Common Sense and Sociological Explanations[1]

Duncan J. Watts
Microsoft Research

Sociologists have long advocated a sociological approach to explanation by contrasting it with common sense. The argument of this article, however, is that sociologists rely on common sense more than they realize. Moreover, this unacknowledged reliance causes serious problems for their explanations of social action, that is, for why people do what they do. Many such explanations, it is argued, conflate understandability with causality in ways that are not valid by the standards of scientific explanation. It follows that if sociologists want their explanations to be scientifically valid, they must evaluate them specifically on those grounds—in particular, by forcing them to make predictions. In becoming more scientific, however, it is predicted that sociologists' explanations will also become less satisfying from an intuitive, sense-making perspective. Even as novel sources of data and improved methods open exciting new directions for sociological research, therefore, sociologists will increasingly have to choose between unsatisfying scientific explanations and satisfying but unscientific stories.

INTRODUCTION

Among sociologists common sense has long been the Rodney Dangerfield of epistemologies—it gets no respect. Although definitions of common sense vary (Taylor 1947; Geertz 1975; Rosenfeld 2011), it is generally associated with the practical knowledge of ordinary people, deployed in everyday situations and, as such, is distinct from the style of theoretical knowledge to

[1] The author acknowledges Matthew Salganik and three *AJS* reviewers for extensive and helpful feedback on earlier drafts of this article. Direct correspondence to Duncan Watts, Microsoft Research NYC, 641 Avenue of the Americas, Seventh Floor, New York, New York 10011. E-mail: duncan@microsoft.com

0002-9602/2015/12002-0001$10.00

which sociologists aspire. Correspondingly, although a handful of sociologists have adopted an admiring view of common sense (Taylor 1947; Mathisen 1989), the prevailing opinion has been critical (Stouffer 1947; Lazarsfeld 1949; Merton 1968; Manis 1972; Black 1979; Boudon 1988; Rosenfeld 2011). Sociologists, for example, are fond of pointing out that much of what passes for common sense is inconsistent and even contradictory.[2] More generally, sociologists have also pointed out that what common sense treats as "facts"—self-evident, unadorned descriptions of an objective reality—often disguises value judgments that depend on the subjective experience of the person making the evaluation as well as on the (supposedly objective) nature of the thing being evaluated (Geertz 1975; Black 1979). Yet precisely because these subjective value judgments are treated as objective—indeed self-evidently true—they are never themselves subject to scrutiny. By surfacing, and critically examining, the unstated assumptions that underpin commonsense statements about reality, sociological thinking is often held up as the antidote to commonsense reasoning (Becker 1998).

The point of this essay is to argue that sociologists rely on common sense much more than they realize. Nor do I mean just that, as Black (1979) and others (Merton 1968; Manis 1972) have argued, sociologists are prone to treating the subjective opinions of survey respondents and other research subjects as objective statements of reality and hence making the same fact-value error as nonsociologists. Rather what I want to argue is that commonsense reasoning pervades sociological theorizing in a fundamental way—at the level of sociological theories of action. In particular, I argue that many such theories—most prominently rational choice theory, but also the many variants of individualism that have pervaded sociology over the past century, as well as more recent additions such as Bourdieu's (2005) field theory and Gross's (2009) pragmatist theory—are all descendants of the same ancestor "theory," which for the sake of argument I will call *rationalizable action*—the claim that actions, whether individual or collective, can be explained in terms of the intentions, beliefs, circumstances, and opportunities of the actors involved.

To be clear, this "ancestor theory" is not itself a sociological theory, but rather is a commonsense or folk theory, deployed in practice by ordinary people in everyday circumstances both to anticipate the behavior of others and also to make sense of behavior they have observed. Nevertheless, I will argue that many sociological theories of action, when stripped of their various formalizations and obfuscations, are effectively versions of rationaliz-

[2] For example, birds of a feather flock together, but opposites also attract. Whereas a sociologist might want to know under what circumstances, or to what extent, the former applied vis-à-vis the latter, common sense simply makes both assertions and leaves it at that. See Mathisen (1989) for many more examples of critiques of common sense in introductory sociology texts.

able action and that even when sociologists advocate theories of action that do not explicitly invoke rationalizable action, they nevertheless fall back upon it when trying to make their findings understandable. Moreover, I will argue that the pervasiveness of rationalizable action among sociological explanations pertains not in spite of its commonsense origins but precisely because of them. In their everyday lives, that is, sociologists necessarily utilize commonsense concepts, and, because these concepts do indeed appear valid on the basis of everyday personal experience, they seem self-evident. As a result, assumptions implicit in the folk theory get incorporated into sociological theories without even being explicitly articulated; thus, even when the theories themselves become contested, the commonsense assumptions on which they rest are left unchallenged (James 1909; Boudon 1988).

If these assumptions were valid, then there would be no problem. But as I will argue, although rationalizability is useful both for making sense of and also anticipating behavior under everyday conditions—namely, in immediate circumstances with immediate feedback—it fails badly under the conditions where it is deployed by sociologists and for the purposes for which they deploy it. More specifically, I argue that explanations that are invoked to make sense of observed behavior in terms of actors' intentions, habits, beliefs, opportunities, circumstances, and so on cannot in general be expected to satisfy the standards of causal explanation; hence, they will generalize poorly to novel scenarios. Correspondingly, I argue that if sociologists want their explanations to be causal, they must place less emphasis on understandability (i.e., sense making) and more on their ability to make predictions. Although sociologists are in disagreement on the relationship between prediction and causality, I maintain that much of this disagreement arises from overly narrow interpretations of prediction. Interpreted appropriately, I maintain that prediction is consistent with a view of causality that is almost universally accepted by sociologists and is therefore a necessary (but not sufficient) condition for an explanation to qualify as scientific. I conclude, however, that as sociologists increasingly evaluate their explanations in terms of scientific validity rather than understandability, these explanations will inevitably become less satisfying; moreover, it will become clear that many questions of interest to sociologists are not directly answerable.

CAUSALITY VERSUS UNDERSTANDABILITY IN EXPLANATION

Over the past century sociologists have proposed many theories of social action—so many theories, in fact, that characterizing them all, or even documenting them exhaustively, would be a daunting task in itself. Without speaking for all theories of social action, therefore, I shall nevertheless

attempt to defend the narrower claim that many of these theories—including rational choice theory (Hedström and Stern 2008) along with the many related strains of individualism (Lukes 1968; Mayhew 1980; Boudon 1987), structural-functionalism (Parsons and Shils 1951), field theory (Bourdieu 2005), and pragmatist theories of action (Whitford 2002; Gross 2009)—are all variants of what I am calling rationalizable action. More generally, I argue that even where sociological theories do not explicitly invoke any version of rationalizable action—as is the case for certain nomological theories—because their findings are generally interpreted with reference to individual psychological states, they do depend on it in practice.

To reiterate, by rationalizable action I mean the claim that individual or collective action can be explained in terms of the intentions, beliefs, circumstances, and opportunities of the actors involved. By design, this definition is extremely broad. For example, it is not restricted to any particular intentions, such as rational instrumental goals, or to beliefs held for rational reasons, or even to beliefs held consciously. Likewise, by circumstances and opportunities, it includes everything from tangible resources like time, money, and human capital to less tangible resources such as social and cultural capital, and even acquired habits of mind or local cultural norms. And finally, its notion of "actors" includes not only individual humans in the strict methodological individualist tradition but also representative individuals, aka "social actors," such as families, firms, political parties, and even nation-states. Any behavior that can be rationalized, in other words, falls under the scope of rationalizable action.

As stated, however, rationalizable action is ambiguous with respect to the meaning of its critical claim to "explain" action. How should this claim be interpreted? Overwhelmingly sociologists—and social scientists in general—are of the view that explanations are inherently causal in nature (Marini and Singer 1988; Hedström 2005; Woodward 2005; Cartwright 2006; Manski 2007; Morgan and Winship 2007; Sloman 2009; Gerber and Green 2012). So tight is the relationship between explanation and causality in social science, in fact, that the terms "explanation" and "causal explanation" are often used interchangeably and without elaboration.[3] In ad-

[3] In an exception, Woodward (2005, p. 4) acknowledges that "explanation" is used in many ways, but he quickly restricts his discussion to causal explanations in part on the grounds that they are of most interest to social science. More typical is Freedman (1991), who refers simply to "causal arguments" without considering any other kind, or Sloman (2009), who notes simply that explanation depends "crucially" on causal understanding and then goes on to treat all explanations as causal explanations. Morgan and Winship (2007), Manski (2007), and others also effectively assume that explanations are causal explanations, and then devote their attention to the proper methods for inferring causality. Noncausal explanations are not considered except as examples of sloppy methodology.

dition, many social scientists and philosophers subscribe to the notion that causal explanations must also be predictive (Lakatos 1980; Marini and Singer 1988; Freedman 1991; Babyak 2004; Manski 2007; Gerber and Green 2012; Schrodt 2013) in the sense that had the causal mechanisms identified by the explanation been known ex ante, that knowledge could have been used to predict (in some sense) the known outcome. This view that prediction is essential to causal explanation was famously articulated by Hempel and Oppenheim (1948, p. 138), who argued that "an explanation is not fully adequate unless its explanans, if taken account of in time, could have served as a basis for predicting the phenomenon under consideration." Indeed, Hempel and Oppenheim went on to emphasize that "it is this potential predictive force which gives scientific explanation its importance: only to the extent that we are able to explain empirical facts can we attain the major objective of scientific research, namely not merely to record the phenomena of our experience, but to learn from them, by basing upon them theoretical generalizations that enable us to anticipate new occurrences and to control, at least to some extent, the changes in our environment."[4]

Sociologists and philosophers have subsequently disputed numerous elements of Hempel and Oppenheim's framework (Lakatos 1980; Ferguson 2000; Woodward 2005; Morgan and Winship 2007; Hedström and Ylikoski 2010), in particular their insistence on covering laws, but also to some extent their insistence on prediction, arguing, for example, that explanations can build causal understanding even when not predictive (Lieberson and Lynn 2002) or that an emphasis on prediction can lead to the rejection of perfectly valid causal explanations (Hedström and Ylikoski 2010). Later I shall argue that much of this disagreement arises from differing interpretations of the term "prediction" and that under the broadest such interpretation most sociologists would agree that causal explanations can and should be evaluated in terms of the predictions they make—just as Hempel and Oppenheim argued. For now, however, it is sufficient to assert only that when sociologists claim to have explained something, the clear implication—if not

