannot provide an empirical test of the exogeneity condition. The reason is that an association between I and O will exist not only when the instrument fails to be exogenous but also when the instrument is valid and yet T is a “collider”, that is, a variable correlated with both I and unobserved factors highly correlated with the putative effect (on the notion of collider, see, in sociology, Elwert and Winship 2014). As to models including several instruments, if their number is higher than those of putative causes included in the model, the so-called “over-identification” test can be performed (see, e.g. Stock and Watson 2010, 353-4). However, as noted by Bollen (2012, 56), this statistical test can only suggest that some instruments are correlated with the error term; it cannot clearly say which of them is violating the exogeneity condition.

It is important to note that the exogeneity conditions must also hold for instrumental variable estimators that are devised to account for potential treatment effect heterogeneity in the population under scrutiny, that is, the so-called LATE (“local average treatment effect”) estimators (for a clear introduction, see Morgan and Winship 2014, 305-15). In addition, proper estimation of LATE requires another assumptions that must be made prior to the data analysis, namely the “monotonicity” assumptions according to which the instrument is supposed to affect all individuals either positively (if they are “compliers”) or negatively (if they are “defiers”) but not both at the same time. The assumption is empirically untestable, however, because, as clarified by Gangl (2010, 37), LATE is based on changes in the expected exposure to treatment, not on actually observed changes of treatment status”. This leads some to stress again the role of theoretical reasoning to complement econometric assumptions so as to deal with heterogeneous causal effects (in Deaton [2010, 430]’s words: “heterogeneity is not a technical problem calling for an econometric solution but a reflection of the fact that we have not started on our proper business, which is trying to understand what is going on”; see also Angrist and Krueger 2001, 78).

To sum up, fundamental assumptions required by instrumental variables to generate

reliable, and causally interpretable, estimates cannot *as a matter of principle* be tested by recourse to data. As noted by [Stock and Watson \(2010, 353, emphasis in original\)](#), “assessing whether the instruments are exogenous *necessarily* requires making an expert judgment based on personal knowledge of the application”, or as remarked by [Deaton \(2010, 432\)](#), it requires “thinking about how and why things work”. This conclusion has also been reached by re-evaluating studies that used naturally occurring events as instruments (see, in particular, the statement on “plausible stories” by [Rosenzweig and Wolpin 2000, 830](#)). Causal inference that relies on instrumental variables has leaps, just as causal inference that relies on ABMs, due to the unavoidable uncertainty as regards the truth of some of the assumptions on which the method relies. Theoretical reasoning, substantive knowledge and/or expert judgments are necessary to reduce such uncertainty.

CAUSAL GRAPHS Causal graphs provide yet another clear example of the same problem. In §4.1, when examining methodological developments at the intersection of philosophy and computer science, we have illustrated the practical difficulties imposed by data scarcity to using causal graphs as a tool for inductively discovering causal relations from data by means of computerized algorithms. Here we want to focus on a more subtle and fundamental issue. Causal graphs, in particular directed acyclic graphs (DAGs), are spreading in sociology, too, in association with the counterfactual view of causality (for a clear introduction, see [Morgan and Winship 2014, ch. 3](#)). In sociology, however, more emphasis is put on using causal graphs as a deductive tool that allows one to clearly express sets of substantive and domain-specific hypotheses, and to establish explicitly under which conditions the associated causal connections can be identified (cf. [Elwert, 2013, 246-7](#)). Given their transparency, DAGs make more visible the presence of assumptions, which are necessary for causal inference and yet untestable *in principle*.

To see this, let us consider “the traditional way to deal with the potential problems raised by noncausal pathways”, namely “conditioning” ([Knight and Winship, 2013, 286](#)). As pedagogically clarified by [Morgan and Winship \(2014, 128-9\)](#), this identification strategy takes the form of “balancing” or “adjusting”, and is implemented in a large variety of statistical methods for causal identification from observational data, ranging from matching techniques to regression-like models (*ibid.*, chs. 5-7). Instead of adopting the defensive strategy of “controlling for everything”, DAGs support a principled definition of the minimal set of variables that is sufficient to make conditioning effective. Among these principles, the most fundamental is the so-called “backdoor criterion” ([Pearl, 1993, 268](#)), according to which a conditioning set Z is sufficient if (i) it blocks all “backdoor” paths from the treatment T to the outcome O (i.e., the non-causal paths that start with an arrow into T), since these introduce confounding, (ii) it does not contain any collider (i.e., an endogenous variable that has two or more causes) lying on backdoor paths, since conditioning on a collider create new conditional dependencies, and (iii) it contains no descendants of T , since these cancel out some portion of the causal effect of T on O (cf. [Elwert 2013, 259](#), and [Morgan and Winship 2014, 109-17, 130-9.](#))

But is the backdoor criterion assumption-free? When discussing the now classical study of the alleged causal effect of Catholic schooling on students’ learning ([Coleman et al., 1982; Hoffer et al., 1985; Coleman and Hoffer, 1987](#)), [Morgan and Winship \(2014,](#)

121) themselves admit the fundamental challenge raised by a conditioning identification strategy: “How do we determine all of the factors that systematically determine whether a student enrolls in a Catholic school instead of a public school? And can we obtain measures of all of these factors?” (cf. their discussion of regression specification, *ibid.*, 219-24). The application of the backdoor criterion requires indeed that the variables that are necessary to block the backdoor paths are observed. But, as noted by [Morgan and Winship \(2014, 122\)](#), “[...] in most cases [...] the directed graph will show clearly that only a subset of the variables in [schooling] that generate confounding are observed, and the confounding that they generate cannot be eliminated by conditioning with the observed data” (see also *ibid.*, 126, 130). Formulated in these terms, the difficulty with the backdoor criterion seems limited to a practical problem of data limitation. We contend, however, that the challenge is much deeper. Indeed the set(s) of variables this principle recommends to condition on in a particular application in order to achieve identification crucially depends on the variables included in the graph. The underlying, fundamental assumption is that all relevant variables are present. In other words, the causal graph must be right. In [Elwert \(2013, 249\)](#)’s words, “when working with DAGs, the analyst (for the most part) needs to assume that the DAG captures the causal structure of everything that matters about a process”. This obviously is a meta-assumption that cannot be tested empirically. Only external, theoretical, substantive, and domain-specific knowledge can help to reduce the uncertainty about the truth of the causal graph under exam.

Formally similar challenges touch other important graphical identification criteria. Among them, the “frontdoor criterion” ([Pearl, 1995, 676](#)) deserves a special attention because it promises to identify causal effects in settings where data are missing on known confounders, and the backdoor criterion cannot be applied (see [Knight and Winship, 2013, 287-8](#)). The idea behind the frontdoor criterion is to condition on a set of intervening variables Z between the treatment T and the outcome O such that (i) the chain of intervening variables captures all the directed paths from T to O , (ii) there is not unblocked paths connecting T to Z , and (iii) conditioning on T blocks all back-door paths connecting Z to O . The first condition is referred to as “exhaustiveness”; the second and the third define “isolatability” (see [Morgan and Winship, 2014, 333-4](#)). As we have seen earlier, according to scholars sharing the “horizontal” view of mechanisms, a chain of intervening variables that is exhaustive and isolated constitutes a “mechanism”. That is why they also speak of “identification by mechanisms” in relation to the frontdoor criterion (cf. [Knight and Winship, 2013](#)).

The question that should be raised at this point is on what ground the exhaustiveness and isolatability assumptions can be justified. The DAG under study cannot by itself say anything about the truth of these two assumptions because, as was the case with the backdoor criterion, the DAG must be *a priori* supposed to be valid for the frontdoor criterion to deliver any reliable conclusion (see [Elwert, 2013, 261](#)). Can empirical data external to the DAG under scrutiny help? In our view, the assumption that the chain of intervening variables describes all causal pathways from T to O is so demanding that no empirical data can be realistically imagined to support them. Similarly, given the complexity of social phenomena, it seems highly implausible that empirical data *alone* can exclude that some unobserved factors open backdoor paths to the set of intervening variables and/or to the outcome of interest. That is probably why [Morgan and Winship \(2014, 337\)](#) ultimately remark, with reference to a weak form of exhaustiveness, that “to

assert this assumption, one typically needs to have a very specific theoretical model of the complete identifying mechanism, even though part of it remain unobserved”, and, with reference to isolatability, Knight and Winship (2013, 293) note that “more generally, the point is that it is necessary to specify the mechanisms involved in a causal process in sufficient detail (or in sufficient depth) so that (some of) the mediating variables involved are isolatable”. But it seems fair to say that both operations are intrinsically theoretical. In the end, as noted by Elwert (2013, 270),

[...] one obvious challenge of working with DAGs is that the true causal DAG is often not known. This is a problem because identification always hinges on the validity of the causal model. If the DAG is incorrect, the identification conclusion drawn from it may be incorrect as well.

Thus, similarly to causal inference from ABMs, causal inference based on causal graphs has leaps: theoretical, substantive, and domain-specific knowledge is required to convincingly show that the assumptions on which the identification strategy relies are satisfied.

4.3 Reliability of the method

In §4.1 and §4.2, taking randomized experiments, instrumental variables, and causal graphs as exemplary data-driven methods for causal inference that illustrate typical identification strategies, we have argued that crucial assumptions that are necessary to causally interpret the results produced by these methods (i) either cannot be verified empirically because of *practical* data limitation, or (ii) are *in principle* untestable because their test would require a perfect and complete knowledge of the myriad of events affecting the phenomenon under exam. As a consequence, we have argued, background knowledge, which is external to the method the researcher is using, *must* be mobilized to complement what data *alone* can tell us. In this respect, data-driven methods and ABM are fundamentally similar. In both cases, causal inference has *leaps*, and theoretical, substantive, and domain-specific knowledge is necessary to make the inference credible.

In this last subsection, we aim to draw all the implications from this twofold argument. In particular, we explain why, in our view, there is no sound argument to the point that data-driven methods are ultimately more reliable methods for causal inference than ABM. For those who endorse the view we disagree with, the “fundamental difference” between the two approaches would arise from three different sources, namely (1) the “formal” nature of the assumptions on which data-driven methods rely; (2) the “materiality” of the systems studied by these methods; and (3) the validity test these methods can perform. Although often conflated (see, e.g., Diez Roux’s long quote we reported in the Introduction), these aspects are best treated separately. We shall tackle them in turn and argue that it is at best unjustified (if not false) to claim that as a matter of principle, data-driven methods are more reliable tools for causal inference than ABM, and more specifically that counterfactuals established in the framework of potential outcomes are more credible than counterfactuals established in the framework of simulated outcomes.

LACK OF A THEORY OF CAUSALITY Let us consider Glymour and Greenland (2008)’s emphasis on the role played by the Markov and faithfulness conditions, which as men-

tioned, are necessary for causal graphs to produce causally interpretable structures:

The rules and assumptions just discussed should be clearly distinguished from the content-specific causal assumptions encoded in a diagram, which relate to the substantive question at hand. These rules serve only to link the assumed causal structure (which is ideally based on sound and complete contextual information) to the associations that we observe. In this fashion, they allow testing of those assumptions and estimation of the effects implied by the graph. (*ibid.*, 191)

This remark is important because it puts very clearly the general point that data-driven methods are based on assumptions—such as SUTVA, exclusion restriction, Markov condition, etc.—that are thought as “formal” in the sense that they do not depend on context-specific knowledge. Rather, they belong to overarching formal theories for causal inference, which make it possible to link model to data, so as to interpret the latter in the light of the former.²⁰ In contrast, according to critics (as illustrated by the statement by Diez Roux we quoted in the [Introduction](#)), ABM would be entirely dependent on “substantive” assumptions, that is, theoretical, domain-specific hypotheses on the particular systems being modeled. Although ABMs may be internally valid models, the method itself lacks a theory by which to establish whether the models are externally valid, too. ABM must exploit knowledge that comes, and cannot but come, from outside the method. To assess how accurate this view really is, we should first consider what sort of justification, if any, the formal assumptions of data-driven methods require.

In order to reliably connect models and data, the assumptions of data-driven methods have to be true. Furthermore, since they aren’t substantive assumptions, which lend themselves to testing (by relying on formal assumptions), but formal assumptions, which by their very nature have to be assumed for testing to be possible, they have to be, in a sense, *a priori* true. But how justified is the view that they are? Our previous discussion on how data limitations affect data-driven methods suggests that it is by no means obvious that the formal assumptions behind data-driven methods can be easily regarded as universally valid.

For instance, to come back to one of the assumptions cited by [Glymour and Greenland \(2008\)](#), the faithfulness assumption is violated if positive and negative effects along different pathways cancel each other out (for a well-known example, see [Hesslow, 1976](#)). Defenders of the faithfulness assumption argue that any parametrization violating faithfulness has Lebesgue measure zero²¹ ([Spirtes et al., 2000](#), 66). However, whether this provides a conclusive argument in favor of the assumption has been questioned.²²

²⁰And clearly, advocates of randomized experiments and instrumental variables would claim they have their own theories, respectively the theory of potential outcomes ([Heckman, 2005](#)) and the theory of structural modeling ([Pearl, 2011](#)).

²¹A Lebesgue measure is a measure of subsets of n -dimensional Euclidean spaces, such as probability spaces over continuous random variables.

²²[Cartwright \(2007b, 68\)](#), for example, observes how the conclusion this argument is meant to support is that “it is unlikely that any causal system to which we consider applying our probabilistic methods will involve genuine causes that are not *prima facie* causes as well. But this conclusion would follow only if there were some plausible way to connect a Lebesgue measure over a space of ordered n -tuples of real numbers with the way in which parameters are chosen or arise naturally for the causal systems that we will be studying. I have never seen such a connection being proposed; that is I think because there is no plausible story to be told”.

An analogous point applies to the other assumption cited by [Glymour and Greenland \(2008\)](#), namely the Markov condition, which is known to fail in the physical realm, where non-deterministic common causes may fail to screen off their effects. For instance, consider the collision of two particles, such that the probability that two protons are emitted is not conditionally independent from the probability that two neutrons are emitted, even though they are (co-occurring) effects of a common cause, namely the collision. In general, it will be no use to assume that such phenomena are confined to the physical realm, precisely because it is unknown to what extent they percolate to the social world.²³

Notice that our argument is by no means intended to establish that these and other assumptions are never reasonable. Our point is rather that there is no *a priori* guarantee that they are necessarily true, or even approximately true, of the causal structures to which they are meant to apply. In this sense, the argument that the truth of ABM's substantive assumptions is context-dependent, whereas the truth of data-driven methods' formal assumptions is not, is not convincing.