[4] Lakatos (1980) also emphasized that what makes a scientific research program "scientific" is its ability to make "new and surprising predictions" that are subsequently borne out. Lakatos specifically contrasts his notion of scientific theories with theories—such as Marxism and Freudianism—that "explain" events only after the fact. Later still, Marini and Singer (1988, p. 347) again emphasized the relation between "the identification of genuine causes" and prediction, noting that "with causal knowledge it is often possible to predict events in the future or new observations and to exercise some measure of control over events. It is knowledge of causes that makes intervention for the production of desired effects possible" (pp. 347–48). Most recently, Schrodt (2013) argued that without a prediction criterion, there is no way to separate a social scientific explanation from mythological explanations of the form used by the ancients to "explain" thunder and lightning in terms of the personal proclivities of Thor.

the explicit assertion—is that it satisfies a somewhat weaker "manipulationist" criterion for causality, meaning that it answers what Woodward (2003, p. 11) calls "a *what-if-things-had-been-different question*: the explanation must enable us to see what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways."[5] Woodward goes further, in fact, to state that if a proposed explanation does not pass this manipulationist test then it is not an explanation at all but is mere description or storytelling (p. 5)—essentially the distinction that Hempel and Oppenheim make for prediction.

What I want to claim next, however, is that although in theory sociologists almost universally associate explanation with causality in Woodward's manipulationsist sense, in practice they frequently invoke a subtly but critically different notion of "explanation," to which Hempel and Oppenheim refer as an "empathetic explanation," meaning "the reduction of something unfamiliar to ideas or experiences familiar to us" (1948, p. 145). Although empathetic explanations are often presented in the form of causal explanations—for example, they may claim to identify causal mechanisms and even speculate about counterfactuals—they are different in that they are evaluated primarily, if not exclusively, in terms of their ability to make sense of some observed behavior or outcome—that is, to render it understandable—by reducing it to intuitive, interpretable claims about human motives, incentives, and opportunities.

Hempel and Oppenheim argued that understanding of the empathetic variety was neither necessary nor sufficient for scientific validity. Nevertheless, understandability has long been associated with causality in sociology, dating at least to Weber's notion of *verstehen* in which an explanation refers to the outcome of some process of internal reflection on the part of the analyst whose job it is to "make sense of" or interpret the observed behavior. Weber, it should be noted, did not specify the precise relationship between causation and understandability, stating only that sociology "is a science concerning itself with the interpretive understanding of social action and thereby with a causal explanation of its course and consequences" (Weber 1968, p. 4). It is unclear from this statement whether Weber believed the interpretive understanding *is* the causal ex-

[5] For example, although Hedström and Ylikoski (2010) reject Hempel's emphasis on law-like statements and strict equivalence of prediction and explanation, they effectively preserve the core requirement that an explanation is not scientific unless it can be used to make reliable, generalizable inferences about cause and effect—in other words, inferences that, had they been known at the time, could have been used to affect the outcome. Kiser and Hecter (1998, p. 790), who describe themselves as realists, strike a similar stance with respect to rational choice theory, arguing that "testing each mechanism's unique empirical implications" is the most useful criterion for deciding among rival explanatory mechanisms. Even historians, who often seek to distance themselves from social science, espouse a similar view of causality in their explanations (Gaddis 2002).

planation, or merely a necessary but insufficient condition; but if the latter, he did not specify what else is required.[6]

The philosopher Donald Davidson (1963, p. 685), however, was explicit, arguing that "rationalization is a species of ordinary causal explanation," in other words, that the reasons that agents give for their actions are the causes of those actions.[7] Interestingly, Davidson did not derive this conclusion from any more fundamental logic or empirical evidence so much as he asserted it to be self-evident—common sense in fact—arguing that since we understand causal explanation, then it follows that if certain reasons are chosen by an agent to make sense of his or her own actions then those reasons are self-evidently causal.[8] Davidson, it should be noted, was referring to rationalizations of quotidian everyday behavior, such as turning on a light switch; but as I shall argue next, the same conflation of scientific and empathetic explanations is common among sociological theories of action as well: an explanation is defended on the grounds of its understandability—in other words, that it is interpretable or that it makes sense[9]—but the reader is encouraged, either explicitly or implicitly, to conclude that the mechanism described is also what caused the outcome.

The Case of Rational Choice Theory

A particularly striking illustration of this unacknowledged epistemological conflation (although as I will argue subsequently, far from the only such illustration) is presented by the case of rational choice theory, which when introduced into sociology and political science from economics in the late 1960s, had clear scientific aspirations. In an early paper, for example, the future Nobel laureate John Harsanyi (1969, p. 514) argued that "theories of rational behavior have a natural tendency to take a hypothetico-deductive form, and to explain a wide variety of empirical facts in terms of a small number of theoretical assumptions (such as assumptions about the actual objectives of people's behavior, about the resources and the informa-

[6] Woodward (2003) is somewhat more explicit, acknowledging that the requirement for "epistemologically accessible" (i.e., understandable) explanations imposes a theoretical constraint on their possible content; however, as with Weber, he does not discuss how or even if this constraint is problematic.

[7] More precisely, Davidson specified "primary reasons" as the causes of actions, where a primary reason is defined as (*a*) a desire for a goal G and (*b*) a belief that action A is a means to attain G.

[8] As Malpas (1996) describes Davidson's reasoning: "Indeed, where an agent has a number of reasons for acting and yet acts on the basis of one reason in particular, there is no way to pick out just that reason on which the agent acts other than by saying that it is the reason that *caused* her action." See also Kalugin (2006).

[9] To clarify, I use the terms "interpretable," "understandable," and "making sense" interchangeably, reflecting their usage elsewhere.

tion available to them, etc.)." Likewise, many other early instantiations of rational choice theory also adhered closely to the criteria of scientific explanation, making strong causal claims in the form of analytically precise models that often yielded sharp, testable predictions (Becker 1976; Coleman and Fararo 1992).

As has been documented exhaustively elsewhere,[10] these early models were heavily criticized by sociologists (Quadagno and Knapp 1992; Elster 1993; Boudon 1998; Somers 1998; Whitford 2002; Elster 2009), psychologists (Tversky and Kahneman 1974; Dawes 2002; Gilovich, Griffin, and Kahneman 2002; Ariely, Loewenstein, and Prelec 2003; Sunstein 2003), political scientists (Green and Shapiro 1994; Walt 1999; Green and Shapiro 2005), and even some economists (Arrow 1986; Conlisk 1996; McFadden 1999) on the grounds that they relied on implausible or empirically invalid assumptions about the preferences, knowledge, and computational capabilities of the actors in question or, alternatively, that they yielded predictions that were also demonstrably at odds with empirical evidence. Partly in response to these criticisms, and partly as a natural consequence of its expansion beyond early market-centered applications, the dominant interpretation of rational choice among sociologists adapted over time to accommodate noneconomic and even prosocial purposes, limited knowledge about future states, and bounded powers of reasoning (Wippler and Lindenberg 1987; Goldthorpe 1998). More profoundly, the very concept of utility maximization, which early rational choice theorists like Becker (1976) and Coleman and Fararo (1992) regarded as the defining characteristic of rational choice, has also gradually been set aside (Kiser and Hechter 1998), as has the weaker assumption of exogenous, stable, and consistent preferences (Foley 2004; Hedström and Stern 2008)—also once regarded as a defining characteristic of rationality (Gintis 2009). Even the requirement, regarded by many as the "hard core" of rationality, that action be forward looking and purposive, as distinct from mere habit or some other unthinking reflex (Kiser and Hechter 1998; Cox 1999), has been questioned by some rational choice theorists for whom backward-looking behavior is permissible (Macy 1993; Macy and Flache 2002). Finally, and arguably most strikingly of all developments, the early commitment by

[10] Within sociology, the debate over rational choice theory has played out over the past 20 years, beginning with an early volume (Coleman and Fararo 1992), in which perspectives from both sides of the debate are represented, and continued in journals like *American Journal of Sociology* (Boudon 1998; Kiser and Hechter 1998; Somers 1998), and *Sociological Methods and Research* (Quadagno and Knapp 1992). Over the same period, a similar debate has also played out in political science, sparked by the publication of Green and Shapiro's (1994) polemic, "Pathologies of Rational Choice Theory." See Friedman (1996) for the responses of a number of rational choice advocates to Green and Shapiro's critique, along with Green and Shapiro's responses to the responses. Other interesting commentaries are by Elster (1993), Goldthorpe (1998), and Whitford (2002).

rational choice theorists to formulate a "hypothetico-deductive" theory of any kind has been effectively abandoned in favor of something rather less precise, like an "approach" or a "paradigm" (Farmer 1992; Kiser and Hechter 1998; Cox 1999) intended more to orient the analyst than to specify any particular rule or set of rules to be followed by the actor.

Whether these changes have increased the empirical relevance of rational choice theory, as advocates have attested (Kiser and Hechter 1998), or have instead created "an ever-expanding tent in which to house every plausible proposition advanced by anthropology, sociology, or social psychology" as critics have alleged (Green and Shapiro 2005, p. 76), is a topic that has been debated endlessly without resolution and that I shall not attempt to resolve either. Rather, I want make the related observation that in the course of defending both its theoretical plausibility and empirical validity, advocates of rational choice theory have effectively shifted from a scientific to an empathetic view of explanation. Even as rational choice theorists have continued to express explicitly scientific aspirations, that is—occasionally drawing analogies with Newtonian mechanics (Farmer 1992; Diermeier 1996; Cox 1999)—the standards of evidence by which they evaluate their explanations have shifted over time from an emphasis on prediction and deduction to an emphasis on understandability and sense making.

This effective conflation of scientific (i.e., causal) and empathetic explanation is illustrated nicely by Farmer (1992), who argues that "we need to hold that actors are purposive (rather as physicists hold that energy is always conserved) in order to apply theories that on the basis of more specific claims about actors' purposes, knowledge, internalized and external constraints, and so on . . . provide explanatory accounts of the processes generating particular social outcomes." At first glance this looks like a standard scientific motivation for rational action: rationality is a quantity, like energy, that exists independently of the observer and can be used to make predictions about the objects (actors) under observation. But whereas the principle of conservation of energy is indeed used by physicists to make predictions about the motion of pendulums and the like, the principle of rationality, according to Farmer, does not need to pass any test of predictive validity. Rather, quoting Buchanan (1985, p. 417), she concludes that "the whole utility maximizing apparatus assumes meaning only in some reconstructive and explanatory sense." In other words, although rational choice explanations are explicitly intended to rest on general causal mechanisms in the manner of physical laws, they are to be evaluated in terms of their understandability.[11]

[11] Nor is Farmer's account in any way unique. Much earlier, Hollis (1977, p. 130) made a similar claim, that "rational action is its own explanation." Coleman (1986, p. 1) subsequently concurred, arguing that "the very concept of rational action is a conception of

Other Examples

Rational choice theory, however, is not the only example of empathetic reasoning creeping into ostensibly scientific (i.e., causal) explanations of social action. Long before rational choice theory became prominent in sociology, for example, Lukes (1968) criticized individualistic explanations of social action, widespread among thinkers since the Enlightenment, for assuming that "the laws of the phenomena of society are, and can be, nothing but the actions and passions of human beings, namely the laws of individual human nature," a perspective that he quoted directly from Mill (1875, 2:469) but attributed to thinkers as varied as Weber, Hayek, Popper, Parsons, and Homans. As with rational choice theory, methodological individualism was explicitly intended to generate explanations of the causal, scientific variety. Lukes, however, argued that the central tenet of methodological individualism—"that all attempts to explain social and individual phenomena are to be rejected . . . unless they refer exclusively to facts about individuals" (Lukes 1968, p. 123)—is either absurd, if say, it requires reducing all behavior to primitive brain states, or else pointless, because it would necessarily include in the definition of "facts about individuals" all relevant features of the social world. Although stated in different terms, Lukes's critique of methodological individualism exposes essentially the same problem that arose in rational choice theory—that explanations that were intended to be causal were in effect being evaluated in terms of their ability to render observed behavior understandable.

Writing more than a decade after Lukes, Mayhew (1980) also excoriated what he called the "individualist, psychologistic perspective" that he saw as dominating American sociology. Like Lukes, moreover, Mayhew's critique lumped together explicitly individualist behavioralists like Homans with avowed structuralists like Parsons, both of whom, in Mayhew's view, concluded little more than "what you do depends on what you want" (p. 353). Once again, that is, theories of action that were superficially scientific turned out to depend in practice on ex post rationalizations of behavior, which in turn undermined their claims to causal explanatory power. According to Mayhew, for example, Parsons's notion of values was that "people have values which tell them what they want (This is what a value is.) So, people do things because they want to. That is the explanation of their behavior. If a person refrains from doing something, this means the person did not want to do that" (Mayhew 1980, p. 353). Likewise for Homans, whose notion of a value proposition was: "'The more valuable to a person is the result of his action, the more likely he is to perform the ac-

action that is 'understandable,' action that we need ask no more questions about." And Goldthorpe (1998) similarly argued that if one does not start from some conception of rational action, it is unclear from what analytic vantage point one can start.

tion.' . . . That is, values lead to action, or: people do things because they want to" (Mayhew 1980, p. 354).

Tellingly, neither Lukes nor Mayhew presented a compelling alternative to rationalizable action, thereby further emphasizing the difficulty of explaining social action without at some point invoking empathetic reasoning—a point that even Mayhew's sympathizers also emphasized (Gannon and Freidheim 1982). Nor has this difficulty been avoided by more recent theories of action. Bourdieu (2005), for example, after delivering a devastating critique of rational choice theory's reliance on what he calls "economic common sense"—the unacknowledged incorporation of historically and culturally specific assumptions about instrumental economic goals and rational calculation into supposedly universal theories of choice—then goes on to propose his own version of rationalizable action, simply a richer, more contextually dependent version. And more recently, Gross (2009), seeking to distinguish his "pragmatist theory of social mechanisms" both from Bourdieu on the one hand, and from rational choice theory on the other hand, lays out yet another variation of rationalizable action. Pragmatists, that is, "view social mechanisms as composed of chains or aggregations of actors confronting problem situations and mobilizing more or less habitual responses" (p. 368). This analytical exercise requires, in turn that, "we grasp how the relevant individuals understand the situations before them and act on those understandings, helping thereby to enact the mechanism" (p. 369).[12] Once again, presumably causal mechanisms are to be evaluated in terms of their understandability.

Finally, the pervasiveness of empathetic explanations is not limited to those that explicitly invoke individual psychological states. Boudon (1987), for example, has argued not only that the individualistic paradigm is the dominant one among sociologists but also that even overt subscribers to other paradigms often effectively invoked individualistic rationalizations in practice, albeit typically without acknowledgment. Nomological theorists, that is, attempt to explain macrosociological phenomena in terms of other macrosociological phenomena via general law-like statements. Nevertheless, when interpreting these statements they are often forced to invoke rationalizable action, precisely in order to be understood. Blau's (1987) contribution to the very same volume—an analysis of interracial marriage in terms of heterogeneity and inequality along multiple crosscutting, social

[12] In proposing a mechanism-based approach to causal explanations, Hedström and Ylikoski (2010, p. 60) express an ecumenical view on theories of action, noting that "for many social scientific purposes, a relatively simple desire-belief-opportunity model will be sufficient. For other purposes, a pragmatist theory of action could be fruitful. Habits, routines, and various unconscious cognitive processes are important parts of modern naturalistic accounts of human cognition, and sociology needs to take such factors into account." Importantly, however, both views are effectively versions of rationalizable action.

dimensions—illustrates this tendency, albeit unintentionally. On the surface, that is, Blau's analysis is what Boudon would call a macrostructural explanation, in the sense that both independent and dependent variables are macrostructural; however, when trying to explain how the resulting theorems make sense, even Blau invokes individual psychological states as explanatory variables (Blau 1987, pp. 80–84).[13]

To conclude, sociological theories of action—whether the individualist explanations critiqued by Lukes and Mayhew, the various strains of rational choice theory, Bourdieu's field theory, or Gross's pragmatism—resemble the commonsense theory of rationalizable action in that they promise scientific explanations based on generalizable causal mechanisms but in practice produce empathetic explanations that render action understandable. Moreover, I claim that this resemblance is no accident—that stripped of their jargon and theoretical pretensions, many sociological theories of action are in effect variants of the commonsense theory. Although different theories, that is, have placed widely differing emphasis on intentions, beliefs, circumstances, and opportunities and although pitched ideological battles have at times been waged over these differences, all such theories ultimately seek explanations of individual or collective action in terms of some combination of these factors. Critically, moreover, the tendency has been to evaluate these explanations in the same way as commonsense rationalizations—namely, in terms of their ability to make sense of the observed outcome. Nor, as I will argue next, can this reliance on commonsense rationalizable action easily be avoided when theorizing about social action or attempting to explain human behavior.

THEORIZING BY MENTAL SIMULATION

The reason we rely on commonsense rationalizations is that because we experience being human, we can and do "explain" human behavior in a way that we cannot for the behavior of electrons, proteins, or planets. When trying to understand the behavior of electrons, for example, the physicist does not start by imagining what it would be like to be the electrons in question. He may have intuitions about theories of electron

[13] In a similar vein, comparative historical sociologists frequently invoke structural variables to account for, say, shifts in political power in different countries (Skocpol 1979), but then explain their findings in terms of an individual "actor"—a country, say, or a population—"responding" to its structural environment, a trick that Hodgson (2007) has labeled methodological collectivism. Referring to Brinton's (1938) metaphor of revolution as a fever, e.g., Tilly (1984, p. 102) remarked, "Despite all the qualifications he attached to it, the idea of a fever suggests that revolution happens to something like a single person—to a society personified."

behavior, and exposure to these theories presumably helps him to understand their behavior in the limited sense of connecting causes and effects in a systematic and empirically verifiable manner. But at no point does he expect to view the world from an electron's perspective—indeed the very notion of such intuition is laughable. In the same way, if a neuroscientist were to propose a theory of human consciousness arising out of primitive neurological processes, she would not expect to have any intuition regarding the "rules" of neuronal behavior. She might propose plausible heuristics, as is common in the biological oscillator literature (Kopell 1988; Winfree 2000; Freeman 2003; Strogatz 2003), and seek to derive testable causal hypotheses regarding their collective dynamics, such as the existence of traveling waves, global synchronization, or chaos. She would seek to understand, in other words, the link between microrules and macrophenomena, but she would have no expectation that these microrules themselves be "understandable" in the sense of being accessible to her own subjective experience.

When the lower-level units are not electrons or neurons but people, however, not only is it possible for us to construct theories solely by introspecting on what it would be like to be them, it is arguably impossible not to. According to what is known in the philosophy of mind literature as simulation theory (Gordon 1986; Goldman 2006), when attempting to predict the behavior of others, or even ourselves in some future or hypothetical situation (e.g. "What would I do if my house were burning?" "How would I react to winning the lottery?"), we effectively create a simulation of the focal decision maker in our mind's eye, substituting our own perception of the situation for theirs and also taking into account whatever relevant information we have about their intentions, beliefs, circumstances, and so on. The behavior of our simulated agent—with the appropriately modified "us" standing in for a hypothetical "them"—is then interpreted as a prediction of what the actual agent will do. Conversely, when given a situation and an observed behavior—a stranger speaking to us as if he were an old friend, a spouse reacting to our unexpected return with surprise and anxiety, even the gambits of politicians that we read about in the daily news—the same process can be deployed in reverse to infer intentions beliefs, and so on, that we can then reconcile with the known outcome and, hence, rationalize it.