Moreover, contrary to the claim that ABM is less reliable than data-driven methods because of the lack of formal assumptions, one may even insist that there are good reasons to reverse the qualitative hierarchy between formal and substantive assumptions. As forcefully argued by prominent statisticians (see, e.g., the leitmotif in [[Freedman 2009](#)] and [[Freedman 2010](#)]), formal assumptions, applicable across contexts, never grant by themselves the reliability of the inference from a data-driven method. This point is also echoed in philosophy by [Cartwright \(2007b, 68\)](#), who (in relation to the faithfulness assumption) stresses that “[i]t is not appropriate to offer the authority of formalism over serious consideration of what are the best assumptions to make about the structure at hand”. We take these remarks to show that even seemingly formal assumptions carry a substantive weight and require justifications that are often context- and domain-specific if they are to reliably connect models to data. Thus, it seems fair to conclude that the existence of formal assumptions is a red herring, and that substantive assumptions are as important to data-driven methods as they are to ABM.

Once this point is fully appreciated, it seems more difficult to make the more specific claim that counterfactuals on simulated outcomes are less realistic than counterfactuals on potential outcomes (for an explicit statement, see the quote by Diez Roux in the [Introduction](#)). Counterfactual claims are notoriously hard to establish because one is never in the position to observe at once how things actually are and how they could be. As a consequence, establishing counterfactual claims requires knowledge of quantities, be they about the response of some individual or about the response of an entire population, that are unobservable *in principle*. In the end, as reminded to us by [Morgan and Winship \(2014, 5\)](#), “individual causal effects cannot be observed or directly calculated at the individual level” (this is what Holland [[1986, 947](#)] labeled the “fundamental problem of causal inference”). Rather, they must be inferred from what happens to populations

²³As noticed again by [Cartwright \(2007b, 107\)](#), “viruses often produce all of a set of symptoms or none; we raise the interest rate to encourage savings, and a drop in the rate of consumption results as well; we offer incentives to single mothers to take jobs and the pressure on nursery places rises; and so forth”. In cases where multiple effects of a probabilistic common cause always co-occur, it will be impossible to recover the correct causal structure by relying on the Markov condition (for a more extensive discussion, see [Cartwright 1999](#); see also [Salmon 1984, 168-9](#)).

to which the individuals belong under *a priori* assumptions, like SUTVA (for a statement on the *a priori* nature of SUTVA, see [Rubin, 1986](#), 961). That is why [Morgan and Winship \(2014, 444\)](#) admit: “[...] we see no way to escape having to assert what-if assumptions about potential outcomes in order to move forward”. Data-driven methods and ABMs rely on different strategies to overcome the unobservability of actual and counterfactual responses but, in both cases, assumptions requiring substantive justification have to be made, only some of which can be empirically verified. There are no differences, based on the nature of the assumptions in ABM and data-driven methods, which make ABM in principle inferior to data-driven methods for establishing counterfactuals. In both cases, ultimately only a complex combination of data and theoretical reasoning can secure convincing causal claims.

LACK OF MATERIALITY At this point, an advocate of the potential outcome approach may object that there is still a fundamental difference between data-driven methods and ABM, which makes the former a more reliable tool for causal inference than the latter. The difference would depend on the “materiality” of the set-up studied by the former, as opposed to the abstract nature of the computer program investigated by the latter. This materiality would allow data-driven methods to generate *novel* causal knowledge, something ABM is structurally unable to do. This argument initially appeared in the literature comparing numerical simulations in physics or economics and lab experiments (for detailed discussions, see [Guala 2002](#), [Winsberg 2003](#), [Winsberg 2003](#), [Frigg and Reiss 2009](#), §5, and [Reiss 2011a](#)). Given that ABM is a special type of numerical simulation and the potential outcome approach aims at importing the experimental protocol within statistical methods for observational data, the materiality objection can easily be extended to the comparison between data-driven methods and ABM.

In particular, the materiality objection to using ABM as a reliable inferential tool would unfold as follows. In data-driven methods, the novelty comes from letting the real system work as a data-generating mechanism, which, under experimental or quasi-experimental conditions, allows one to generate evidence that narrows down the class of admissible hypotheses. In contrast, an ABM is not made of the same “stuff” as the target it is designed to study. As a consequence, it simply cannot generate the kind of novel evidence needed to eliminate uncertainties on the nature of the mechanism, no matter how much theory/data are fed into them. The evidence provided by an ABM can at most expose the (deductive) implications of available knowledge. No tweak in a fictional system can, by itself, provide evidence for any empirical hypothesis (for a similar point about models more generally, see [Hausman, 1992](#); [Grüne-Yanoff, 2009b](#)). Thus, no matter the limitations of data-driven methods, the behavior of an ABM simply isn’t as relevant to its target as the behavior of a sample is to the behavior of the target population. To be sure, it is not uncommon to find analogous claims even among agent-based modelers. For instance, [Gilbert and Troitzsch \(2005, 13\)](#) claim that, although “[s]imulation is akin to experimental methodology”, the two are not the same in that “while in an experiment one is controlling the actual object of interest [...] in a simulation one is experimenting with a model rather than the phenomenon of interest”.

A number of philosophical papers have recently challenged the materiality objection arguing that materiality *is not enough* to establish the superiority of data-driven methods, especially experiments, over computer simulations (see, in particular, [Barberousse](#)

et al. 2009, Parker 2009; Morrison 2015, ch. 6). An important line of argument in this literature is that what really matters for allowing credible inferences is “whether the experimental and target systems were actually similar in the ways that are relevant, given the particular question to be answered about the target system” (Parker, 2009, 493). The point is that the materiality of the system does not by itself guarantee this similarity. When experiments and data-driven methods leave out important substantive elements like interaction structures, detailed time ordering of events, and dynamic and multi-level feedback loops, the fact they rely on a sample of real subjects, that is, their materiality, does not make them *ipso facto* more reliable inferential devices than ABMs, which, although working only with virtual subjects coded by the computer program, typically do take into account all these aspects (for a similar remark applied to mathematical *versus* ABM, see Page 2008). Thus, according to the latter view, what ultimately must be assessed to judge the quality of the inferences made thanks to an ABM is “whether or not, to what extent, and under what conditions, a simulation reliably mimics the [...] system of interest” (Winsberg, 2003, 115). The argument from materiality is, in this respect, question-begging (on this point, see also Morrison 2015, 243). Like Winsberg, we believe that it is not material similarity in and by itself that grants the reliability of scientific inferences, but the (external) *validity* of the experiment, or the simulation, with respect to its intended target. We have previously established that, in suitable conditions, an ABM can be taken as an approximate replica of its target (see §3.1). In particular, when empirical information can be fed into it (empirical calibration) and independent datasets can be used to assess its outcomes (empirical validation), an ABM can be thought of as a reliable inferential device because it works as a “mimicking device”, to borrow a metaphor from Mary Morgan (2012, 337).

In this respect, experiments, data-driven methods, and computer simulations have to face the same challenge. As was the case with formal assumptions, materiality, too, is a red herring. It is in fact neither sufficient (see data-driven methods) nor necessary (see ABM) to ground the sort of similarity between model and target that is required for warranting the model’s (external) validity.

LACK OF VALIDITY TEST Even if one agrees that substantive justification is required for the assumptions of ABM and data-driven methods alike, and that materiality *per se* does not assure greater reliability of data-driven methods, one may still object that, in the end, data-driven methods grant credible inferences because, unlike ABM, in this case well-established procedures for testing the model’s validity exist.

We have acknowledged in §3.2 that it is still an open issue in the ABM field what counts as the best way of comparing simulated outcomes and empirical data describing the real-world outcomes under scrutiny. Still, with regard to this objection, we would like to reply that the testing methodology on which data-driven methods and ABM rely can in fact be interpreted along the same lines. As argued by Parker (2008), the aim of validity tests on simulations is, in analogy with statistical tests over experimental designs, to ensure that simulation results constitute good evidence for some real-world hypothesis *H*. This, in turn, depends on whether “(i) the results fit *H*; and (ii) it is unlikely that the simulation study would deliver results that fit so well with *H*, if *H* is false” (*ibid.*, 374). As a consequence, Parker argues, simulation hypothesis testing is ultimately animated by an error-statistical approach according to which

[...] in order to claim that simulation results provide good evidence for some hypothesis of interest, we would be required to show that the potential sources of error were unlikely to have been present or to have impacted the results by more than a specified amount, rather than just that the evidence collected so far is consistent with their absence or their having minimal impact. (*ibid.*, 382)

The procedures of empirical calibration and validation, as well as the operations of sensitivity, robustness, dispersion and model analysis, illustrated by the research discussed in §3.3, constitute the toolbox agent-based modelers can draw upon to track and quantify these sources of error. Although we are well aware that there are practical difficulties in performing such tests, our point is that it is unclear on what grounds one could deny that validity tests performed by data-driven methods and simulations follow in principle similar logics, and claim that external validity is harder to establish in the latter case (for a claim of this sort with respect to simulations *versus* experiments, see, e.g., [Morgan 2003](#), 231).

Obviously, we agree that, no matter how well calibrated, validated or tested an ABM is, it remains possible that the model lacks external validity. The mechanism postulated by the modeler may not capture the most relevant one. Indeed, the postulated mechanism may just happen to reproduce the data. The data may also be the result of a plurality of mechanisms, whose operations and relevance for the phenomenon is simply masked by the postulated mechanism. No (serious) modeler can completely exclude that the evidence provided by the simulation systematically underdetermines causal inference. As we have seen in §3.3, the impact of certain non-calibrated formal assumptions concerning, for instance, functional forms can be assessed by sensitivity and robustness analysis. However, this is not enough to guarantee the model's validity. Are the modeled behavioral/interaction rules correct? Are they executed in the correct order and at the correct rate? Even when empirical data are rich, some substantive assumptions are likely to remain untested.

But don't statistically-oriented modelers, especially from within the potential outcomes framework, face analogous challenges? As discussed at length in §4.1 and §4.2, some formal assumptions of data-driven methods can be partially tested with empirical data—e.g., the relevance condition in instrumental variable estimators—whereas others cannot, by construction—e.g., the exogeneity assumption in instrumental variable estimators or, more generally, the meta-assumption that the causal graph under scrutiny is correct. Thus, background knowledge cannot completely exclude that the conclusion reached by means of data-driven methods is systematically biased or confounded. That is why the most rigorous scholars recommend to perform sensitivity analysis to assess to what extent empirical estimates from data-driven methods are robust against confounders (on this point, recall Spirtes and Glymour [1997, 562]'s remark, which we reported in §4.1; for an overview on the topic, see [Gangl 2013](#), 385-90; for an application, see [VanderWeele 2011](#)). However, they also highlight that sensitivity analysis itself “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). Substantive assumptions are necessary to the inference, which in turn are not always verified or verifiable.

Thus, it seems fair to conclude that statistically-oriented scholars, simulationists and—one may add—experimentalists all struggle with analogous issues of precision,

accuracy and calibration. In each methodology, a complex and—it should be admitted—still loosely codified combination of background substantive knowledge, external empirical evidence, and testing statistical procedures is necessary. This combination of elements is typically mobilized in any serious application that aims to eliminate uncertainty as regards possible sources of error.

The question, then, is: On what ground can this combination of elements justify causal inference? The justification, we maintain, may be rationalized by a pattern of reasoning known in philosophy of science as “inference to the best explanation” (Lipton, 2004). The central idea is that “the phenomenon that is explained [...] provides an essential part of the reason for believing the explanation is correct” (Lipton, 2009, 629). In short, the hypothesis must be true, because if it were, it would provide the best explanation of the data.

In the case of ABM, the explanation that the truth of the ABM’s generative hypothesis provides for the match between empirical data and the simulation of a well-calibrated, validated and tested ABM is so “lovely”—to paraphrase Lipton—that it is rational to believe that the hypothesis is true. In turn, when this reasoning is at work, it makes the counterfactual scenarios where something changes in the simulated mechanism credible in reality, too. Analogously, in the case of data-driven methods, background knowledge in a given context makes it implausible that evidence of a robust dependence is not due to a causal relation, which is at the same time responsible for the different responses under different hypothetical circumstances. In both cases, the reasoning goes as follows: the explanation that the truth of the causal claim provides for the data is so “lovely” that it is rational to believe that the causal claim is true. This reasoning is, we maintain, as legitimate in the case of simulated outcomes as it is in the case of potential outcomes.

5 METHODOLOGICAL SYNERGY

In philosophy, it is common to distinguish between monistic and pluralist views on causation (see, e.g., Williamson, 2006): roughly, whereas the monist believes that there is one notion of cause, the pluralist believes that there are several notions. The latter thesis, labeled “conceptual pluralism” (Reiss, 2011b, 908), is endorsed by, among others, Hall (2004, §6) and Longworth (2006). Conceptual pluralism, in turn, should be distinguished from evidential pluralism, namely the (weaker) thesis that there are various kinds of evidence for causality (Russo and Williamson, 2007). In §3 and §4, we took conceptual pluralism at face value. We discussed pros and cons of ABM and data-driven methods, on the assumption that the views of causality that animate them—respectively the production view and the dependence view—are correct. In this section, we shall move beyond conceptual pluralism, which in our opinion has undesirable and implausible implications, and endorse an evidential pluralism about the kinds and sources of evidence of causality, which will be the backdrop of our view on the relation between different methods for causal inference. More precisely, we shall explain why, given the limitations shared by the two classes of methods, it would be profitable to shift the debate from defending the superiority of particular methods, and thus of specific understandings of causality and mechanism, to discussing how ABM and data-driven methods can provide a variety of evidential sources that compensate for each other’s weaknesses.