Although mundane, this ability to explain behavior—both in the scientific sense of identifying its causes and the empathetic sense of understanding it—via mental simulation is one that is thought to be unique to humans (Gilbert 2006), allowing us not only to learn from past experiences as other animals do but also to extrapolate from them to entirely novel or even hypothetical situations. Other animals can anticipate the behavior of

others, but only humans can reason about their behavior by first forming representations of their mental states, an ability that is thought to develop in children around the age of four (Wimmer and Perner 1983). It is probably also not an exaggeration to claim that without this ability any but the simplest forms of social organization would be impossible. Any form of planning or prediction, in fact—whether performed by friends, family members, and coworkers or by managers, marketers, and policy makers—is at least in part an exercise in anticipating the behavior of others in response to certain incentives, information, and constraints.

Given how useful mental simulation is for "theorizing" about human behavior in everyday life, it is not surprising that social scientists make use of the same facility when theorizing about human behavior more formally. And indeed they do, routinely. When the social scientist notices some initially puzzling behavior—for example, that public high school teachers alter their students' answers on standardized tests so as to boost their class scores (Jacob and Levitt 2003), a crime for which they could easily lose their jobs—they attempt to explain it by, in effect, putting themselves in the position of the teacher. What are her incentives for cheating? What, in her mind, are the chances that she will be caught? And what other circumstances—like low pay, inadequate teaching resources, and other job-related frustrations—might undermine the usual social norms of fair play and integrity? The data, in other words, may tell us that cheating is taking place, but only by reconstructing the relevant details of the teacher's situation in our minds, and effectively experiencing what it would be like to be her, can we understand the observed behavior. Just as in everyday life, that is, social scientists can and do explain social action simply by imagining what it would be like to be a particular actor in a particular situation and predicting how their hypothetical actor would respond.[14]

As natural and helpful as this method of explanation is, its reliance on what is essentially a commonsense intuition comes with certain unacknowledged consequences. In particular, because our everyday use mental simulation applies both to making predictions about how some actor or actors will behave and equally to making inferences about the details of the situations or actors that enable us to rationalize some behavior that we

[14] Levitt, in fact, has said precisely, "whenever I try to answer a question, I put myself in the shoes of the actors and I ask myself, 'What would I do if I were in that situation?'" (see Levitt and Dubner 2006, p. 247). Historians also use the same trick. Collingwood, in fact, argued explicitly that "history cannot be scientifically written unless the historian can re-enact in his own mind the experience of the people whose actions he is narrating"—a view that has been advocated essentially unchanged by contemporary historians (Gaddis 2002, p. 124).

have observed, and because in everyday life we switch between these two modes of prediction and sense making instinctively and without awareness, they seem to us to be flip sides of the same coin, differing only in terms of the temporal vantage point—ex ante versus ex post—at which they are carried out. Given that this is how we experience our own thought processes, it seems self-evident that predictive explanations are accessible to our interpretative faculties, and hence that the explanations that help us to make sense of human behavior ex post do correspond to causal mechanisms—just as Davidson (1963) argued.

The jump from commonsense theories of action to sociological theories is then straightforward. As James (1909) and later Boudon (1988) have argued, commonsense notions, even when not universally valid, are generally valid in everyday situations, and the plausibility they inherit from their general, everyday validity lends them an aura of universal validity. The result is that "a theory can easily be perceived as true when it is false, or as more valid than it deserves to be, if it includes beside its explicit statements implicit unnoticed commonsensical statements, which, although valid in everyday life, are not of universal validity" (Boudon 1988, p. 1). So it is, I would argue, with rationalizable action. In their everyday lives, social scientists necessarily utilize rationalizable action as a part of their mental simulation toolkit; and because the explanations they construct on an everyday basis are useful both for making sense of behavior and also for predicting and even altering it, the same conflation of empathetic and causal explanation gets incorporated into sociological theorizing without ever being explicitly acknowledged. Consequently, the conflation of understandability and causality in sociological theories has remained unchallenged even as the specific theories themselves have become contested.

Of course, if understandability and causality were effectively interchangeable, then this epistemic conflation would be unproblematic. What I want to argue next, however, is that the validity of this assumption is an illusion—that just because an explanation makes sense of some observed outcome is in fact no guarantee that it also corresponds to any generalizable causal mechanisms or was even the cause of that particular outcome in any meaningful sense. Correspondingly, reasons that seem salient ex ante will fail to make accurate predictions, while the reasons that ex post seem to have been predictively accurate, had they been known at the time, could not necessarily have been known at the time even in principle. As I will argue, in everyday life these errors are either sufficiently minor or else can be corrected sufficiently quickly with feedback, that we do not notice them. Sociologists' theories of action, however, are meant to apply much more broadly than everyday behavior, and under these circumstances the

urge to theorize via mental simulation can undermine the scientific validity of the resulting explanations.

THREE PROBLEMS WITH RATIONALIZABLE ACTION AS CAUSAL EXPLANATION

In particular, I will describe three problems, which I call the "frame problem," the "indeterminacy problem," and the "outcome problem." Each is distinct from the others, but all have the consequence of breaking the equivalence between understandability and causality in sociological explanations of action.

The Frame Problem

The first problem is that when we deploy our mental simulation apparatus to project ourselves into a particular situation, our brains do not immediately respond with a long list of questions about the messy details that our conscious minds are forced to answer before they can simulate the experience—rather these details are simply "filled in" by the unconscious mind, which draws on its assembled library of images, emotions, stereotypes, role models, and other stylized memorabilia (Schacter 2001; Gilbert 2006; Marcus 2008). And because this "filling in" process happens instantaneously and effortlessly, we are typically unaware that it is even taking place. We therefore treat the imagined details in exactly the same way as the specified details—that is, as defining features of the situation—and we calibrate our imagined reaction accordingly. In effect we treat these details as unimportant desiderata that can be ignored by our mental simulations without consequence. But as a great deal of work over the past 30 years in psychology (Gilovich, Griffin, and Kahneman 2002), and a more recent body of work in behavioral economics (Camerer, Loewenstein, and Rabin 2003), has demonstrated the way in which people form preferences, the rules they invoke when making decisions, and how they evaluate the outcomes they experience can all depend on precisely these details often with surprising consequences. Survey respondents who are given a green pen when asked to write down their favorite sports drink disproportionately list Gatorade (Berger and Fitzsimons 2008); shoppers in wine stores disproportionately buy German wine when German music is playing in the background (North, Hargreaves, and McKendrick 1997); and bidders in an auction pay more when they are asked to think of a high number beforehand (Chapman and Johnson 1994; Ariely et al. 2003). But what will they buy, and how much will they pay for it, if they are listening to music in a green room and also thinking of a number? Unfortunately, it is not clear. By their design, experimental settings emphasize one potentially relevant

factor at a time; but in real life many such factors are liable to be present to various extents, and how they all interact with each other for different kinds of decisions is a question that is far from being resolved (Gilbert and Malone 1995).

Even more troubling, the list of potentially relevant factors is itself unknown and, in fact, probably unknowable. Philosophers and cognitive scientists have long worried about what they call the "frame problem," which, roughly speaking, asks how a decision maker can determine what is relevant to their situation when deciding what to do (Dennett 1984; Fodor 2006). At face value, the answer seems self-evident, but as artificial intelligence researchers have discovered, it is notoriously difficult to codify in a noncircular manner: that is, determining what is relevant about the current situation requires one to associate it with some set of comparable situations; yet determining which situations are comparable depends on knowing which of their features to compare. So deep is this circularity, according to the philosopher Jerry Fodor, that attempts to "solve" the frame problem invariably succeed only in restating it in some other form (Fodor 2006).

The crux of the problem, Fodor argues, derives ultimately from the "local" nature of computation, which—at least as currently understood—takes some set of parameters and conditions as given and then applies some sort of operation on these inputs that generates an output. In the case of rational choice theory, for example, the "parameters and conditions" might be captured by the utility function, and the "operation" would be some optimization procedure; but one could imagine other conditions and operations as well, including heuristics, habits, and other nonrational approaches to problem solving (Gigerenzer, Todd, and ABC Research Group 1999). The point is that no matter what kind of computation one tries to write down, one must start from some set of assumptions about what is relevant, and that decision is not one that can be resolved in the same (i.e., local) manner. If one tried to resolve it, for example, by starting with some independent set of assumptions about what is relevant to the computation itself, one would simply end up with a different version of the same problem (what is relevant to that computation?), just one step removed. Of course, one could keep iterating this process and hope that it terminates at some well-defined point. In fact, one can always do this trivially by exhaustively including every item and concept in the known universe in the basket of potentially relevant factors, thereby making what at first seems to be a global problem local by definition. Unfortunately, this approach succeeds only at the expense of rendering the computational procedure intractable.

In fact, one way to view the evolution of rational choice theory over the past 40 years is as a series of encounters with the frame problem, each of

which leads to an increasingly expansive redefinition of "rational," with a correspondingly less precise definition of computation. Early proposals like those of Harsanyi and Becker began with a very narrow set of postulates about what is relevant, and well-defined notions of computation, yielding the appearance of well-defined, elegant theory. But very quickly it became clear that those postulates were inadequate to capture more than a narrow range of empirically observable human behavior, at which point a new generation of theorists revisited the *what is relevant?* question and returned a broader, more inclusive answer. Yet in applying this broader set of assumptions, analysts discovered two problems: first, that in increasing the scope of the local domain, the computational process had become increasingly difficult to pin down precisely; and second, that the inclusion of additional factors had not solved the original problem, leading to a demand for yet more inclusive notions of rationality. Viewed from this perspective, the steady evolution of claims made by rational choice theorists about what they were attempting to accomplish—starting with very clear ideas about what ought to count as a theory, and ending in denials that rational choice was ever intended to be a "theory" in the first place—was inevitable. Cognitive scientists, artificial intelligence researchers, and philosophers of mind have not been able to solve the frame problem, so it should come as no surprise that rational choice theorists have not been able to solve it either.[15]

Why do we not notice the frame problem in everyday life? One possible explanation is that mental simulation works better in practice than it should in theory because it is applied under conditions that are especially forgiving for prediction, namely concrete, immediate circumstances that are experienced repeatedly. For example, as Gilbert and Mallone (1995) point out, although mental simulations generally suffer from correspondence bias, meaning that they attribute observed behavior to intrinsic features when the correct explanation is situational, it is also often the case that intrinsic and situational features are highly correlated in practice—because, for example, people select into situations that complement their psychological predispositions (e.g., a socially awkward individual choosing a job that does not require much interaction with coworkers), or are repeatedly predicting someone's behavior in the same situation (e.g., a coworker at the office)—hence, the predictions can be accurate even if the simulation itself is flawed. A second explanation, pointed out by Gordon (1986), is that in everyday situations, the practice of theorizing about others

[15] Interestingly, the field of artificial intelligence eventually adopted a fundamentally different, and ultimately far more successful, approach to intelligence, manifested in the modern field of machine learning (Bishop 2006), which emphasizes statistical modeling of data over explicit cognitive representations.

using simulation is aided greatly by real-time feedback; thus, even when our mental simulations do make erroneous predictions, these errors can be corrected quickly.[16]

Although these explanations differ, they have the same consequence: that as serious a threat as the frame problem poses to mental simulation in theory, it may be relatively benign in everyday practice. The flip side of this explanation, however, is that the frame problem is likely not benign when mental simulation is used to theorize about human behavior that is not the everyday behavior of known alters in familiar environments. The further one gets from a familiar actor and a familiar situation, that is, the greater the potential for error to creep in, and the less feedback one receives on one's predictions the less opportunity there is to correct the errors that are made. Unfortunately, it is precisely these remote and unfamiliar conditions under which mental simulation, in the guise of rationalizable action, is typically deployed by sociologists.

The Indeterminacy Problem

A second problem is that whereas rationalizable action applies most obviously to individuals, sociologists also apply it in practice to collective behavior by invoking, often implicitly, a representative agent or individual whose intentions, beliefs, and so on are invoked to explain the actions of the collective. One of the signature accomplishments of mathematical social science, however, has been to show that when individuals in a collective interact—whether because they are behaving strategically (Olson 1965; Schelling 1978), are responding to informational constraints (Bikhchandani, Hirshleifer, and Welch 1992; Salganik, Dodds, and Watts 2006), or are simply benefiting from coordination (Arthur 1989)—their collective behavior can display emergent effects that are not reducible to the attributes of the individuals themselves (Kirman 1992). To illustrate, consider an extremely simple "threshold" model of binary choice, analyzed by Granovetter (1978), that describes a hypothetical crowd on the brink of a riot. In the model, each member of the crowd follows a simple rule: "I will riot if a critical threshold of others riot, but otherwise I will remain calm." Assuming also a distribution of individual thresholds, ranging from spontaneous

[16] Gordon gives the example of a waiter in an unfamiliar restaurant approaching his table and addressing him in a Slavic language. Immediately his mental simulation apparatus can construct more than one possible explanation: perhaps the waiter has mistaken him for another customer, a fellow countryman who appreciates speaking in his own language; but perhaps the waiter is a spy who suspects him of being his contact. Perhaps, in fact, he is a spy, in which case the second explanation may seem quite plausible. Nevertheless, after responding in English that he does not understand, the waiter quickly confesses that he has mistaken him for someone else, and the puzzle is resolved.

rioters (whose threshold is zero) to laggards, Granovetter showed that a manipulation as subtle as changing the threshold of a single actor could potentially yield collective outcomes that are essentially as distinct as possible: on the one hand, a largely orderly crowd, and on the other hand, an all-out riot. Faced with such dramatically different outcomes, any explanation based on rationalizable action via a representative agent would necessarily locate the cause in the intrinsic predispositions of "the crowd," yet we know from the formulation that the crowds in question are indistinguishable in any meaningful way. Equally troubling for rationalizations of collective behavior, one can easily show that the converse applies: that very different distributions of individual thresholds can yield indistinguishable collective outcomes.

As simple as it is, Granovetter's model provides a powerful insight: that whenever social interactions are involved in collective behavior, the relation between individual attributes and collective outcomes is fundamentally indeterminate in the sense that indistinguishable attributes may lead to wildly divergent outcomes, and conversely, rather different attributes may yield indistinguishable outcomes. Note that this "indeterminacy problem" is distinct from the frame problem, in two ways. First, whereas the frame problem is generally understood to pertain to individual behavior, the indeterminacy problem applies to collective behavior. And second, whereas the frame problem is concerned with what is relevant to a given situation, the indeterminacy problem concerns the nonuniqueness of the relation between micro (i.e., individual) features and macro outcomes. One does not need to have the indeterminacy problem to worry about the frame problem, and one is still stuck with it even if the frame problem could be somehow resolved. Like the frame problem, however, the indeterminacy problem poses a fundamental challenge to the premise of rationalization action that causal explanations of observed outcomes can be obtained exclusively in terms of the intentions, beliefs, and so on of the actors involved. If in fact, any given set of intentions can lead to many distinct outcomes, and any given outcome can be arrived at from many distinct sets of intentions, then given any one realization of an outcome there may be little that can be said about the intentions and so forth that lead to it.

The Outcome Problem

A third, and also insidious, problem with rationalizable action is its casual handling of the notion of outcome. We routinely talk about "outcomes" without feeling any confusion about what we mean, and when we either look forward to the future or back to the past, it seems clear to us which outcomes are relevant to what intentions. When I submit a paper to a journal, presumably the outcome I have in mind is getting it published;

thus, my intention is to write it in the way that maximizes the probability of that outcome. Yet it is also the case that the value we assign to past events, and even our recollection of how we felt about them at the time, is frequently modified by events that happened subsequently: a rejection from our first choice of journal leads us to rewrite the paper in what later seems to a more interesting way; a missed promotion is later construed as the stimulus leading to a more fulfilling career; and a failed relationship, distressing at the time, is replaced by a healthier one. Nor are our revisionist tendencies restricted to positive reevaluations: as all too many lottery winners can attest, what initially seems like a blessing may in fact be a curse. In many familiar circumstances, in fact, we look back upon events that we would have evaluated as good or bad "outcomes" ahead of time, and possibly even at the time itself, but which in retrospect appear not as outcomes at all, but rather as intermediate steps in the process leading up to the "real" outcome—that is, the point at which we are now making the evaluation.

Even worse, the same reasoning suggests that there is no guarantee, and indeed no way to know, that our current evaluation is any more correct than evaluations we might have made in the past.[17] Precisely this point, in fact, was made by Danto (1965) with respect to historiography: because the significance of any event is invariably a function of subsequent events, and because one can never be sure what those events will be, or when they will happen—possibly well after the initial event itself—history cannot be told at the time it is happening, even in principle. To characterize the invention of HTML as the invention of the World Wide Web, for example, is possible only once we know both that the Web is itself an object of historical importance, and also that the particular way in which it developed—web browsers, search engines, e-commerce, social networking sites—built on HTML and not on some other language. Because "the Web" is the sum of all these cumulative contributions which played out over the course of over 20 years, the historical significance of HTML could not have been evaluated at the time of its happening—a point that Sewell (1996) has also argued forcefully.

[17] A traditional Taoist story tells of an old farmer who had worked his crops for many years. One day his horse ran away. Upon hearing the news, his neighbors came to visit. "Such bad luck," they said sympathetically. "May be," the farmer replied. The next morning the horse returned, bringing with it three other wild horses. "How wonderful," the neighbors exclaimed. "May be," replied the old man. The following day, his son tried to ride one of the untamed horses, was thrown, and broke his leg. The neighbors again came to offer their sympathy on his misfortune. "May be," answered the farmer. The day after, military officials came to the village to draft young men into the army. Seeing that the son's leg was broken, they passed him by. The neighbors congratulated the farmer on how well things had turned out. "May be," said the farmer (quotation from http://truecenterpublishing.com/zenstory/maybe.html).

The troubling corollary of this observation is that explanations of historical events that purport to describe merely "what was happening" at the time are fundamentally contaminated by the historians' knowledge of what happened after. For example, to describe the invention of HMTL as the "invention of the Web" relies on what Danto calls a narrative sentence: a sentence that appears to be a description of something happening at the time but that invokes a kind of foreknowledge, a prophetic ability to describe what was happening with the benefit of future knowledge. Danto, moreover, argued that narrative sentences are not merely a useful shortcut for historians, but inextricable from historical descriptions: that, in a sense, historiography without narrative sentences would not satisfy the primary goal of historians to make sense of the past. Thus, he concluded, the dependence of "what happened" on what has happened since arises not only for the trivial reason that events cannot be described until after they have taken place but for the deeper reason that "what happened" in a historical sense inextricably incorporates subsequent events.

Danto's arguments were directed primarily at historians, but they are also relevant to sociologists' explanations of events in the past. If it is not possible even to describe a particular event without reference to subsequent events, that is, then explanations of what caused that event are also potentially contaminated by information about outcomes beyond the "outcome" in question. Even worse, because there is never any definitive endpoint at which the significance of events can be finally determined, explanations of the past are potentially open to revision indefinitely. Unlike the frame problem, moreover, which is defined for individual theories of mind, or the indeterminacy problem, which is relevant primarily to collective behavior, this "outcome problem" applies equally to individual and collective outcomes and is distinct from either problem. Even if the future and the past were linked in a unique way, that is, and even if there were no problem determining which factors were relevant to a given situation, one would still be stuck with the problem of determining at what point in the future one should evaluate the consequences of some action.

To summarize, the combination of the frame problem, the indeterminacy problem, and the outcome problem violates the assumption implicit in theories of rationalizable action that explanations that render observed action understandable can be safely regarded as causal. It may be true that intentions lead to plans, plans lead to actions, and actions lead to outcomes. And it may also be true that having observed the outcome, it is possible for an observer to rationalize it in terms of some combination of intentions, beliefs, circumstances, and opportunities of the actors involved. But if these same actors typically do not know (or cannot know) what it is about the future states they are selecting among that is relevant to the choice at hand, if the outcomes that are realized are the indeterminate product of

many such individuals' interactions with each other, and if the evaluation of outcomes themselves can depend on events that take place subsequently, then it is not clear that the intentions, plans, actions, and so on that seem relevant at the time of explanation could even have been knowable ex ante. Ex post explanations of individual or collective action that are evaluated on the grounds of understandability are therefore at best not self-evidently causal and at worst unlikely to be.