We motivate this normative recommendation through four dialectical steps. First,

we argue that acknowledging the legitimacy of a variety of methods for causal inference does not suffice to justify a radical form of pluralism but rather prompts the question of how scholars with different methodological orientations can still fruitfully communicate about causation. Second, we answer this question by endorsing a pragmatist theory of evidence that, we maintain, is key to reconstructing the way causal claims are supported and warranted by the scientific community. Third, we elaborate on this view, and argue for the integration of methods that produce evidence of a productive kind and of a dependence kind, and rationalize the role of ABM in this methodological synergy. Fourth, we further qualify our claim by explaining under what conditions a variety of evidential sources are effective for causal inference and why ABM and data-driven methods satisfy these conditions.

AGAINST CAUSAL EXCLUSIVISM The two previous sections have approached ABM and data-driven methods from the point of view of the intuitions about causality and mechanisms which animate them: a production view of causality and a vertical view of mechanisms for the former; a dependence view of causality and a horizontal view of mechanisms for the latter (we discussed these conceptual distinctions in §1.1 and §1.2). In particular, we asked under which conditions these two classes of methods can produce convincing evidence of causality along the lines of the specific understanding of causality supported by their internal machinery. Our conclusion was that, although both ABM and data-driven methods are *in principle* self-sufficient to produce convincing evidence of causality in their own terms, *in practice* both families of methods face challenges concerning data limitation and testability of assumptions. Both approaches make, necessarily, use of substantive assumptions, not all of which are verifiable. Both approaches have at their disposal validity tests, which are however partly based on conventions and incapable to conclusively rule out errors. In both cases, the evidence produced is likely to be imperfect. In both cases, causal claims ultimately require background, subject-specific, and expert knowledge that necessarily comes from outside the method and the data under scrutiny. Thus, there appears to be no knock-down argument, which would prove the superiority of either the ABM approach to causal inference or data-driven methods of causal inference.

The unavailability of such an argument, in turn, could be read as a justification of conceptual pluralism, namely the view that there are distinct and equally-legitimate notions of cause. In our view, however, this move would have obvious drawbacks. If conceptual pluralism is correct, users of a given notion of cause would be right in claiming an exclusive authority on how their own notion of cause ought to be used. By the same token, though, they would lose any authority on how the notion of cause is used in other communities. Worse still, seemingly appropriate disagreements on causal knowledge would turn out to be, on reflection, inappropriate. Since the meaning of the concepts involved are different, causal claims on the same matter, but formulated by different communities, which operate with different standards, could not contradict one another. They would rather be orthogonal to one another.

In our view, this causal “exclusivism” is not only normatively detrimental, insofar as it would, if endorsed, obstruct critical analyses of the limitations of each method for causal inference and thus impede the search for synergies that aim to compensate for them. It seems also descriptively inaccurate, insofar as it clashes with the observation

of how scientists from different communities *de facto* react to each other's results. It is certainly true that sometimes scientists talk past each other or ignore each other's results merely on the ground of the method that was used to obtain them. However, scientists also treat each other's results seriously, even when obtained by methods they do not favor, and as calling for further tests by other methods, tests that may lead them to retract previously endorsed conclusions. We propose that the scientists' attitude towards the evidence be understood in terms of a "default and challenge" model of justification (Reiss 2015, 354-5; cf. Brandom 1994, 177): scientists have a default trust in the evidence produced by other scientists, unless there are domain- or case-specific reasons to distrust it or challenge it. This is, we believe, the most direct indicator that the existing plurality of causal languages and practices are *not disconnected*.

To the extent that science is a successful enterprise, the scientific default-and-challenge process is truth- or at least consensus-conducive. This, in turn, requires that there be a level at which, irrespective of one's preference towards a particular method, successful communication and genuine disagreement are possible (on this point, see Casini, 2012). This, we argue, is the level at which "evidence" has been produced that can speak in favor or against a given causal claim, and scientists deliberate on how much "warrant" it lends to the claim. This view naturally raises two questions. First, what kind of evidence is worth collecting? And second, what gives to a piece of evidence a warrant-increasing value, in spite of the apparent difference between the causal views that different researchers subscribe to?

PRAGMATISM ABOUT EVIDENCE Both questions have been recently addressed by Julian Reiss (2015)'s pragmatist theory of evidence. The theory is motivated by the rejection of the view that there is a gold standard for causal inference (see, e.g., Diez Roux's long statement we quoted in the [Introduction](#)), which in turn would depend on mistaking (one's view on) what causality *is* with what constitutes *evidence* for it (Reiss 2015, 349; cf. Reiss 2012).

As to the first question—viz. what kind of evidence is worth collect-ing—the theory says there are several (well-known) kinds of facts one is (defeasibly) entitled to expect under the supposition that a causal claim is true, namely: the obtaining of correlations between putative cause and effect; the obtaining of changes in the effect after interventions on the cause; the fact that the cause is a necessary or sufficient condition for the effect; the existence of a mechanism for the relation; the existence of a continuous process triggered by the mechanism that leads from the cause to the effect. Studies that produce evidence for these facts provide evidence for causality—*irrespective* of one's convictions on what causality essentially is—, in two possible ways. First, they may *directly* support a causal hypothesis. Second, they may be incompatible with what one is entitled to expect under the supposition that some alternative hypothesis is true, thereby *indirectly* supporting the causal hypothesis. That the evidence may be collected by using certain methods is, for Reiss, a contingent fact, of no deep significance for the justification of causal inference. (Reiss' critical target is, in particular, the view that privileges experiments as the gold standard for collecting evidence.) What matters is how the evidence contributes to the plausibility of the causal claim in relation to relevant alternatives.

In the context of our discussion, one may paraphrase the above proposal as fol-

lows. What matters to causal inference is generating evidence that bears on assumptions, whose truth would allow one to safely draw the inference from data. In particular, the warrant for the inference comes from evidence that can confirm assumptions under which the causal claim is plausible or disconfirm assumptions under which it is implausible. This, by itself, does not entail that one should collect the evidence by using one method rather than another. In fact, the assumptions that would warrant the reliability of a method are among those very facts one would need to establish prior to deciding to use the method to collect evidence. Testing those assumptions requires further evidence, which requires further assumptions, and so on. This leads to our second question: what confers to a piece of evidence the capacity to warrant a causal claim?

By reference to the case of the inference to the causal effect of smoking on cancer, [Reiss \(2015, 355-7\)](#) suggests that a causal claim is warranted when a number of facts, expectable under the supposition that the claim is true, are together sufficient to convincingly rule out alternative explanations of those facts. For instance, it is known that smoking is strongly correlated to cancer, that more frequent smokers have a higher risk of developing cancer, that stopping has a beneficial effect, that cancer develops some time after taking up smoking, etc. Thus, it is plausible to conclude that smoking (rather than, say, some genetic factor) *causes* cancer. Clearly, this is a case of inference to the best explanation, in the sense introduced in §4.3: the truth of the causal claim is inferred, because it provides the best explanation of the evidence among a pool of alternative hypotheses, by either being directly supported by the evidence, or by being indirectly supported by evidence that rules out the alternatives.

Naturally, disagreement can arise on how one should move from collecting evidence to warranting a causal claim, especially when the results of studies that follow different methods pull in different directions. Different communities may prefer evidence coming from different sources. As we have shown earlier, ABMs produce evidence of one kind, based on generative mechanisms, whereas data-driven methods produce evidence of other kinds, based on other criteria, which are in a way or another based on a different-making intuition. However, we share Reiss' conviction that favoring a specific method for causal inference does not amount to a justification for believing that only a kind of evidence is relevant for establishing a causal claim. What matters to causal inference is not so much the source of the evidence but how it increases one's confidence in some relation being non-spurious, given constraints due to limited data and uncertain background assumptions. Various kinds of evidence may be relevant in different ways.

EVIDENTIAL VARIETY From the vantage point of the pragmatist theory of evidence, it is easier to make a further step, and understand why, although different researchers may place different weight on evidence coming from different sources, still *all* kinds of evidence can in principle bear on one's confidence that a causal claim is warranted. In this regard, [Russo and Williamson \(2007\)](#) have put forward a thesis, referred to as "evidential pluralism", to the point that evidence of both difference making and production is—at least in the context of biomedical studies—necessary to causal inference.²⁴

²⁴In passing, note that [Russo and Williamson \(2007\)](#) and subsequent discussions call evidence of productive relations "mechanistic". This entails a particular commitment to one understanding of mechanisms. We, in contrast, have preferred to be neutral on what a "mechanism" is.

Similar claims in the context of macroeconomics and econometrics have been made by, respectively, [Claveau \(2011, 2012\)](#) and [Moneta and Russo \(2014\)](#).

In short, the justification for evidential pluralism is the following (see, e.g., [Illari, 2011](#)). The uncertainty about the causal nature of a given relation may be reduced in (at least) two ways. On the one hand, evidence of credible entities and activities (i.e., a mechanism in a vertical sense, in the terminology introduced in §1.1) influences one's belief that some relation is genuinely causal. Intuitively, evidence of robust relations not backed by evidence of the entities and activities responsible for them does not eliminate the doubt that the relation is *confounded*. On the other hand, evidence of a robust relation, too (i.e., a mechanism in a horizontal sense, in the terminology introduced in §1.1), influences one's belief that the relation is causal. Intuitively, evidence of how entities interact to produce some behavior does not guarantee that their organized operation is not *masked* in the broader context of further (possibly unknown) modes of interaction and outer influences. Thus, if one accepts this justification of evidential variety, it becomes clear that mechanistic evidence (vertical sense) is not more fundamental than dependence evidence, or vice versa: rather, the two kinds of evidence complement each other. Let us now paraphrase this thesis in relation to evidence from ABM and data-driven methods.

The strength of the evidence produced by an ABM resides in its power to reduce the uncertainty on whether the source of some observed relation as depicted by the model does in fact capture the most likely and/or relevant social mechanism responsible for the relation. In principle, a careful use of the method, along the lines sketched in §4.1, can offer strong evidence for causality. In practice, as pointed out in §4.2, such a use is hampered by data and knowledge limitations, which may not be fully counterbalanced by the theoretical explorations described in §4.3. In such cases, data-driven methods can surely help. Indeed, ABMs make use of empirical knowledge got by other means, such as controlled experiments or observational studies, to formulate assumptions and calibrate the model or, more generally, to modify or refine a previous model. At the same time, although evidence coming from data-driven methods is useful for ABM, this does not make generative evidence dispensable. When the assumptions of data-driven methods that would control for confounding and bias are not well confirmed by data or justified by contextual, substantive considerations, ABM allows one to explore those scenarios over which macro- and/or individual-level data are missing. From the point of view of data-driven methods, ABM can in such cases be thought as a theory-driven sensitivity technique. It can be used to design a variety of mechanisms, making different hypotheses on different sets of confounders, and assess under which of these mechanisms the expected (or observed) correlations persist. In this way, ABM can increase one's confidence that a relation is not spurious, and that confounding is negligible.

In sum, ABM and data-driven methods provide necessary and complementary evidence of causality. On the one hand, ABM adds specificity to data-driven causal inference in that it formalizes, and deduces consequences from, specific sequences of events that may have generated the connections identified by data-driven methods, in particular when the latter's assumptions are hard to realize or difficult to check. On the other hand, data-driven methods add plausibility to ABMs in that they provide empirical information that constrains the modeler's theoretical creativity on both input and output side. Thus, the thesis of evidential variety ultimately fosters a virtuous circularity between

the two methods (for a similar proposal in the analytical sociology literature, see [Manzo 2014b](#), figure 1).

ELIMINATIVE DIVERSITY This recommendation requires a further qualification, though. After all, ABM and data-driven methods may produce evidence that is various but this evidence may be affected by qualitatively similar biases and errors. Under what conditions is the evidential variety produced by ABM and data-driven methods relevant to causal inference?

In our view, the condition under which evidential variety is effective is that the evidence-gathering methods, in particular the sets of assumptions on which the methods rely, are “independent”. The intuitive motivation is the following: independent sources provide more “diverse” evidence; and if diverse evidence agrees, the evidence is more confirmatory ([Bovens and Hartmann 2003](#), 96-7). To illustrate with an example from our earlier discussion of specific ABMs, when [Ajelli et al. \(2010, 2\)](#) compare a data-rich ABM with a more parsimonious meta-population model of disease diffusion, they make a remark to the same point: “The good agreement of the two approaches reinforces the message that computational approaches are stable with respect to different data integration strategies and modeling assumption”.

More precisely, as recently argued (in a Bayesian framework) by [Schupbach \(2015\)](#), the role of diverse evidence is to make it less plausible that a robust result is explained by any alternative hypothesis than it is by the hypothesis at stake. Accordingly, the kind of diversity that is required for the evidence of a robust result to be confirmatory is “eliminative diversity”: the evidence is (more) eliminatively diverse if it is (more) capable to rule out salient competitors of the hypothesis. Methods based on (partially) diverse assumptions are more likely to produce this sort of diverse evidence. As Schupbach observes with reference to the application of the Volterra principle to population dynamics,

[...] when seeking to confirm the Volterra Principle [...] diverse models may be identical but for some modest difference. By utilizing these [...] diverse models, we rule out confounding hypotheses pertaining to our result left standing by any subset of the models used alone. Notably, we alleviate worries that our result is an artifact of a simplifying assumption common to some subset of our models by duplicating that result using a new model that does not share that assumption. (*ibid.*, 314)

It is precisely this qualified sort of “independence”—viz. (partial) independence among background assumptions—that makes it possible that the combination of data-driven models and ABMs provide a stronger warrant for causal claims. Although in both cases assumptions are made that in practice are not fully testable, if the facts that one is entitled to expect under the supposition that the causal claim is true are reproduced by a model that does *not* make those untestable assumptions (albeit perhaps making other, untested assumptions), one is more justified to infer to the truth of the causal claim. In general, there is no reason to believe that the assumptions of an ABM—about how agents behave and interact and how such behaviors and interactions generate aggregates of interest—are going to be systematically related to the assumptions of data-driven methods—about how to causally interpret observed patterns among such aggregates—in such a way that the total evidence produced is not eliminatively diverse. On the contrary, often we have

reason to believe that the two sets of assumptions are sufficiently independent as to produce evidence that is eliminatively diverse in the required way. This justifies our plea for methodological synergy.