REFRAMING EXPLANATION AS PREDICTION

If understanding in an empathetic sense is an unreliable guide for sociological explanations, then how should sociologists proceed? There is no globally applicable solution, but a number of partial solutions have been proposed over the years in sociology and economics, as well as in related fields such as statistics and computer science, that if taken seriously would help guard against faulty methodological practices. At a conceptual level, moreover, they would force sociologists to confront the difference between empathetic and causal explanation and thereby to produce more scientifically rigorous if not more satisfying explanations. Although the implications of these methods may be controversial in some respects, the methods themselves all start from the same point of (near) universal agreement—namely, Woodward's manipulationist criterion that explanations must answer a what-if-it-had-been-different question—and then proceed to lay out different but related standards of evidence for such claims to be taken seriously.

The most straightforward such approach, in theory if not in practice, is for sociologists to rely more on experimental methods. In particular, field experiments (Harrison and List 2004; Gerber and Green 2012), in which causal effects can be identified by virtue of random assignment, have long been regarded as the gold standard of evidence in medicine (Madigan et al. 2014) and are becoming increasingly popular among social scientists, including sociologists (Pager, Western, and Bonikowski 2009; van de Rijt et al. 2014). Less clean but also useful are natural experiments and quasi-experiments (Gerber and Green 2012), which exploit naturally occurring randomness—such as draft lottery—or other salient features of the environment, such as geographical boundaries, cutoffs, or errors (Sorensen 2007), that are plausible substitutes for random assignment. Finally, laboratory experiments can also be used to identify causal effects, as has long been common in social psychology (Asch 1953; Milgram 1969) and, more recently, in behavioral economics (Camerer et al. 2003). With notable exceptions (e.g., Cook et al. 1983), lab experiments have been less prevalent in sociology; however, web-based "virtual labs" have recently demonstrated the feasibility of experiments with as many as tens of thousands

of participants, suggesting that "experimental macrosociology" (Hedstrom 2006) may be more realistic than previously thought (Zelditch 1969).

Although experimental methods could and arguably should receive more attention from sociologists, an inevitable limitation is that for many questions of interest to sociologists experiments are either difficult or impossible to implement in practice. Overwhelming in field experiments, for example, the unit of analysis is the individual, whereas sociologists are often interested in collective entities such as organizations, markets, or cultures. In such cases, "treatments" may be unfeasible, or the available population of comparable entities (e.g., countries) may be too small, or both. Even when individuals are the unit of analysis—as in life course studies—random assignment may be impractical or unethical, and other requirements of identification, such as noninterference and excludability (Gerber and Green 2012), may be violated. Another frequent criticism of field experiments is that even where they are possible, it can be difficult to generalize from one field setting to another. Lab experiments, finally, tend to suffer from even greater problems of external validity than field experiments, while natural and quasi-experiments, although sometimes useful in situations where designed experiments are unfeasible, are limited to instances in which approximations to random assignment occur naturally.

A natural alternative to experiments, therefore, is that offered by the counterfactual model of causal inference (Rubin 1974; Morgan and Winship 2007) applied to nonexperimental data—an approach that most naturally applies to "large *N*" observational studies. Correspondingly, the avalanche of digital data that is being generated both by online activities and also parallel efforts to digitize existing archival records, combined with massively faster computational methods and methodological advances in statistics (Gelman and Hill 2007) and econometrics (Manski 2007) is opening up many exciting opportunities for sociologists to explore questions of long-standing interest with better tools and data (Lazer et al. 2009). As the extensive literature on identification and estimation makes clear, however, making valid causal inferences based on nonexperimental data is problematic in ways that do not necessarily diminish with the increasing size of data sets. Of particular concern is that substantive conclusions are overly dependent on modeling assumptions—for example, that causal effects can be identified by conditioning on observable covariates, or that errors associated with model specification are ignorable—and even experienced researchers can make elementary mistakes (Young 2009). The result, as a growing number of critics have pointed out (Sobel 2000; Ioannidis 2005; Madigan et al. 2014), is that many causal claims based on quantitative analysis of observational data—including those published in leading journals—are at best unsupported by the evidence. As with experimental meth-

ods, moreover, many of the questions of interest to sociologists—including those that motivate single-case studies (Abbott 1992) and small-*N* comparative studies (Mahoney 2000), but also the kind of analytical sociology advocated by Hedström (2005) and others—involve data that are simply unsuitable for statistical modeling.

An alternative both to running experiments and also to estimating statistical models that is simpler to implement in practice and more general in applicability is to evaluate explanations in the manner suggested by Hempel and Oppenheim—namely, their ability to predict.[18] As noted earlier, however, whereas social scientists almost universally agree that valid explanations should be causal, there is much less agreement on whether causal explanations need to be predictive. To be clear, the disagreement is not over whether predictive accuracy alone is sufficient to establish causality: elementary statistics shows clearly that it is not. No matter how well type of schooling predicts lifetime income, for example, the prediction in itself tells us little about the causal effect on income of private schools versus the confounding effects of selection. Indeed, it is precisely this "reflection" problem (Manski 2007) that the methods of causal inference were designed to address. Rather the argument is about whether or not prediction is even a necessary condition for causality—a view that some researchers (Freedman 1991; Manski 2007) find self-evident, but that others (Lieberson and Lynn 2002; Hedström and Ylikoski 2010) emphatically dispute. What I will argue next, however, is that much if not all of this disagreement derives from differing definitions of "prediction" and that under a suitably broad definition essentially all sociologists would agree that valid causal explanations do in fact make predictions. Specifically, there are three points of confusion: first, that prediction implies deterministic prediction; second, that prediction is necessarily about the future; and, third, that predictions can be made only about specific events or outcomes.

The first objection is illustrated by Lieberson and Lynn (2002), who argue that sociology should abandon its historical aspiration to the predictive accuracy of physics and instead model itself on evolutionary theory, on the grounds that the latter emphasizes explanation over prediction. While agreeing that the biological sciences provide, on the whole, a more appropriate benchmark for sociology than physics, I would argue that Lieberson and Lynn's rejection of prediction presumes that predictions are necessarily deterministic, in the sense of predicting specific events (e.g., the

[18] As Freedman (1991, p. 293) has advocated: "Take the model as a black box and test it against empirical reality. Does the model predict new phenomena? Does it predict the results of interventions? Are the predictions right?"

return of Halley's comet) with near certainty. Overwhelmingly, however, modern applications of prediction, including in physics, but also in statistics, computer science, and econometrics, conform to a much looser, probabilistic standard of prediction, according to which one must merely demonstrate that the probability of an event Y increases in the presence of some other factor X relative to its absence. And by this standard, the explanations by which Lieberson and Lynn judge evolutionary theory successful are indeed predictive. For example, although evolution does not predict how any particular species will look at some point in the future, or which particular species will dominate at any point in time, it does make statements of the form "when A is observed we can predict that X is more likely to occur than without A, but still extremely unlikely" (Scriven 1959, p. 480).

A second objection to prediction as a standard for evaluating explanations is that predictions are inherently about the future, whereas much of what sociologists seek to explain concerns the past. Once again, however, this objection rests on an overly narrow definition of prediction, which might more precisely be called "forecasting." In fact, although social scientific theories could, and perhaps should, be subjected to tests of forecasting accuracy—as Tetlock (2005) and others (e.g., Schrodt 2013) have argued—especially where they are relevant to policy, forecasts are just one special class of predictions. A more inclusive definition of prediction can easily encompass past events, where the defining criterion is simply that the data about which one is making a "prediction" cannot be the same data one has used to motivate the explanation in the first place—an approach that is known as out-of-sample testing.[19] Although many specific methods have been introduced in statistics and also in fields like computer science (Kohavi 1995; Bishop 2006; Provost and Fawcett 2013), where prediction is a central concern, what is important for the current argument is that in all such methods the model is estimated or "fit" on one set of data—usually called the "training" data—and then is evaluated exclusively on a distinct "test" or "holdout" set (Provost and Fawcett 2013). Obviously future events can be used as test data, but past events can also be used so long as they are not used in the process of generating the explanation itself.

Out-of-sample testing is similar to the standard practice of hypothesis testing but differs in two important respects. First, although in theory hypothesis testing requires that hypotheses be specified in advance of the test, in social science this rule is rarely enforced and almost certainly is

[19] As Babyak (2004, p. 414) explains, "If you use a sample to construct a model, or to choose a hypothesis to test, you cannot make a rigorous scientific test of the model using that same sample data. This, by the way, is the real statistical meaning of the term post-hoc—it does not refer to afterward in terms of time. Rather, it refers to looking at the data to decide which tests or parameters will be included in the analysis and interpretation."

routinely violated in practice (Leamer 1983; Young 2009).[20] The result is that many "findings" are likely artifacts of model overfitting (Sarle 1995; Babyak 2004) in which the model is effectively "explaining" the noise in the sample observations in addition to the hypothesized function. By insisting on strictly out-of-sample tests of model performance, many common problems of hypothesis testing can be avoided. Second, out-of-sample testing also shifts attention from the sign and significance of coefficients—the overwhelming focus of standard hypothesis testing in social science—to the overall performance of the model: How much of the observed variance does it account for? Even if the signs of the coefficients are in the predicted direction, they are highly significant and large enough to be substantively interesting, that is, and even if one's model only accounts for 10% of the variance, then it is still the case that most of what has been observed is unexplained by the proposed hypothesis.[21]

A final objection to prediction as a necessary condition for causality is that causal mechanisms may be proposed to explain not only what Hedström and Ylikoski call "empirical facts" but also "stylized facts," which may not be sufficiently quantifiable to be subjected to standard performance metrics. While agreeing with the usefulness of explaining stylized facts, I would argue that any mechanism-based explanation that would pass Hedström and Ylikoski's test of empirical validity does in fact also make predictions—just predictions of a different sort than what they have in mind.[22] For example, Hedström and Ylikoski approvingly cite Watts and Strogatz's (1998) "small-world" network model as an example of mechanism-based-explanation for the stylized fact that path lengths in even large social networks are surprisingly short. But although the Watts-Strogatz model did not make predictions about specific outcomes—the "empirical facts" to which Hedström and Ylikoski allude—it nevertheless did make a clear and testable prediction—namely, that any mechanism that generated some amount of local clustering and any nonzero amount of

[20] In medical trials, by contrast, it is common to require that hypotheses be recorded before the trial is conducted, a requirement that prevents pharmaceutical companies from running many trials and only reporting "positive" results (Young 2009).