To conclude, let us note that the justification we propose for our plea for evidential variety already appears—albeit implicitly—in the (rare) writings that attempt to directly compare ABM and data-driven methods. For instance, with reference to epidemiological studies of obesity, [Marshall and Galea \(2014, 97\)](#) state:

[...] currently missing from the literature are comparative studies in which investigators interrogate an epidemiologic question with different types of causal inference models, including those that are agent-based. These may fruitfully be the focus of future work. [...] Constructing and calibrating an ABM with these same data would permit direct comparisons of the assumptions made by each method and would also reveal specific situations in which the ABM approach may provide novel and important public health insights to curb obesity.

Our goal in this section has been to make explicit why this reasoning is sound, and thus why methodological synergy is worth pursuing.

DISCUSSION AND CONCLUSION

Sociologists have insisted time and again on the importance of developing an autonomous reflection on causal analysis for sociology. Commenting on Liberson’s skepticism against randomized experiments as the gold-standard for sociological causal analysis, the British sociologists John [Goldthorpe \(2001, 8\)](#), for instance, has noted: “[...] sociologists have to find their own ways of thinking about causation, proper to the kinds of research that they can realistically carry out and the problems that they can realistically address”. More recently, Jeremy Freese has remarked that “[...] the professional philosophical literature on causality is often surprisingly unhelpful [...]”. No uncontroversial general philosophical account of causality exists, and social researchers have plenty of our own work to do while we wait” ([Freese and Kevern, 2013, 28](#)).

By relying on selected works from a variety of disciplines, including the philosophy of the social sciences, our paper aimed at contributing to this meta-theoretical project about causality for sociology. To this end, we chose a specific methodological entry, however. The question that guided us concerns whether a specific simulation technique, namely ABM, can produce causally-relevant evidence. This seemed to us a fruitful point of view for three main reasons (we will discuss some of its limitations below). First, while ABM is entering the accepted toolbox of quantitative sociologists, markedly different views are still expressed as to its capacity to contribute to empirical research. Thus, a principled and systematic assessment of its causal value seemed necessary. Second, given the link between ABM and mechanism-based explanations, raising the issue of the causal value of this technique allows one to connect three research streams (namely, on causation, social mechanisms, and simulation methods) that are rarely considered at once. Finally, since assessing the causal relevance of agent-agent simulations requires that alternative methodological approaches are also considered, a systematic, and explicit, comparison between ABM and data-driven methods becomes possible in relation to the specific task of causal inference.

Upon endorsing this comprehensive perspective, we first inquired the sources of the observed disagreement about the causal value of ABM. We identified two of them. On the one hand, we noted, claims on the alleged incapacity of this method to contribute to causal analysis compared to data-driven methods rely on specific views on what causality and mechanisms are. In particular, dependence accounts of causality—according to which causes make a non-spurious difference to their effects—were contrasted to production accounts—according to which causes bring about their effects via a continuous chain of events. We showed that this opposition is reinforced by the fact that it squares with two opposing understandings of mechanisms, with the followers of dependence accounts of causality regarding a mechanism as a chain of intervening variables (we have labeled this perspective the “horizontal view” of mechanisms) and the followers of production accounts regarding a mechanism as a set of entities and activities that dynamically produce an outcome through iterative aggregation steps (we have labeled this perspective the “vertical view”). Our diagnosis was that scholars who endorse the former view tend to deny the role of ABM for causal analysis whereas those who accept the idea of “generative causality” tend to formulate a more favorable assessment. On the other hand, we noted, there exists a second, more transversal reason for the difference in appreciations of the causal value of ABM. This reason is the variety of types of ABMs that populate the literature across several disciplines. Among these types, speculative ABMs are still by far more frequent than theoretically plausible (meaning solidly grounded in existing, confirmed theories), empirically-calibrated and validated ones. This has led many observers to take a pattern of current applications for an intrinsic limitation of the technique, and to judge ABM as fundamentally incapable to integrate appropriate empirical information. As a consequence, even among scholars who overtly endorse a production understanding of causation, ABM is seen as promising for exploring the “causal adequacy” of a mechanism but not to prove that it is actually at work. As recently put by [Goldthorpe \(2016, ch. 9\)](#), “[...] to show the generative sufficiency of a mechanism is not to show that it is in fact this mechanism that is in some particular instance at work”.

Having accepted this variety of views on causality and mechanism and having acknowledged the full range of technical options offered by ABMs, we then devoted the second part of our study to identifying under which conditions this method can contribute to causal analysis. In particular, we argued, an ABM can produce evidence that is causally-relevant (from a production/vertical view of causality/mechanism) when it acts as a “mimicking device” (according to a metaphor by Mary Morgan), that is, when it generates the high-level regularities of interest (meaning that empirical validation is performed) under assumptions and conditions that were in turn based on empirical data proving their realism (meaning that empirical calibration is also performed). To the potential objection that these conditions are difficult to fulfill *in practice* because of the lack of sufficiently fine-grained data, we replied that ABM can rely on systematic sensitivity, robustness, and uncertainty analysis to quantify the variability of the model’s outcome, in particular for those pieces of the model for which empirical information is partial or missing, and, on the other hand, on several, although not yet formalized, strategies to describe the dynamic through which the model generated the outcome of interest. Both the quantification of the outcome’s stability and the transparency of the model’s dynamics, we argued, are key to increasing an observer’s confidence in the external validity of

the mechanism postulated by the model.

As testified by several quotations reported in the text, scholars animated by a dependence account of causality (and an horizontal view of mechanisms) would counter-object to us that, as long as some pieces of the internal machinery of an ABM only rely on theoretical explorations, the model's capacity to tell us whether the postulated mechanism actually works in the real world—and thus its capacity to produce causally-relevant evidence of a production kind—remains very limited. In order to show that ABM is not exceptional in this respect, we pushed further our analysis and scrutinized some data-driven methods—namely randomized experiments, instrumental variables, and causal graphs—that are usually regarded as good methods for causal inference (from the point of view of dependence/difference-making accounts of causation). We showed that these methods, too, ultimately justify their causal conclusions by arguments that do not exclusively rely on empirical data. The simple reason for this is that, similarly to ABM, the assumptions required to causally interpret the observed connections cannot be tested empirically because of the lack of data or because it is impossible to empirically adjudicate the truth of the assumptions. Thus, we concluded, both ABM and data-driven methods can reach their respective goals, namely convincing an observer of, respectively, the external validity of the postulated chains of events and the non-spurious character of the detected dependence, only by combining partial empirical information with theoretical, subject-specific knowledge that is exogenous to the data directly used to calibrate/validate the model or document the putative connection.

As our introduction to §4 shows, we are well aware that several statisticians and sociologists have already forcefully argued that the causal interpretation of statistical estimates requires more than empirical data alone (for a recent review of this debate, see [Goldthorpe 2016](#), ch. 8). However, we believe that the original and, we suspect, more controversial part of our analysis is the implication that we drew from this argument, namely that upon scrutiny, there is no cogent argument to the point that data-driven methods are superior to other methods, namely simulation-based methods, in establishing causal claims. In economics, by partly building on the work of Nancy Cartwright, which we also considered, Angus Deaton has recently defended this argument with respect to randomized experiments. Based on an analysis of “ideal conditions” and “practical problems” of implementation of this method, he has claimed:

Randomized controlled trials cannot automatically trump other evidence, they do not occupy any special place in some hierarchy of evidence, nor does it make sense to refer to them as “hard” while other methods are “soft”. These rhetorical devices are just that; metaphor is not argument, nor does endless repetition make it so. ([2010](#), 426)

We proposed to extend this view to statistical methods for observational data and simulation methods, namely ABM.

Over and above the fact that all methods for causal inference ultimately rely on a complex mix of theory and data—such that, when properly understood, any of them can be claimed to be more “empirical” than others—there is a deeper reason that motivates our rejection of the view that one class of methods—and the specific understanding of causality/mechanisms behind them—would be more fundamental, and thus more appropriate, to establish credible causal claims. This reason, which we discuss in the paper's last section (§5), is that the evidence produced by each family of methods, no matter

how firmly grounded in theory and data, can only partly capture the information needed to reduce the uncertainty about the causal nature of a given relation. On the one hand, evidence of credible sequence of events sustained by interactions among (credible) entities and activities (i.e., a mechanism in a vertical sense) influences one's belief that some relation is genuinely causal. Intuitively, evidence of robust relations not backed by evidence of the entities and activities responsible for them does not eliminate the doubt that the relation is confounded. On the other hand, evidence of a robust relation, too (i.e., a mechanism in a horizontal sense), influences one's belief that the relation is causal. Intuitively, in fact, evidence of how entities interact to produce some behavior does not guarantee that their organized operation is not masked in the broader context of further (possibly unknown) modes of interaction and outer influences. That is why, we argued, mechanistic evidence (in the vertical sense) should not be seen as more fundamental than difference-making evidence, or vice versa. Rather, the two kinds of evidence complement each other. This argument has ultimately motivated our plea for a methodological synergy between randomized experiments, ABM, and statistical methods for observational data—a synergy in virtue of which different but complementary kinds of evidence, equally necessary to establish a causal claim, are generated.

Thus, in our view, the most general contribution of this paper consists in shifting the focus of current methodological debates from proposing arguments that defend the superiority of particular methods, arguments based on an often implicit endorsement of particular notions of causality and mechanism, to discussing how different methods, hence different understanding of causation and mechanisms, can be combined to produce different kinds of evidence that can complement each other's weaknesses. The British sociologist John Goldthorpe seems to go in the same direction when, with regard to the specific task of proving the empirical adequacy of a postulated mechanism (in our terminology, a vertical mechanism), he claims:

What is important is that the actual operation of mechanisms should be tested in as many ways as is possible and the results obtained be considered in relation to each other. It should not be expected that any particular test will produce 'clinching results', at least not of a positive kind, but at best 'vouching' results – to take up Cartwright's (2007: ch. 3) useful distinction; and greatest weight has then to be given to how far results from different tests do, or do not, 'fit together'. (Goldthorpe, 2016, ch. 9)

By building on a principled and systematic assessment of the causal value of a specific simulation technique, namely ABM, and comparing it to data-driven methods, our paper attempted to generalize this synergistic view to causal analysis as a whole.

To conclude, we would like to briefly discuss three points that we perceive as major potential objections to our analysis. First of all, it may be argued that our starting question itself, namely what the causal value of ABM is, is neither here nor there. From a philosophical point of view, this objection would arise from the widespread view that causation and constitution are different kinds of relations (cf. Craver and Bechtel, 2007). Whilst causation obtains between spatiotemporally distinct entities (cause and effect are distinct events or states) and is asymmetric (intuitively, effects depend on causes), constitution obtains between spatiotemporally overlapping entities (parts and wholes) and is non-asymmetric (in a qualified sense, parts depend on wholes and vice versa). If, as we argued, ABMs study how changes in dynamic loops between micro- and macro-level variables trigger changes in other macro-level variables, then ABMs study constitutional

rather than causal relations. Consequently, ABMs are in the business of generative explanation, and not of causal inference. From within sociology, one may make a parallel point, by arguing that we ignored the distinction between “description” and “explanation”. As recently restated by [Goldthorpe \(2016, ch. 1\)](#), description concerns detecting, or making “visible”, probabilistic regularities in the social world, whereas explanation concerns providing narratives, which make “transparent” how these regularities were brought about. One thing is to *see* regularities. To this end, multivariate statistics is very good. Another thing is to *see through* them. To this end, theory is needed. With this distinction in mind, one may conclude that, if ABMs, as we argued, aim to establish which low-level chains of events lead to observed robust connections between aggregates, then ABMs are not a method for causal inference but a device for generating (potential) explanations. In other words, so the objection, whilst ABMs are concerned with explaining causal states of affairs on the assumption that they obtain, a proper method for causal inference is concerned with establishing whether causal states of affairs obtain in the first place.

Whilst we understand the rationale behind this objection, we believe that it fails to appreciate the way we have proposed to re-conceptualize causal analysis. Our point is that several understandings of causation are *equally* legitimate and that the evidence produced from within these various approaches is *equally* necessary to reduce the uncertainty about the non-accidental nature of a given observed connections among happenings. From this point of view, both randomized experiments and statistical techniques for observational data on the one hand, and ABMs on the other hand, are methods for causal inference. In the first family of methods, causal inference is about generalizing correlations from the sample under scrutiny to a target population. Within ABMs, in contrast, causal inference is about extrapolating the mechanism neatly described *in silico* to the real world. The first kind of causal inference is tied to a dependence/difference-making view of causality. Description is an activity that emanates from this view. The second kind of causal inference, in contrast, is tied to a production view of causality. Explanation is an activity associated to this understanding of causality. Our ultimate suggestion is to stop trying to establish a hierarchy between intuitions about causation, and between descriptive and explanatory activities. Our argument is based on the conviction that the internal structure of different methods makes them more appropriate to pursue different sorts of causal analysis. However, this difference should be understood in collaborative, and not conflicting, terms, insofar as different kinds of evidence are necessary to establish causal claims.

A second objection that one may address to our analysis is that, no matter how appealing this synergistic view is, the division of labor that it implies on a methodological level is not clearly spelled out. From a philosophical point of view, this objection would arise from the observation that, although we have argued in favor of “evidential variety”, we have not formulated any specific recipe for combining different types of evidence. This would reveal, philosophers may argue, a more fundamental problem: our analysis lacks a fully developed theory of evidence. While we regard this objection as a strong one, we invite (charitable) readers to consider two elements. First, philosophers themselves are in disagreement about the utility of available philosophical theories of evidence for empirically-minded scientists ([Achinstein, 2000](#); [Cartwright et al., 2010](#)); moreover, among philosophers, the problem of how integrating different types of evi-

dence is an open one ([Williamson, 2015](#)). Second, from a sociological point of view, our analysis in fact implies a clear *modus operandi* for causal analysis in social sciences.

According to this *modus operandi*, establishing credible causal claims would ideally amount to loop the following steps. First, one should employ experimental and/or statistical methods for observational data to show that, given the best data available, the probability that the connection between a set of happenings of interest is observed by chance is very low. This step is performed from within a dependence/difference-making account of causality. Second, one should formulate a clear set of hypotheses about the underlying chains of actions, interactions, and their constraints, which are believed to be responsible for the observed connection. This step is performed from within the production understanding of causality and the vertical view of mechanisms. Third, the hypotheses should be translated into a formal model, whose behavior should be simulated in order to see if the dynamics triggered by the model is able to generate the observed connection. For this step, we have argued that ABMs are an especially powerful tool because of the granularity they allow for theoretical mechanism design. Fourth, the simulation should be constrained by as much empirical information as possible in order to see if, when as many pieces of the simulation machinery as possible are based on empirical regularities, the model's dynamics is still able to generate the observed connection of interest.

In fact, our plea for evidential variety leads to a specific research path in which dependence/difference-making and production accounts of causality (as well as horizontal/vertical views of mechanism) interpenetrate. This is especially visible in the third and fourth steps: the empirical validation and calibration of an ABM indeed requires information obtained from data-driven methods. It seems to us that this research path is consistent with existing proposals from empirically-minded scholars. In particular, [Goldthorpe \(2016\)](#)'s recent manifesto for sociology as a population science defends a research program that combines multivariate statistical techniques, which generate sophisticated descriptions of macro-level probabilistic regularities, with theoretical reasoning, which offers narratives at the level of actions and interactions that explain the robust patterns detected. As compared to Goldthorpe's proposal, our own puts greater emphasis on ABM as a tool to provide not only proofs of the "generative" sufficiency of a set of hypotheses but also to produce, through empirical calibration, evidence that the simulated narrative is empirically plausible. In this respect, our ideal sequence of research steps amounts to a protocol that, although not formulated in causal terms, appears in some variants of the research program now known as analytical sociology (for an overview, see [Manzo, 2014b](#), ch. 1).

This leads to the last potential objection that we would like to discuss. In short, this objection would attack our investigation for its limited scope. Although for different reasons, both quantitative and qualitative scholars may formulate this critique. Among quantitative scholars, those more attracted by experimental and statistical methods may object that we considered only a few data-driven methods in order to build our argument about the similarity between these methods and ABM in terms of comparable incapacity to establish causal claims by relying only on the empirical data under scrutiny. To this, our reply would be that, although we admit that our choice cannot do full justice of the variety of data-driven methods, our selection covers all the basic types of identification strategies, and thus seems sufficient to illustrate the general point of interest:

no matter what specific method is chosen, data-driven methods, too, repeatedly rely on assumptions that, because data limitation and/or logical construction, cannot be tested empirically. The fractal nature of this problem, meaning that for checking/testing some assumptions other assumptions are needed, is effectively illustrated by [Berk \(2010, 482\)](#) when, commenting on regression, he states: “For both the diagnostics and the remedies, new and untestable assumptions are required even before one gets to a number of thorny technical complications”.

As to the smaller community of quantitative scholars who, in addition to experimental and statistical methods, routinely work with analytically tractable mathematical models (like differential equations or game-theoretical models), they may complain about our exclusive focus on a specific form of computer-based modeling, namely ABMs. To this, we would reply that our admittedly restricted focus seems justified given that ABM has only recently entered the sociological toolbox, which explains, we believe, why a principled and systematic assessment of its causal value is still missing. In addition, we have explained the deep technical reasons that give ABMs a strong capacity to relax unrealistic assumptions, which leads to higher granularity in mechanism design, as well as a great deal of flexibility in handling the dynamic process of moving from lower to higher levels of analysis. For this reason, as compared to other forms of mathematical and computational modeling (for such a comparison, see, e.g., [Macy and Flache, 2009](#)), ABMs seems a more flexible method for causal inference, among those that are based on a production view of causality, which justifies our focused analysis of its potential for causal analysis in comparison with data-driven methods of causal inference.

Finally, more qualitatively-oriented scholars may also object to the limited scope of our investigation, on the ground that the value of ABM for causal inference is silent about the role that qualitative evidence may play to establish credible causal claims. This limitation arises indeed from the fact that the primary target we had in mind when writing this paper were statistically-minded sociologists who—more often in informal settings than in print—deny that ABMs have any value for causal analysis. Thus, our chosen focus should not be seen as in opposition to the attempt to combine qualitative and quantitative sources of evidence. Quite the contrary, we believe that qualitative evidence, namely from ethnographic observations, in-depth interviews, document or historical archives, can play a fundamental role, in particular at the stage of empirical calibration of an ABM, thereby helping to adjudicate between different low-level assumptions that may, within the virtual world of the simulation, lead to similar results. In other words, qualitative evidence may help to reduce the problem of “equi-finality” (or “multiple realizability”), thereby increasing the credibility of claims about the external validity of the simulated mechanisms. That is why a recent stream of research in the literature on ABMs focuses on how to inject qualitative evidence within these models ([Edmonds, 2015](#); [Ghorbani et al., 2015](#)). On the other hand, it is also meaningful to remark that prominent qualitative scholars, too, in particular from within the “process tracing” approach (for a philosophical discussion, from the point of view of causal inference, see [Steel 2007](#), ch. 8), acknowledge that simulating a qualitatively-inspired narratives through ABMs can help “to check the plausibility of inferences about causal mechanisms derived from process tracing” ([Bennett and Checkel, 2014](#), ch. 1). These scholars, too, defend the view that “diversity and independent evidence are useful in testing explanation”. Thus, although we had a quantitative target in mind and although

we focused on a specific simulation-based technique, our argument to the point that methodological synergy and evidential variety are necessary to support credible causal inference in fact aligns with arguments put forward by well-developed and identified qualitative approaches.

Notwithstanding these limitations—and others the reader may find—we believe that our argument may give more solid and transparent conceptual foundations to a new methodology, viz. ABM, and foster a more synergistic approach to causal inference in sociology. Research examples going into this direction are still rare (for a recent exception, see [Bruch, 2014](#)). It would be a welcome result if our meta-theoretical but at the same time methodologically-informed analysis succeeded in informing further investigations along the same lines.

REFERENCES

- Abbott, A. (1988). Transcending General Linear Reality. *Sociological Theory*, 6:169–186.
- Abbott, A. (1997). Seven Types of Ambiguity. *Theory and Society*, 26:357–391.
- Abbott, A. (1998). The Causal Devolution. *Sociological Methods and Research*, 27(2):148–181.
- Abbott, A. (2001). *Time Matters: On Theory and Method*. University of Chicago Press.
- Abbott, A. (2007). Mechanisms and Relations. *Sociologica*, 2:1–22.
- Abend, G., Petre, C., and Sauder, M. (2013). Styles of Causal Thought: An Empirical Investigation. *American Journal of Sociology*, 119:602–654.
- Achinstein, P. (2000). Why Philosophical Theories of Evidence Are (And Ought to Be) Ignored by Scientists. *Philosophy of Science*, 67:S180–S192.
- Ajelli, M., Gonalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., Merler, S., and Vespignani, A. (2010). Comparing Large-Scale Computational Approaches to Epidemic Modeling: Agent-based versus Structured Metapopulation Models. *BMC Infectious Diseases*, 10(190):1–13.
- Ajelli, M., Merler, S., Pugliese, A., and Rizzo, C. (2011). Model Predictions and Evaluation of Possible Control Strategies for the 2009 A/H1N1v Influenza Pandemic in Italy. *Epidemiology & Infection*, 139(1):68–79.
- Alexander, J. M. (2007). *The Structural Evolution of Morality*. Cambridge: Cambridge University Press.
- Angrist, J. D. and Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *Quarterly Journal of Economics*, 106(4):979–1014.
- 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.
- Angrist, J. D. and Pischke, J. (2010). The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics. *Journal of Economic Perspectives*, 24:3–30.
- Antonakis, J., Bendahan, S., Jacquart, P., and Lalive, R. (2010). On Making Causal Claims: A Review and Recommendations. *The Leadership Quarterly*, 21:1086–1120.
- Arthur, W. B. (2006). Out-of-Equilibrium Economics and Agent-Based Modeling. In Tesfatsion, L. and Judd, K. L., editors, *Handbook of Computational Economics. Agent-based Computational Economics*, volume 2, pages 1551–1564. North Holland: Elsevier.
- Arthur, W. B., LeBaron, B., Palmer, B., and Taylor, R. (1997). Asset Pricing under Endogenous Expectations in an Artificial Stock Market. In Arthur, W. B., Durlauf, S. N., and Lane, D. A., editors, *Economy as an Evolving Complex System II*, volume XXVII, pages 15–44. Santa Fe Institute Studies in the Science of Complexity, Reading, MA: Addison-Wesley.
- Auchincloss, A. H. and Roux, A. V. D. (2008). A New Tool for Epidemiology: The Usefulness of Dynamic-Agent Models in Understanding Place Effects on Health. *American Journal of Epidemiology*, 168(1):1–8.
- Axelrod, R. (1997). *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*. Princeton: Princeton University Press.
- Axtell, R. L. (2000). Why Agents? On the Varied Motivations for Agent Computing in the Social Sciences. Technical report, Center on Social and Economic Dynamics, Brookings Institution, Washington, DC. Working Paper 17.

- Axtell, R. L. (2001). Effects of Interaction Topology and Activation Regime in Several Multi-Agent Systems. In Moss, S. and Davidsson, P., editors, *Multi-Agent-Based Simulation: Lecture Notes in Computer Science*, pages 33–48. Berlin: Springer.
- Axtell, R. L. and Epstein, J. M. (1996). *Growing Artificial Societies: Social Science From the Bottom Up*. Brookings Institution Press.
- Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., Chakravartya, S., Hammonda, R., Parkera, J., , and Parkera, M. (2002). Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences*, 99:7275–7279.
- Axtell, R. L., Epstein, J. M., and Peyton Young, H. (2006). The Emergence of Classes in a Multi-agent Bargaining Model. In Epstein, J. M., editor, *Generative Social Science: Studies in Agent-based Computational Modeling*. Princeton: Princeton University Press, student edition.
- Barberousse, A., Franceschelli, S., and Imbert, C. (2009). Computer Simulations as Experiments. *Synthese*, 169:557–574.
- Barringer, S. N., Eliason, S. R., and Leahey, E. (2013). A History of Causal Analysis in the Social Sciences. In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 9–26. Dordrecht: Springer.
- Bechtel, W. and Abrahamsen, A. (2005). Explanation: A Mechanist Alternative. *Studies in the History and Philosophy of the Biological and Biomedical Sciences*, 36:421–441.
- Bennett, A. and Checkel, J. T. (2014). Process Tracing. From Philosophical Roots to Best Practices. In Bennett, A. and Checkel, J. T., editors, *Process Tracing: From Metaphor to Analytic Tool*. Cambridge: Cambridge University Press.
- Berk, R. (2010). What You Can and Can’t Properly Do With Regression. *Journal of Quantitative Criminology*, 26:481–487.
- Berk, R. A., Brown, L., George, E., Pitkin, E., Traskin, M., Zhang, K., and Zhao, L. (2013). What You Can Learn from Wrong Causal Models. In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 403–424. Dordrecht: Springer.
- Bianchi, F. and Squazzoni, F. (2015). Agent-based Models in Sociology. *Computational Statistics*, 7(4):284–306.
- Billari, F. and Prskawetz, A., editors (2003). *Agent-Based Computational Demography: Using Simulation to Improve our Understanding of Demographic Behaviour*. Heidelberg: Physica Verlag.
- Billari, F. C., Prskawetz, A., Diaz, B. A., and Fent, T. (2007). The “Wedding-Ring”: An Agent-Based Marriage Model Based on Social Interaction. *Demographic Research*, 17(3):59–82.
- Birks, D., Townsley, M., and Stewart, A. (2012). Generative Explanations of Crime: Using Simulation to Test Criminological Theory. *Criminology*, 50(1):221–254.
- Boero, R., Bravo, G., Castellani, M., and Squazzoni, F. (2010). Why Bother with What Others Tell You? An Experimental Data-Driven Agent-Based Model. *Journal of Artificial Societies and Social Simulation*, 13(3):6.
- Boero, R. and Squazzoni, F. (2005). Does Empirical Embeddedness Matter? Methodological Issues on Agent-Based Models for Analytical Social Science. *Journal of Artificial Societies and Social Simulation*, 8(4):6.
- Bollen, K. A. (2012). Instrumental Variables in Sociology and the Social Sciences. *Annual Review of Sociology*, 38:37–72.
- Boudon, R. (1974). *Education, Opportunity, and Social Inequality*. New York: John Wiley & Sons, Inc.
- Boudon, R. (1976). Comment on Hausers review of Education, Opportunity, and Social Inequality. *American Journal of Sociology*, 81(5):1175–1187.
- Boudon, R. (1979). Generating Models as a Research Strategy. In Rossi, P. H., editor, *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld*, pages 51–64. New York: The Free Press.
- Boudon, R. (2012). Analytical Sociology and the Explanation of Beliefs. *Revue européenne des sciences sociales*, 50(2). doi: 10.4000/ress.2165.
- 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.
- Bovens, L. and Hartmann, S. (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.
- Brandom, R. B. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge: Harvard University Press.
- Breen, R. and Karlson, K. B. (2013). Counterfactual Causal Analysis and Nonlinear Probability Models. In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages

- 167–187. Dordrecht: Springer.
- Brenner, T. and Werker, C. (2007). A Taxonomy of Inference in Simulation Models. *Computational Economics*, 30:227–244.
- Brown, D. and Robinson, D. (2006). Effects of Heterogeneity in Residential Preferences on an Agent-Based Model of Urban Sprawl. *Ecology and Society*, 11(1):46.
- Bruch, E. (2014). How Population Structure Shapes Neighborhood Segregation. *American Journal of Sociology*, 119(5):1221–1278.
- Bruch, E. and Atwell, J. (2015). Agent-Based Models in Empirical Social Research. *Sociological Methods and Research*, 44(2):186–221.
- Bruch, E. and Mare, R. (2006). Neighborhood Choice and Neighborhood Change. *American Journal of Sociology*, 112:667–709.
- Bruch, E. and Mare, R. (2009). Preferences and Pathways to Segregation: Reply to van de Rijt, Siegel, and Macy. *American Journal of Sociology*, 114:1181–1198.
- Bryan, M. L. and Jenkins, S. P. (2015). Multilevel Modelling of Country Effects: A Cautionary Tale. *European Sociological Review*. doi:10.1093/esr/jcv059.
- Carley, K. M. (2002). Computational Organization Science: A New Frontier. *PNAS*, 99(3):7257–7262.
- Cartwright, N. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press.
- Cartwright, N. (2004). Causation: One Word, Many Things. *Philosophy of Science*, 71:805–819.
- Cartwright, N. (2007a). Are RCTs the Gold Standard? *Contingency and Dissent in Science*, CPNSS, LSE. Technical Report 01/07.
- Cartwright, N. (2007b). *Hunting Causes and Using Them*. Cambridge: Cambridge University Press.
- Cartwright, N., Goldfinch, A., and Howick, J. (2010). Evidence-Based Policy: Where Is Our Theory of Evidence? *Journal of Childrens Services*, 4(4):6–14.
- Casini, L. (2012). Causation: Many Words, One Thing? *Theoria*, 27(74):203–219.
- Casini, L. (2014). Not-so-minimal Models: Between Isolation and Imagination. *Philosophy of the Social Sciences*, 44(5):646–672.
- Casini, L., Illari, P., Russo, F., and Williamson, J. (2011). Models for Prediction, Explanation and Control: Recursive Bayesian Networks. *Theoria*, 26(70):5–33.
- Castellano, C., Fortunato, S., and Loreto, V. (2009). Statistical Physics of Social Dynamics. *Reviews of Modern Physics*, 81:591–646.
- Cederman, L. E. (2005). Computational Models of Social Forms: Advancing General Process Theory. *American Journal of Sociology*, 110(4):864–893.
- Chattoe-Brown, E. (2014). Using Agent Based Modelling to Integrate Data on Attitude Change. *Sociological Research Online*, 19(1):16.
- Chavali, A. K., Gianchandani, E. P., Tung, K. S., Lawrence, M. B., Peirce, S. M., and Papin, J. A. (2008). Characterizing Emergent Properties of Immunological Systems with Multi-cellular Rule-based Computational Modeling. *Trends in Immunology*, 29:589–599.
- Clarke, B., Leuridan, B., and Williamson, J. (2014). Modeling Mechanisms with Causal Cycles. *Synthese*, 191:1651–1681.
- Claveau, F. (2011). Evidential Variety as a Source of Credibility for Causal Inference: Beyond Sharp Designs and Structural Models. *Journal of Economic Methodology*, 18(3):233–253.
- Claveau, F. (2012). The Russo–Williamson Theses in the Social Sciences: Causal Inference Drawing on Two Types of Evidence. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43:806–813.
- Cointet, J.-P. and Roth, C. (2007). How Realistic Should Knowledge Diffusion Models Be? *Journal of Artificial Societies and Social Simulation*, 10(3):5.
- Coleman, J. S. (1986). Social Theory, Social Research, and a Theory of Action. *American Journal of Sociology*, 91:1309–1335.
- Coleman, J. S. and Hoffer, T. (1987). *Public and Private Schools: The Impact of Communities*. New York: Basic Books.
- Coleman, J. S., Hoffer, T., and Kilgore, S. (1982). *High School Achievement: Public, Catholic and Private Schools Compared*. New York: Basic Books.
- Courgeau, D., editor (2003). *Methodology and Epistemology of Multilevel Analysis. Approaches from Different Social Sciences*. Dordrecht: Kluwer.
- Cox, D. R. (1992). Causality: Some Statistical Aspects. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 155(2):291–301.
- Cox, D. R. (2006). *Principles of Statistical Inference*. Cambridge: Cambridge University Press.
- Craver, C. and Bechtel, W. (2007). Top-Down Causation Without Top-Down Causes. *Biology*

- and *Philosophy*, 22:547–563.
- Craver, C. F. (2007). *Explaining the Brain*. Oxford: Oxford University Press.
- De Grauwe, P. (2010). Top-Down versus Bottom-Up Macroeconomics. *CESifo Economic Studies*, 56(4):465–497.
- de Marchi, S. and Page, S. E. (2014). Agent-Based Models. *Annual Review of Political Science*, 17:1–20.
- Deaton, A. (2010). Instruments, Randomization, and Learning about Development. *Journal of Economic Literature*, 48:424–455.
- Deffuant, G., Weisbuch, G., Amblard, F., and Faure, T. (2003). Simple Is Beautiful ... and Necessary. *Journal of Artificial Societies and Social Simulation*, 6(1):6.
- Delre, S. A., Jager, W., Bijholt, T. H. A., and Janssen, M. A. (2010). Will It Spread or Not? The Effects of Social Influences and Network Topology on Innovation Diffusion. *Journal of Product Innovation Management*, 27:267–282.
- Demeulenaere, P., editor (2011). *Analytical Sociology and Social Mechanisms*. Cambridge: Cambridge University Press.
- Diez Roux, A. V. (2014). The Virtual Epidemiologist—Promise and Peril. *American Journal of Epidemiology*, 181(2):100–102.
- DiMaggio, P. and Garip, F. (2011). How Network Externalities Can Exacerbate Intergroup Inequality. *American Journal of Sociology*, 116(6):1887–1933.
- Doreian, P. (1999). Causality in Social Network Analysis. *Sociological Methods and Research*, 30(1):81–114.
- Duffy, J. (2006). Agent-Based Models and Human Subject Experiments. In Tesfatsion, L. and Judd, K. L., editors, *Handbook of Computational Economics. Agent-Based Computational Economics*, volume 2, pages 949–1011. North Holland: Elsevier.
- Duflo, E., Glennerster, R., and Kremer, M. (2008). Using Randomization in Development Economics Research: A Toolkit. In Schultz, T. P. and Strauss, J., editors, *Handbook of Development Economics*, pages 3895–3962. Amsterdam and Oxford: Elsevier, North-Holland.
- Dugundji, E. R. and Gulyás, L. (2008). Sociodynamic Discrete Choice on Networks in Space: Impacts of Agent Heterogeneity on Emergent Outcomes. *Environment and Planning B: Planning and Design*, 35:1028–1054.
- Durlauf, S. N. and Ioannides, Y. M. M. (2010). Social Interactions. *Annual Review of Economics*, 2:451–478.
- Edmonds, B. (2015). Using Qualitative Evidence to Inform the Specification of Agent-Based Models. *Journal of Artificial Societies and Social Simulation*, 18(1):18.
- Edmonds, B. and Moss, S. J. (2005). From Kiss to KIDS – an ‘Antisimplistic’ Modelling Approach. In et al., P. D., editor, *Multi Agent Based Simulation 2004*, volume 3415 of *Lecture Notes in Artificial Intelligence*, pages 130–144. Berlin: Springer.
- Elster, J. (2009a). Excessive Ambitions. *Capitalism and Society*, 4(2). doi: 10.2202/1932-0213.1055.
- Elster, J. (2009b). *Alexis de Tocqueville: The First Social Scientist*. New York: Cambridge University Press.
- Elwert, F. (2013). Graphical Causal Models. In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 245–273. Dordrecht: Springer.
- Elwert, F. and Winship, C. (2014). Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. *Annual Review of Sociology*, 40:31–53.
- Epstein, J. M. (1999). Agent-Based Computational Models and Generative Social Science. *Complexity*, 4(5):41–60.
- Epstein, J. M. (2006). *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton: Princeton University Press.
- Epstein, J. M. and Axtell, R. (1996). *Growing Artificial Societies. Social Science from the Bottom Up*. Cambridge, MA: MIT press.
- Fagiolo, G., Moneta, A., and Windrum, P. (2007). A Critical Guide to Empirical Validation of Agent-Based Models in Economics: Methodologies, Procedures, and Open Problems. *Computational Economics*, 30:195–226.
- Fararo, T. J. (1969a). The Nature of Mathematical Sociology. *Social Research*, 36:75–92.
- Fararo, T. J. (1969b). Stochastic Processes. In Borgatta, E. F., editor, *Sociological Methodology*. San Francisco: Jossey-Bass.
- Fararo, T. J. and Butts, C. T. (1999). Advance in Generative Structuralism: Structured Agency and Multilevel Dynamics. *Journal of Mathematical Sociology*, 24:1–65.
- Fararo, T. J. and Kosaka, K. (1976). A Mathematical Analysis of Boudon’s IEO Model. *Social Science Information*, 15(2-3):431–475.

- Farmer, J. D. and Foley, D. (2009). The Economy Needs Agent-Based Modelling. *Nature*, 460:685–686.
- Ferber, J., Michel, F., and Baez, J. (2005). AGRE: Integrating Environments with Organizations. In Weyns, D., Parunak, V. D., and Michel, F., editors, *Environments for Multi-Agent Systems*, pages 44–56. Berlin: Springer.
- Fioretti, G. (2013). Agent-based Simulation Models in Organization Science. *Organizational Research Methods*, 16:227–242.
- Fountain, C. and Stovel, K. (2014). Turbulent Careers: Social Networks, Employer Hiring Preferences, and Job Instability. In Manzo, G., editor, *Analytical Sociology: Actions and Networks*, pages 339–370. Chichester: John Wiley & Sons.
- Freedman, D. A. (2005). Linear Statistical Models for Causation: A Critical Review. In Everitt, B. and Howell, D. C., editors, *Encyclopedia of Statistics in Behavioral Science*. Hoboken, NJ: John Wiley & Sons.
- Freedman, D. A. (2009). *Statistical Models: Theory and Practice*. Cambridge: Cambridge University Press.
- Freedman, D. A. (2010). *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. Cambridge University Press.
- Freedman, D. A. and Humphreys, P. (1996). The Grand Leap. *British Journal for the Philosophy of Science*, 47:113–23.
- Freedman, D. A. and Humphreys, P. (1999). Are There Algorithms That Discover Causal Structure? *Synthese*, 121:29–54.
- Freese, J. and Kevern, J. A. (2013). Types of Causes. In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 27–41. Dordrecht: Springer.
- Frias-Martinez, E., Williamson, G., and Frias-Martinez, V. (2011). An Agent-Based Model of Epidemic Spread Using Human Mobility and Social Network Information. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, pages 57–64.
- Frigg, R. and Reiss, J. (2009). The Philosophy of Simulation: Hot New Issues or Same Old Stew? *Synthese*, 169(3):593–613.
- Gabbriellini, S. and Torroni, P. (2014). Arguments in Social Networks. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-agent Systems (AAMAS’13)*, pages 119–1120.
- Gallegati, M. and Kirman, A. P., editors (1999). *Beyond the Representative Agent*. Aldershot & Lyme, NH: Edward Elgar.
- Gangl, M. (2010). Causal Inference in Sociological Research. *Annual Review of Sociology*, 36:21–47.
- Gangl, M. (2013). Partial Identification and Sensitivity Analysis. In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 377–402. Dordrecht: Springer.
- Gelman, A. and Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.
- Gerring, J. (2008). The Mechanismic Worldview: Thinking Inside the Box. *British Journal of Political Science*, 38(1):161–179.
- Ghorbani, A., Dijkema, G., and Schrauwen, N. (2015). Structuring Qualitative Data for Agent-Based Modelling. *Journal of Artificial Societies and Social Simulation*, 18(1):2.
- Gilbert, N. and Abbott, A., editors (2005). *Social Science Computation*. Special issue of *American Journal of Sociology*, 110(4).
- Gilbert, N. and Troitzsch, K. (2005). *Simulation for the Social Scientist*. Maidenhead: Open University Press, 2nd edition.
- Gintis, H. (2009). *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton: Princeton University Press.
- Gintis, H. (2013). Markov Models of Social Dynamics: Theory and Applications. *ACM Transactions on Intelligent Systems and Technology*, 4(3):53.
- Glennan, S. (1996). Mechanisms and the Nature of Causation. *Erkenntnis*, 44:49–71.
- Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science*, 69(3):S342–S353.
- Glymour, M. and Greenland, S. (2008). Causal Diagrams. In Rothman, K., Greenland, S., and Lash, T., editors, *Modern Epidemiology*, pages 183–209. Philadelphia, PA: Lippincott Williams & Wilkins.
- Goldthorpe, J. H. (2001). Causation, Statistics and Sociology. *European Sociological Review*, 17(1):1–20.
- Goldthorpe, J. H. (2016). *Sociology as Population Science*. Cambridge: Cambridge University