[21] It is of course true that in some cases an R^2 of 0.10 can be useful, either because even slightly better predictions on average can yield meaningful differences in payoffs (as in high frequency trading or online advertising) or because it may correspond to very accurate predictions about a relatively small fraction of cases for which it can be extremely informative (as with certain genetic predispositions to illness). It is one thing, however, to defend an R^2 of 0.10 as nevertheless useful; it is another not to mention it.

[22] While Hedström and Ylikoski (2010, p. 55) state that mechanism-based explanation "severs the close connection between explanation and prediction" they nevertheless remain globally in agreement with Woodward's manipulationist view of causality, stressing that mechanism-based explanations "still rely on causal generalization" and require "an ability to make correct what-if inferences."

long-range random linking would lead to a small-world network topology.[23] Relatedly, mechanism-based explanations may make testable predictions about patterns or distributions of empirical facts even when they do not make predictions about facts themselves.[24] Salganik et al. (2006), for example, proposed that in markets for cultural goods, the presence of social influence causes both increased inequality in the distribution of success and also increased unpredictability in the success of particular goods; moreover, they also predicted that unpredictability and inequality should increase with the strength of the social signal. Thus, although they did not make predictions about the success of individual songs—indeed, their point was precisely that individual success is inextricably unpredictable—Salganik et al. did make clear predictions about the distribution of success.[25]

To summarize, the claim that prediction is a necessary (but not sufficient) feature of causal explanation is consistent with a view of causality that is almost universally accepted by sociologists—even sociologists who have explicitly denied the necessity of prediction. The resolution of the apparent conflict is that prediction must be defined suitably—that is, in the broad sense of out-of-sample testing, allowing both for probabilistic predictions and for predictions about stylized facts or patterns of outcomes. Defined in this manner, moreover, prediction can be used to evaluate not only statistical models of "large-*N*" data but also mathematical or agent-based models (Hedström 2005), small-*N* comparative studies (Mahoney 2000), rational choice explanations of historical events (Kiser and Hector 1998), or even mental models based on intuition and experience. Although the details would differ depending on the type of explanation in question, in all cases the procedure would be roughly: (1) construct a "model" based on analysis of cases (A, B, C, . . .); (2) deploy the model to make a prediction about case *X*, which is in the same class as (A, B, C, . . .) but was not used to inform the model itself; (3) check the prediction.[26]

[23] Specifically Watts and Strogatz (1998, p. 442) wrote, "Although small-world architecture has not received much attention, we suggest that it will probably turn out to be widespread in biological, social and man-made systems, often with important dynamical consequences." This prediction has subsequently been validated on hundreds of networks spanning many substantive domains and orders of magnitude in size (Goel, Muhamad, and Watts 2009; Ugander et al. 2011).

[24] Cartwright (2006, p. 242) has made a similar point, arguing that "predictions from our simple model about the kinds of effects that result from this kind of change may be relatively accurate even if no prediction about the level of the price itself will be."

[25] They then tested these predictions in a web-based "virtual lab" experiment involving over 14,000 participants. Subsequent field experiments have found corroborating evidence for cumulative advantage, although these experiments were not designed to test for unpredictability (Muchnik, Aral, and Taylor 2013; van de Rijt et al. 2014).

[26] Consider, e.g., Kiser and Hechter's (1998) claim that a wide range of historical events—including the rise of Christianity (Stark 1996), the outcomes of revolutions

To clarify the point further, what would it mean for a causal explanation not to satisfy this requirement? What would it mean, in other words, to claim to have provided a causal explanation for some outcome but also insist that one's explanation makes no predictions other than about that precise outcome, and hence that out-of-sample testing is inapplicable? As Mitchell (2004) points out, causal claims of this kind are rather common. Using the example of the collapse of Enron, Mitchell argues that many subsequent explanations were used to impute lessons about corporate governance—lessons that could only be meaningful if the explanations were indeed causal in the manipulationist "had *X* been different" sense. When challenged to demonstrate the generalizability of their explanations, however, proponents responded that they were relevant only to the case in question—hence no test of generalizability should be required. Are these claims simply disingenuous? Mitchell believes not. Rather he speculates that proponents genuinely believe their explanations to be causal in the standard counterfactual sense that "had *X* been different" (where *X* is the purported cause) the collapse would have been averted or mitigated. Their error, however, is that the counterfactual in question is purely hypothetical: a mental simulation, conducted in the mind of the analyst, of what the world would have looked like had *X* been different. Given the impossibility of verifying any such counterfactual empirically, and given that it is inevitably subject to all the problems of mental simulation discussed earlier, the resulting "explanation" is in fact nothing of the sort but, rather, is what Mitchell (2004) calls a "causal story"—a description of events that invokes the language of causality but none of its evidentiary basis.

Another way to see the problem with causal stories is that they answer what Gelman and Imbens (2013) call "reverse causal questions": "Why" questions, such as "why did Enron collapse?" or "why do incumbents get more contributions than challengers?" Gelman and Imbens distinguish these questions from questions of "forward causal inference," which are the "what if" questions of counterfactual causal reasoning, such as "what is the effect of private schooling on income" or "what is the effect of po-

(Lindenberg 1989) and of labor strikes (Brown and Boswell 1995), shifts in medieval marriage norms (Ermakoff 1997), and the persistence of the QWERTY keyboard (David 1985)—have been successfully explained using a rational choice framework. Because in all these cases the analyst has at his or her disposal a very large set of potentially relevant factors (model features) to explain only a small number of events, the result is almost certainly a "model" that overfits the data—an error that would be exposed by requiring it to account for other, out-of-sample cases. In a similar manner, arguments over causal explanations in small-*N* studies (Mahoney 2000) could also be avoided, or at least restricted to instances in which the explanations pass some out-of-sample predictive test.

litical advertising on turnout." Although both types of questions appear natural, Gelman and Imbens point out that only "what if" questions are directly answerable via accepted methods of causal inference. By contrast, "a reverse causal question does not in general have a well-defined answer, even in a setting where all possible data are made available" (p. 6). In a nutshell, the problem with reverse causal questions is that they generally admit many potential answers—as is evidenced by the many "explanations" that Mitchell documents for the collapse of Enron—any one of which could be tested individually as a forward hypothesis. No method, however, exists for testing all possible explanations simultaneously, especially as one can never be sure that all possible explanations have been included. By invoking imagined counterfactuals, causal stories sweep this causal messiness under the rug of understandability, but the appearance is deceptive. By forcing causal stories to make out-of-sample predictions, the rug can be pulled up and the deception revealed.

BEYOND COMMON SENSE

The purpose of this essay has been to explain a curious pattern: that sociologists in theory care deeply about causality in their explanations but in practice evaluate them on the grounds of understandability—assuming, in effect, that understandability and causality are one and the same. I have proposed that this conflation is not grounded either in theory or empiricism but rather in common sense. Just as in everyday life, that is, when sociologists think about why people do what they do (explanation), what they might do in the future (prediction), and possibly how to make them do something differently (intervention), their mental simulation apparatus—so useful for navigating everyday social interactions—will spontaneously generate answers that invoke the intentions, beliefs, circumstances, and opportunities of the actors involved. These answers, moreover, will have the form of causal explanations in that they specify mechanisms that connect potential causes to known outcomes. In fact, however, these explanations are of the empathetic, not causal, variety—a distinction that, although possibly ignorable in everyday situations, is unlikely to be so for the kind of large-scale, spatially distributed, or temporally distant phenomena that are of primary interest to sociologists. Nevertheless, because sociologists—in their professional capacities as in their everyday lives—are generally called on to explain things only ex post, their ability to tell plausible causal stories is so powerful that it seems superfluous to subject them to tests of scientific validity.

Explanations invoking rationalizable action, in other words, whether of the informal everyday variety or the formal theoretical kind, attain their apparent power from a kind of self-fulfilling prophecy: because rationali-

zations are understandable in a way that requires no further justification, we are motivated to seek explanations of this form, and because our mental simulation apparatus is capable of rationalizing essentially any conceivable behavior, we are always able to find explanations that render the observed outcome understandable.[27] Thus satisfied, we are rarely prompted to question the unstated assumption that the details of the explanations we construct could in principle have been identified as relevant ex ante, hence predictive. That is not to say that rationalizations are useless for sociology—as Gelman and Imbens (2013) have argued, they can be viewed as a source of plausible hypotheses[28]—just that in the absence of further testing, they are no more likely to be causal than numerous other potential mechanisms, both plausible and implausible, that might have quite different implications for prediction and intervention.

The good news is that through closer attention to causal inference, experimental methods, and out-of-sample testing, it ought to be possible to improve the scientific validity of sociologists' explanations. The bad news is that attention to these issues will do more than highlight the difference between scientific validity and understandability—it will also reveal that they are in tension with one another. Precisely because they do not have to satisfy any of the standards of causal inference or prediction, explanations that are evaluated solely on the basis of understandability can be satisfying in ways that scientifically valid explanations cannot be. Causal stories, for example, can be extremely rich and detailed precisely because the counterfactuals in question are hypothetical and so limited only by the analyst's imagination. Models that do not have to worry about overfitting can include potentially as many features as they have observations and hence can trivially achieve the appearance of perfect within-sample prediction (Provost and Fawcett 2013). And explanations that are constructed exclusively with the goal of making sense of something can be optimized for plausibility. It seems inevitable, therefore, that the more rigorously a hypothesis is tested and the more data it is tested against, the weaker it must be in order to survive scrutiny (Manski 2007).