- Press. Forthcoming.
- Gonzales-Bailon, S. and Murphy, T. E. (2013). Social Interactions and Long-Term Fertility Dynamics. A Simulation Experiment in the Context of the French Fertility Decline. *Population Studies*, 67(2):135–155.
- Gorski, P. (2009). Social Mechanisms and Comparative-Historical Sociology: A Critical Realist Proposal. In Hedström, P. and Wittrock, B., editors, *Frontiers of Sociology*, pages 147–194. Leiden: Brill.
- Gould, R. V. (2002). The Origins of Status Hierarchies: A Formal Theory and Empirical Test. *American Journal of Sociology*, 107(5):1143–1178.
- Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 37(3):424–438.
- Grimm, V., Berger, U., Bastiansen, F., Eliassen, S., Ginot, V., Giske, J., Goss-Custard, J., Grand, T., Heinz, S. K., Huse, G., Huth, A., Jepsen, J. U., Jørgensen, C., Mooij, W. M., Müller, B., Peér, G., Piou, C., Railsback, S. F., Robbins, A. M., Robbins, M. M., Rossmanith, E., Rüger, N., Strand, E., Souissi, S., Stillman, R. A., Vabø, R., Visser, U., and DeAngelis, D. L. (2006). A Standard Protocol for Describing Individual-Based and Agent-Based Models. *Ecological Modelling*, 198(1-2):115–126.
- Gross, N. (2009). A Pragmatist Theory of Social Mechanisms. *American Sociological Review*, 74:358–379.
- Grüne-Yanoff, T. (2009a). The Explanatory Potential of Artificial Societies. *Synthese*, 169(3):539–555.
- Grüne-Yanoff, T. (2009b). Learning from Minimal Economic Models. *Erkenntnis*, 70(1):81–99.
- Guala, F. (2002). Models, Simulations, and Experiments. In Magnani, L. and Nersessian, N., editors, *Model-Based Reasoning: Science, Technology, Values*, pages 59–74. New York: Kluwer.
- Hägerstrand, T. (1965). A Montecarlo Approach to Diffusion. *European Journal of Sociology*, 6(1):43–67.
- Hägerstrand, T. (1970). What About People in Regional Science? *Papers of the Regional Science Association*, 24(1):6–21.
- Hall, N. (2004). Two Concepts of Causation. In Collins, J., Hall, N., and Paul, L. A., editors, *Causation and Counterfactuals*, pages 225–276. Cambridge, MA: MIT Press.
- Hansen, L. P. and Heckman, J. J. (1996). The Empirical Foundations of Calibration. *Journal of Economic Perspectives*, 10(1):87–104.
- Harrington, J. J. and Chang, M.-H. (2005). Co-evolution of Firms and Consumers and the Implications for Market Dominance. *Journal of Economic Dynamics and Control*, 29(1-2):245–276.
- Hassan, S., Pavón, J., Antunes, L., and Gilbert, N. (2010). Injecting Data into Agent-Based Simulation. In Takadama, K., Deffuant, G., and Cioffi-Revilla, C., editors, *Simulating Interacting Agents and Social Phenomena: The Second World Congress Springer, Tokyo (2010)*, volume 7 of *Springer Series on Agent Based Social Systems*, pages 179–191. Tokyo: Springer.
- Hauser, R. (1976). Review Essay. On Boudon’s Model of Social Mobility. *American Journal of Sociology*, 81(1):911–928.
- Hausman, D. (1992). *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Hayward, S. (2006). Agent-based Modelling with Wavelets and an Evolutionary Artificial Neural Network: Applications to CAC 40 Forecasting. In Chatterjee, E., Chakrabarti, E., and Bikas, K. R., editors, *Econophysics of Stock and Other Markets*, pages 163–174. Milan: Springer.
- Heckbert, S., Baynes, T., and Reeson, A. (2010). Agent-Based Modeling in Ecological Economics. *Annals of the New York Academy of Sciences*, 1185:39–53.
- Heckman, J. J. (2005). The Scientific Model of Causality. *Sociological Methodology*, 35(1):1–97.
- Hedström, P. (2005). *Dissecting the Social. On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
- Hedström, P. (2009). Studying Mechanisms to Strengthen Causal Inferences in Quantitative Research. In Box-Steffensmeier, J. M., Brady, H. E., and Collier, D., editors, *The Oxford Handbook of Political Methodology*, pages 319–335. Oxford: Oxford University Press.
- Hedström, P. and Bearman, P., editors (2009). *The Oxford Handbook of Analytical Sociology*. Oxford: Oxford University Press.
- Hedström, P. and Manzo, G., editors (2015). *Agent-Based Modeling: Advances and Challenges*. Special issue of *Sociological Methods and Research*, 44(2).
- Hedström, P. and Swedberg, R. (1998). Social Mechanisms: An Introductory Essay. In Hedström, P. and Swedberg, R., editors, *Social Mechanisms: An Analytical Approach to Social*

- Theory*, pages 1–31. Cambridge: Cambridge University Press.
- Hedström, P. and Ylikoski, P. (2010). Causal Mechanisms in the Social Sciences. *Annual Review of Sociology*, 36:49–67.
- Helbing, D. (2012). *Social Self-Organization. Agent-based Simulations and Experiments to Study Emergent Social Behavior*. Berlin: Springer.
- Hesslow, G. (1976). Discussion: Two Notes on the Probabilistic Approach to Causality. *Philosophy of Science*, 43:290–292.
- Hoffer, T., Greeley, A. M., and Coleman, J. S. (1985). Achievement Growth in Public and Catholic Schools. *Sociology of Education*, 58:74–97.
- 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 Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 331–352. Dordrecht: Springer.
- Hoover, K. (2008a). Does Macroeconomics Need Microfoundations? In Hausman, D., editor, *Philosophy of Economics*, pages 315–333. Cambridge: Cambridge University Press.
- Hoover, K. (2008b). Microfoundations and the Ontology of Macroeconomics. In Ross, D. and Kincaid, H., editors, *Oxford Handbook of Philosophy of Economics*, pages 386–409. Oxford: Oxford University Press.
- Hoover, K. (2012). Causal Structure and Hierarchies of Models. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43:778–786.
- Hummon, N. P. and Fararo, T. J. (1995a). Actors and Networks as Objects. *Social Networks*, 17:1–26.
- Hummon, N. P. and Fararo, T. J. (1995b). The Emergence of Computational Sociology. *Journal of Mathematical Sociology*, 20(2-3):79–87.
- Humphreys, P. (2009). The Philosophical Novelty of Computer Simulation Methods. *Synthese*, 169(3):615–626.
- 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: Cambridge University Press.
- Izquierdo, L. R., Izquierdo, S. S., Galán, J. M., and Santos, J. I. (2009). Techniques to Understand Computer Simulations: Markov Chain Analysis. *Journal of Artificial Societies and Social Simulation*, 12(1):6.
- Janssen, M. A. and Jager, W. (2001). Fashions, Habits and Changing Preferences: Simulation of Psychological Factors Affecting Market Dynamics. *Journal of Economic Psychology*, 22:745–772.
- Janssen, M. A. and Jager, W. (2003). Simulating Market Dynamics: Interactions between Consumer Psychology and Social Networks. *Artificial Life*, 9:343–356.
- Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., and Bühlmann, P. (2012). Causal Inference Using Graphical Models with the R Package pcalg. *Journal of Statistical Software*, 47(11).
- Kalter, F. and Kroneberg, C. (2014). Between Mechanism Talk and Mechanism Cult: New Emphases in Explanatory Sociology and Empirical Research. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66:S91–S115.
- Karlsson, G. (1958). *Social Mechanisms: Studies in Sociological Theory*. Stockholm: Free Press.
- Kirman, A. (1992). Whom or What Does the Representative Individual Represent? *Journal of Economic Perspectives*, 6(2):117–136.
- Kirman, A. (2010). The Economic Crisis is a Crisis for Economic Theory. *CESifo Economic Studies*, 56(4):498–535.
- Kistler, M. (2002). Causation in Contemporary Analytical Philosophy. In Esposito, C. and Porro, P., editors, *Quaestio. Annuario di Storia della Metafisica*, volume 2, pages 635–668. Turnhout Belgium: Brepols.
- Knight, C. R. and Winship, C. (2013). The Causal Implications of Mechanistic Thinking: Identification Using Directed Acyclic Graphs (DAGs). In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 275–299. Dordrecht: Springer.
- Korb, K. B. and Nicholson, A. E. (2011). *Bayesian Artificial Intelligence*. Boca Raton, FL: CRC Press, 2nd edition.
- Law, A. M. (2007). *Simulation Modeling and Analysis*. New York: McGraw-Hill.
- Leombruni, R. and Richiardi, M. (2005). Why are Economists Skeptical about Agent-Based

- Simulations? *Physica A*, 355:103–109.
- Lipton, P. (2004). *Inference to the Best Explanation*. London: Routledge, 2nd edition.
- Lipton, P. (2009). Causation and Explanation. In Beebe, H., Menzies, P., and Hitchcock, C., editors, *The Oxford Handbook of Causation*, pages 619–631. Oxford: Oxford University Press.
- Little, D. (2012). Analytical Sociology and The Rest of Sociology. *Sociologica*, 1:1–47.
- Lizardo, O. (2012). Analytical Sociologys Superfluous Revolution. *Sociologica*, 1:1–12.
- Longworth, F. (2006). Causation, Pluralism and Moral Responsibility. *Philosophica*, 77(1):45–68.
- Lucas, R. (1976). Econometric Policy Evaluation: A Critique. In Brunner, K. and Meltzer, A., editors, *The Phillips Curve and Labor Markets*, volume 1 of *Carnegie-Rochester Conference Series on Public Policy*, pages 19–46. Amsterdam: North-Holland.
- Lux, T. and Marchesi, M. (1999). Scaling and Criticality in a Stochastic Multi-Agent Model of a Financial Market. *Nature*, 397:498–500.
- Machamer, P., Darden, L., and Craver, C. (2000). Thinking about Mechanisms. *Philosophy of Science*, 67:1–25.
- Macy, M. and Flache, A. (2009). Social Dynamics from the Bottom-Up: Agent-based Models of Social Interaction. In *The Oxford Handbook of Analytical Sociology*, chapter 11. Oxford: Oxford University Press.
- Macy, M. and Sato, Y. (2008). Reply to Will and Hegselmann. *Journal of Artificial Societies and Social Simulation*, 11(4):11.
- Macy, M. W. and Willer, R. (2002). From Factors to Actors: Computational Sociology and Agent-Based Modeling. *Annual Review of Sociology*, 28:143–166.
- Magliocca, N. R., Brown, D. G., and Ellis, E. C. (2014). Cross-Site Comparison of Land-Use Decision-Making and Its Consequences across Land Systems with a Generalized Agent-Based Model. *PLoS ONE*, 9(1):e86179.
- Mahoney, J. and Goertz, G. and Ragin, C. C. (2013). Causal Models and Counterfactuals. In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 75–90. Dordrecht: Springer.
- Mahoney, J. (2000). Strategies of Causal Inference in Small-N Analysis. *Sociological Methods and Research*, 28(4):387–424.
- Mahoney, J. (2001). Beyond Correlational Analysis: Recent Innovations in Theory and Method. *Sociological Forum*, 16(3):575–593.
- Mahoney, J. (2008). Toward a Unified Theory of Causality. *Comparative Political Studies*, 41(4-5):412–436.
- Manski, C. F. (2003). *Partial Identification of Probability Distributions*. New York: Springer.
- Manski, C. F. (2007). *Identification for Prediction and Decision*. Cambridge: Harvard University Press.
- Manski, C. F. (2013). Identification of Treatment Response with Social Interactions. *Econometrics Journal*, 16(1):S1–S23.
- Manzo, G. (2007). Variables, Mechanisms, and Simulations: Can the Three Methods Be Synthesized? *Revue française de sociologie*, 48(5):35–71.
- Manzo, G. (2010). Analytical Sociology and its Critics. *European Journal of Sociology*, 51(1):129–170.
- Manzo, G. (2013). Educational Choices and Social Interactions: A Formal Model and a Computational Test. *Comparative Social Research*, 30:47–100.
- Manzo, G., editor (2014a). *Analytical Sociology: Actions and Networks*. Chichester: John Wiley & Sons.
- Manzo, G. (2014b). Data, Generative Models, and Mechanisms: More on the Principles of Analytical Sociology. In Manzo, G., editor, *Analytical Sociology: Actions and Networks*, chapter 1. Chichester: John Wiley & Sons.
- Manzo, G. and Baldassarri, D. (2015). Heuristics, Interactions, and Status Hierarchies: An Agent-based Model of Deference Exchange. *Sociological Methods and Research*, 44(3):329–387.
- Marini, M. and Singer, B. (1988). Causality in the Social Sciences. In Clogg, C., editor, *Sociological Methodology*, pages 347–409. San Francisco: Jossey-Bass.
- Marshall, B. D. and Galea, S. (2014). Formalizing the Role of Agent-Based Modeling in Causal Inference and Epidemiology. *American Journal of Epidemiology*, 181(2):92–99.
- Mäs, M. and Flache, A. (2013). Differentiation without Distancing. Explaining Bi-Polarization of Opinions without Negative Influence. *PLoS ONE*, 8(11):e74516.
- Mathieu, P., Beaufils, B., and Brandouy, O. (2005). *Artificial Economics: Agent-based Methods*

- in *Finance, Game Theory and Their Applications*, volume 564. Berlin: Springer Science & Business Media.
- Menzies, P. (2012). The Causal Structure of Mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43:796–805.
- Miller, J. H. and Page, S. E. (2004). The Standing Ovation Problem. *Complexity*, 9(5):8–16.
- Miller, J. H. and Page, S. E. (2007). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton, NJ: Princeton University Press.
- Moneta, A. and Russo, F. (2014). Causal Models and Evidential Pluralism in Econometrics. *Journal of Economic Methodology*, 21(1):54–76.
- Morgan, M. S. (2003). Experiments without Material Intervention: Model Experiments, Virtual Experiments and Virtually Experiments. In Radder, H., editor, *The Philosophy of Scientific Experimentation*, pages 217–235. Pittsburgh, PA: University of Pittsburgh Press.
- Morgan, M. S. (2012). *The World in the Model: How Economists Work and Think*. Cambridge University Press.
- Morgan, S. L. (2005). *On the Edge of Commitment: Educational Attainment and Race in the United States*. Stanford, CA: Stanford University Press.
- Morgan, S. L., editor (2013). *Handbook of Causal Analysis for Social Research*. Springer.
- Morgan, S. L. and Winship, C. (2014). *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge University Press, 2nd edition.
- Morrison, M. (2015). *Reconstructing Reality: Models, Mathematics, and Simulations*. Oxford University Press.
- Mouchart, M. and Russo, F. (2011). Causal Explanation: Recursive Decompositions and Mechanisms. In Illari, P., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 317–337. Oxford: Oxford University Press.
- Muldoon, R. (2007). Robust Simulations. *Philosophy of Science*, 74(5):873–883.
- Nikolai, C. and Madey, G. (2009). Tools of the Trade: A Survey of Various Agent Based Modeling Platforms. *Journal of Artificial Societies and Social Simulation*, 12(2):2.
- Opp, K. (2007). Peter Hedström: Dissecting the Social. On the Principles of Analytical Sociology. *European Sociological Review*, 23:115–122.
- Opp, K. (2013). What is Analytical Sociology? Strengths and Weaknesses of a New Sociological Research Program. *Social Science Information*, 52(3):329–360.
- Oreskes, N., Shrader-Frechette, K., and Belitz, K. (1994). Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. *Science*, 263(5147):641–646.
- O’Sullivan, D. (2008). Geographical Information Science: Agent-Based Models. *Progress in Human Geography*, 32:541–550.
- O’Sullivan, D. and Perry, G. L. W. (2013). *Spatial Simulation: Exploring Pattern and Process*. Chichester: Wiley.
- Page, S. E. (2008). Agent Based Models. In Blume, L. E. and Durlauf, S. N., editors, *The New Palgrave Dictionary of Economics*. Basingstoke, UK: Palgrave Macmillan, 2nd edition.
- Parker, J. and Epstein, J. (2011). A Distributed Platform for Global-Scale Agent-Based Models of Disease Transmission. *ACM Trans. Model Comput Simul.*, 22(1):2–33.
- Parker, W. (2008). Computer Simulation Through an Error-Statistical Lens. *Synthese*, 163:371–384.
- Parker, W. (2009). Does Matter Really Matter? Computer Simulations, Experiments, and Materiality. *Synthese*, 169:483–496.
- Pawson, R. (1989). *A Measure for Measures: a Manifesto for Empirical Sociology*. London: Routledge.
- Pearl, J. (1993). Comment: Graphical Models, Causality, and Interventions. *Statistical Science*, 8(3):266–269.
- Pearl, J. (1995). Causal Diagrams for Empirical Research. *Biometrika*, 82(4):669–710.
- Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press, 2nd edition.
- Pearl, J. (2010). Causal Inference. In Guyon, I., Janzing, D., and Schölkopf, B., editors, *Causality: Objectives and Assessment (NIPS 2008)*, volume 6, pages 39–58.
- Pearl, J. (2011). The Structural Theory of Causation. In Illari, P., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 697–727. Oxford: Oxford University Press.
- Pinyol, I. and Sabater-Mir, J. (2013). Computational Trust and Reputation Models for Open Multi-agent Systems: a Review. *Artificial Intelligence Review*, 40(1):1–25.
- Psillos, S. (2007). What is Causation? In Choksi, B. and Natarajan, C., editors, *The Episteme Reviews: Research trends in Science, Technology and Mathematics Education*, pages 11–34. Bangalore: Macmillan India.

- Railsback, S. F. and Grimm, V. (2011). *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton: Princeton University Press.
- Reichenbach, H. (1956). *The Direction of Time*. Berkeley and Los Angeles: University of California Press, 1971 edition.
- Reiss, J. (2009). Causation in the Social Sciences. Evidence, Inference, and Purpose. *Philosophy of the Social Sciences*, 39(1):20–40.
- Reiss, J. (2011a). A Plea for (Good) Simulations: Nudging Economics Toward an Experimental Science. *Simulation & Gaming*, 42(2):243–264.
- Reiss, J. (2011b). Third Time's a Charm: Causation, Science and Wittgensteinian Pluralism. In Illari, P., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 907–927. Oxford: Oxford University Press.
- Reiss, J. (2012). Causation in the Sciences: An Inferentialist Account. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(4):769–777.
- Reiss, J. (2013). *Philosophy of Economics. A Contemporary Introduction*. New York: Routledge.
- Reiss, J. (2015). A Pragmatist Theory of Evidence. *Philosophy of Science*, 82:341–362.
- Rolfe, M. (2014). Social Networks and Agent-Based Models. In Manzo, G., editor, *Analytical Sociology: Actions and Networks*, pages 237–260. Chichester: John Wiley & Sons.
- Rosenzweig, M. R. and Wolpin, K. I. (2000). Natural Experiments in Economics. *Journal of Economic Literature*, 38:827–874.
- Rubin (1980). Comment on 'Randomization Analysis of Experimental Data in the Fisher Randomization Test' by Basu. *Journal of the American Statistical Association*, 75:591–593.
- Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688–701.
- Rubin, D. B. (1986). Which Ifs Have Causal Answers (Comment on 'Statistics and Causal Inference' by Paul W. Holland). *Journal of the American Statistical Association*, 81:961–962.
- Russo, F. and Williamson, J. (2007). Interpreting Causality in the Health Sciences. *International Studies in the Philosophy of Science*, 21(2):157–170.
- Sakoda, J. M. (1971). The Checkerboard Model of Social Interaction. *Journal of Mathematical Sociology*, 1(1):119–132.
- Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton, NJ: Princeton University Press.
- Saltelli, A. (2000). What Is Sensitivity Analysis? In Saltelli, A., Chan, K., and Scott, E., editors, *Sensitivity Analysis*, pages 3–13. Chichester: Wiley.
- Saltelli, A., Chan, K., and Scott, E., editors (2000). *Sensitivity Analysis*. Chichester: Wiley.
- Sampson, R. J. (2011). Neighborhood Effects, Causal Mechanisms and the Social Structure of the City. In Demeulenaere, P., editor, *Analytical Sociology and Social Mechanisms*, pages 227–249. Cambridge: Cambridge University Press.
- Sargent, T. (1993). *Bounded Rationality in Macroeconomics*. Oxford University Press.
- Sawyer, R. K. (2003). Artificial Societies: Multiagent Systems and the Micro-Macro Link in Sociological Theory. *Sociological Methods and Research*, 31:325–363.
- Sawyer, R. K. (2004). Social Explanation and Computational Simulation. *Philosophical Explorations*, 7(3):219–231.
- Sawyer, R. K. (2011). Conversation as Mechanism: Emergence in Creative Groups. In Demeulenaere, P., editor, *Analytical Sociology and Social Mechanisms*, pages 78–98. Cambridge: Cambridge University Press.
- Schelling, T. C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1:143–186.
- Schupbach, J. N. (2015). Robustness, Diversity of Evidence, and Probabilistic Independence. In Mäki, U., Votsis, I., Ruphy, S., and Schurz, G., editors, *Recent Developments in the Philosophy of Science: EPSA13 Helsinki*, pages 305–316. Heidelberg: Springer.
- Shalizi, C. R. (2006). Methods and Techniques in Complex Systems Science: An Overview. In Deisboeck, T. S. and Kresh, J. Y., editors, *Complex Systems Science in Biomedicine*, pages 33–114. New York: Springer.
- Shoham, Y. and Leyton-Brown, K. (2009). *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge: Cambridge University Press.
- Silverman, E., Bijak, J., Hilton, J., Cao, V. D., and Noble, J. (2013). When Demography Met Social Simulation: A Tale of Two Modelling Approaches. *Journal of Artificial Societies and Social Simulation*, 16(4):9.
- Sims, C. (1980). Macroeconomics and Reality. *Econometrica*, 48(1):1–48.
- Smith, E. R. and Conrey, F. R. (2007). Agent-Based Modeling: A New Approach for Theory Building in Social Psychology. *Personality and Social Psychology Review*, 11(1):87–104.

- Snijders, T. A. B. and Steglich, C. E. (2015). Representing Micro-Macro Linkages by Actor-Based Dynamic Network Models. *Sociological Methods and Research*, 44:222–271.
- 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.
- Sobkowicz, P. (2009). Modelling Opinion Formation with Physics Tools: Call for Closer Link with Reality. *Journal of Artificial Societies and Social Simulation*, 12(1):11.
- Sørensen, A. B. (1976). Models and Strategies in Research on Attainment and Opportunity. *Social Science Information*, 15(1):71–91.
- Spirtes, P. (2010). Introduction to Causal Inference. *Journal of Machine Learning Research*, 11:1643–1662.
- Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. Cambridge MA: MIT Press, 2nd edition.
- Spirtes, P. and Glymour, C. and Scheines, R. (1997). Reply to Humphreys and Freedman’s Review of Causation, Prediction, and Search. *British Journal for the Philosophy of Science*, 48(4):555–568.
- Squazzoni, F. (2012). *Agent-Based Computational Sociology*. Chicester: Wiley.
- Steel, D. (2004). Social Mechanisms and Causal Inference. *Philosophy of the Social Sciences*, 34(1):55–78.
- Steel, D. (2007). *Across the Boundaries: Extrapolation in Biology and Social Science*. Oxford: Oxford University Press.
- Steiger, D. and Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, 65(3):557–586.
- Stock (2001). Instrumental Variables in Statistics and Econometrics. In Smelser, N. and JaB, P., editors, *International Encyclopedia of the Social and Behavioral Sciences*, pages 7577–7582. Kidlington: Elsevier Science Ltd.
- Stock, J. H. and Watson, M. W. (2010). *Introduction to Econometrics*. Boston: Addison Wesley, 3rd edition.
- Stock, J. H. and Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In K., A. D. W., editor, *Identification and Inference for Econometric Models*, pages 80–108. New York: Cambridge University Press.
- Stock, J. H. and Wright, J. H. and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, 20(4):518–529.
- Stonedahl, F. and Wilensky, U. (2010). Evolutionary Robustness Checking in the Artificial Anasazi Model. In *Proceedings of the 2010 AAAI Fall Symposium on Complex Adaptive Systems*. Arlington, VA. Archived at <http://www.webcitation.org/6H9Rdec9p>.
- Tesfatsion, L. (2002). Agent-based Computational Economics: Growing Economies From the Bottom Up. *Artificial Life*, 8(1):55–82.
- Tesfatsion, L. (2006). Agent-Based Computational Economics: A Constructive Approach to Economic Theory. In Tesfatsion, L. and Judd, K. L., editors, *Handbook of Computational Economics. Agent-Based Computational Economics*, volume 2, pages 831–880. North Holland: Elsevier.
- Thiele, J. C., Kurth, W., and Grimm, V. (2014). Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models: A Cookbook Using NetLogo and R. *Journal of Artificial Societies and Social Simulation*, 17(3):11.
- Thorne, B. C., Bailey, A. M., and Peirce, S. M. (2007). Combining Experiments with Multi-Cell Agent-Based Modeling to Study Biological Tissue Patterning. *Brief Bioinform.*, 8(4):245–257.
- Thorngate, W. and Edmonds, B. (2013). Measuring Simulation-Observation Fit: An Introduction to Ordinal Pattern Analysis. *Journal of Artificial Societies and Social Simulation*, 16(2):14.
- Todd, P. M., Billari, F. C., and Simao, J. (2005). Aggregate Age-at-marriage Patterns From Individual Mate-search Heuristics. *Demography*, 42:5559–5574.
- Van de Rijt, A., Siegel, D., and Macy, M. (2009). Neighborhood Chance and Neighborhood Change: A Comment on Bruch and Mare. *American Journal of Sociology*, 114:1166–1180.
- VanderWeele, T. J. (2011). Sensitivity Analysis for Contagion Effects in Social Networks. *Sociological Methods and Research*, 40(2):240–255.
- VanderWeele, T. J. and An, W. (2013). Social Networks and Causal Inference. In Morgan, S. L., editor, *Handbook of Causal Analysis for Social Research*, pages 353–374. Dordrecht: Springer.
- Varenne, F. (2009). Models and Simulations in the Historical Emergence of the Science of

- Complexity. In Aziz-Alaoui, M. A. and Bertelle, C., editors, *System Complexity to Emergent Properties*, pages 3–21. Springer.
- Wang, J., Zhang, L., Jing, C., Ye, G., Wu, H., Miao, H., Wu, Y., and Zhou, X. (2013). Multi-scale Agent-based Modeling on Melanoma and Its Related Angiogenesis Analysis. *Theoretical Biology & Medical Modelling*, 10(41).
- Wang, X. and Sobel, M. E. (2013). New Perspectives on Causal Mediation Analysis. In Morgan, S., editor, *Handbook of Causal Analysis for Social Research*, pages 215–242. Dordrecht: Springer.
- Watts, D. J. (2014). Common Sense and Sociological Explanations. *American Journal of Sociology*, 120(2):313–351.
- Weisberg, M. (2006). Robustness Analysis. *Philosophy of Science*, 73:730–742.
- Willer, D. and Walker, H. A. (2007). *Building Experiments: Testing Social Theory*. Stanford: Stanford University Press.
- Williamson, J. (2006). Causal Pluralism versus Epistemic Causality. *Philosophica*, 77(1):69–96.
- Williamson, J. (2013). How Can Causal Explanations Explain? *Erkenntnis*, 78(2):257–275.
- Williamson, J. (2015). Deliberation, Judgement and the Nature of Evidence. *Economics and Philosophy*, 31:27–65.
- Winsberg, E. (2003). Simulated Experiments: Methodology for a Virtual World. *Philosophy of Science*, 70:105–125.
- Winship, C. (2009). Time and Scheduling. In Hedström, P. and Bearman, P., editors, *The Oxford Handbook of Analytic Sociology*, pages 498–520. Oxford: Oxford University Press.
- Winship, C. and Sobel, M. (2004). Causal Inference in Sociological Studies. In Hardy, M. and Bryman, A., editors, *A Handbook of Data Analysis*, pages 480–504. London: Sage Publications.
- Woodward, J. (2002). What is a Mechanism? A Counterfactual Account. *Philosophy of Science*, 69:S366–S377.
- Woodward, J. (2003). *Making Things Happen. A Theory of Causal Explanation*. New York: Oxford University Press.
- Woodward, J. (2013). Mechanistic Explanation: Its Scope and Limits. *Proceedings of the Aristotelian Society*, 87(1):39–65.
- Wooldridge, M. (2000). *Reasoning about Rational Agents*. Cambridge, MA: MIT Press.
- Wooldridge, M. (2009). *An Introduction to MultiAgent Systems*. Chichester: John Wiley & Sons, 2nd edition.
- Wunder, M., Suri, S., and Watts, D. J. (2013). Empirical Agent Based Models of Cooperation in Public Goods Games. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC ’13, pages 891–908. New York: ACM.
- Wurzer, G., Kowarik, K., and Reschreiter, H., editors (2015). *Agent-Based Modeling and Simulation in Archaeology*. Berlin: Springer.
- Xia, H., Wang, H., and Z., X. (2011). Opinion Dynamics: A Multidisciplinary Review and Perspective on Future Research. *International Journal of Knowledge and Systems Science*, 2(4):72–91.
- Ylikoski, P. and Aydinonat, N. E. (2014). Understanding with Theoretical Models. *Journal of Economic Methodology*, 21(1):19–36.
- Young, P. (2006). Social Dynamics: Theory and Applications. In Judd, K. and Tesfatsion, L., editors, *Handbook of Computational Economics*, volume II, chapter 22. Amsterdam: North Holland.
- Zhang, L., Chen, L., and Deisboeck, T. (2009). Multiscale, Multi-resolution Brain Cancer Modeling. *Mathematics and Computers in Simulation*, 79(7):2021–2035.