[27] Elster (1993, pp. 183, 189) has made a similar argument about rationality: that it is appealing in a "hermeneutic" or "interpretive" sense, meaning that for whatever reason, humans like to find reasons for things, including their own behavior, and rationality—however it is defined—satisfies this need.

[28] Specifically, Gelman and Imbens argue that asking a why question implies a deficiency in one's previous model of reality, in which, say, Enron should not have collapsed, or all candidates should get the same level of support. In turn, the deficiency generates new candidate hypotheses (perhaps it was the culture of greed, or lack of oversight, that caused Enron's collapse; perhaps voters like to support a winner, or incumbents have higher name recognition) that can then be evaluated by the methods of forward causal inference.

To illustrate the problem consider a recent study by Brand and Xie (2010), who studied the effect of attending college on the subsequent earnings of graduates in two distinct longitudinal panels—one national and one from Wisconsin—finding that individuals who benefit the most from college are those least likely to attend. Their approach of testing their hypothesis on multiple data sets is admirable and does arguably increase confidence in their main finding; however, it also has the effect of weakening the finding itself. For example, although the finding is strongly supported in some cases (for men in the national panel and women in the Wisconsin panel), other effects (for women in the national panel and for men in the Wisconsin panel) are essentially zero. Moreover, depending on which panel one looks at, one could conclude that the effect is stronger either for men (national panel) or for women (Wisconsin). Considered together, therefore, the only robust conclusion is that evidence for positive selection is scarce—a weaker finding than could have been concluded if only one data set had been considered.[29]

I predict that this pattern will generalize. In particular, as more and larger data sets become available, and as methods such as out-of-sample testing become increasingly familiar to sociologists, it will become increasingly apparent how much variance our explanations can actually account for. I predict that it will be less than we would like—and certainly less than the discussion section of many existing papers, both quantitative and qualitative, would lead an innocent reader to believe. Even less satisfying, increasing awareness of the limitations of causal inference will force sociologists to confront the problem that in some instances, as for example when trying to explain the cause of a truly unique historical event—the stunning success of Apple, the collapse of Enron, or the meltdown of the global financial system in 2008—the only honest conclusion may be that no answer is possible. As these trends play out, sociologists will increasingly have to choose between scientifically rigorous but empathetically unsatisfying explanations and satisfying but unscientific stories. Viewed optimistically, however, these same trends offer sociologists, especially those willing to work with computer scientists and researchers in related disciplines, an unprecedented opportunity to finally realize Stouf-

[29] A similar result emerges from a more recent study of long-run learning in games of cooperation (Mason, Suri, and Watts 2014), in which the authors reanalyze data from three separate series of "virtual lab" experiments conducted over a two-year period. Although they find evidence of learning in all three cases, the specifics vary in ways that force weaker conclusions than if only one example had been studied. Much the same conclusion can also be reached from the results of meta-analytical studies, which consistently attribute the majority of total variation in results to model error, not sampling error (Young 2009)—meaning, in effect, that many explanations are presented with a higher degree of certainty than could be supported if more data were available.

fer's (1947, p. 12) call to "use *uncommon* sense . . . to formulate our thinking so that if it is wrong we can be proved wrong . . . to design scientifically controlled experiments which those in a hurry will call trivial . . . [to] work with a timeless patience in forging sociological theories which eventually can be applied by social actionists for the betterment of mankind."

REFERENCES

Abbott, Andrew. 1992. "What Do Cases Do? Some Notes on Activity in Sociological Analysis." Pp. 53–82 in *What Is a Case*, edited by Charles Ragin and Howard Becker. New York: Cambridge University Press.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "Coherent Arbitrariness: Stable Demand Curves without Stable Preferences." *Quarterly Journal of Economics* 118 (1): 73–105.

Arrow, Kenneth J. 1986. "Rationality of Self and Others in an Economic System." *Journal of Business* 59 (4): S385–S99.

Arthur, W. Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-In by Historical Events." *Economic Journal* 99 (394): 116–31.

Asch, Solomon E. 1953. "Effects of Group Pressure upon the Modification and Distortion of Judgments." Pp. 151–62 in *Group Dynamics: Research and Theory*, edited by D. Cartwright and A. Zander. Evanston, Ill.: Row, Peterson.

Babyak, Michael A. 2004. "What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models." *Psychosomatic Medicine* 66 (3): 411–21.

Becker, Gary S. 1976. *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.

Becker, Howard S. 1998. *Tricks of the Trade: How to Think about Your Research While You're Doing It*. Chicago: University of Chicago Press.

Berger, Jonah, and Gráinne Fitzsimons. 2008. "Dogs on the Street, Pumas on Your Feet: How Cues in the Environment Influence Product Evaluation and Choice." *Journal of Marketing Research* 45 (1): 1–14.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992–1026.

Bishop, Christopher M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer.

Black, Donald. 1979. "Common Sense in the Sociology of Law." *American Sociological Review* 44 (1): 18–27.

Blau, Peter M. 1987. "Contrasting Theoretical Perspectives." Pp. 71–85 in *The Micro-Macro Link*, edited by Jeffrey Alexander, Bernard Giesen, Richard Münch, and Neil J. Smelser. Berkeley and Los Angeles: University of California Press.

Boudon, Raymond. 1987. "The Individualistic Tradition in Sociology." Pp. 45–70 in *The Micro-Macro Link*, edited by Jeffrey C. Alexander, Bernhard Giesen, and Richard Munch. Berkeley and Los Angeles: University of California Press.

———. 1988. "Common Sense and the Human Sciences." *International Sociology* 3 (1): 1–22.

———. 1998. "Limitations of Rational Choice Theory." *American Journal of Sociology* 104 (3): 817–28.

Bourdieu, Pierre. 2005. *The Social Structures of the Economy*. Malden, Mass.: Polity.

Brand, Jennie E., and Yu Xie. 2010. "Who Benefits Most from College? Evidence for Negative Selection in Heterogeneous Economic Returns to Higher Education." *American Sociological Review* 75 (2): 273–302.

Brinton, Crane. 1938. *The Anatomy of Revolution*. New York: Norton.
Brown, Cliff, and Terry Boswell. 1995. "Strikebreaking or Solidarity in the Great Steel Strike of 1919: A Split Labor Market, Game-Theoretic, and QCA Analysis." *American Journal of Sociology* 100:1479–1519.
Buchanan, James M. 1989. "Rational Choice Models in the Social Sciences." Pp. 37–50 in *Explorations into Constitutional Economics*. College Station: Texas A&M University Press.
Camerer, C. F., G. Loewenstein, and M. Rabin. 2003. *Advances in Behavioral Economics*. Princeton, N.J.: Princeton University Press.
Cartwright, Nancy. 2006. "From Causation to Explanation and Back." Pp. 230–45 in *The Future of Philosophy*, edited by Brian Leiter. Oxford: Clarendon Press.
Chapman, Gretchen B., and Eric J. Johnson. 1994. "The Limits of Anchoring." *Journal of Behavioral Decision Making* 7 (4): 223–42.
Coleman, James S. 1986. *Individual Interests and Collective Action: Selected Essays*. Cambridge: Cambridge University Press.
Coleman, James S., and Thomas J. Fararo. 1992. *Rational Choice Theory: Advocacy and Critique*. Thousand Oaks, Calif.: Sage.
Conlisk, John. 1996. "Why Bounded Rationality?" *Journal of Economic Literature* 34 (2): 669–700.
Cook, Karen S., Richard M. Emerson, Mary R. Gillmore, and Toshio Yamagishi. 1983. "The Distribution of Power in Exchange Networks: Theory and Experimental Results." *American Journal of Sociology* 89:275–305.
Cox, Gary W. 1999. "The Empirical Content of Rational Choice Theory: A Reply to Green and Shapiro." *Journal of Theoretical Politics* 11 (2): 147–69.
Danto, Arthur C. 1965. *Analytical Philosophy of History*. Cambridge: Cambridge University Press.
David, P. A. 1985. "Clio and the Economics of QWERTY." *American Economic Review* 75 (2): 332–37.
Davidson, Donald. 1963. "Actions, Reasons, and Causes." *Journal of Philosophy* 60 (23): 685–700.
Dawes, R. M. 2002. *Everyday Irrationality: How Pseudo-Scientists, Lunatics, and the Rest of Us Systematically Fail to Think Rationally*. Boulder, Colo.: Westview.
Dennett, Daniel C. 1984. "Cognitive Wheels: The Frame Problem of AI." Pp. 129–51 in *Minds, Machines and Evolution*, edited by C. Hookaway. Cambridge: Cambridge University Press.
Diermeier, Daniel. 1996. "Rational Choice and the Role of Theory in Political Science." Pp. 59–70 in *The Rational Choice Controversy: Economic Models of Politics Reconsidered*, edited by Jeffrey Friedman. New Haven, Conn.: Yale University Press.
Elster, Jon. 1993. "Some Unresolved Problems in the Theory of Rational Behavior." *Acta Sociologica* 36:179–90.
———. 2009. *Reason and Rationality*. Princeton, N.J.: Princeton University Press.
Ermakoff, I. 1997. "Prelates and Princes: Aristocratic Marriages, Canon Law Prohibitions, and Shifts in Norms and Patterns of Domination in the Central Middle Ages." *American Sociological Review* 62:405–22.
Farmer, Mary K. 1992. "On the Need to Make a Better Job of Justifying Rational Choice Theory." *Rationality and Society* 4 (4): 411–20.
Ferguson, Niall. 2000. *Virtual History: Alternatives and Counterfactuals*. New York: Basic Books.
Fodor, Jerry. 2006. "How the Mind Works: What We Still Don't Know." *Daedalus* 135 (3): 86–94.
Foley, Duncan K. 2004. "Rationality and Ideology in Economics." *Social Research: An International Quarterly* 71 (2): 329–42.
Freedman, David A. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology* 21 (2): 291–313.

Freeman, W. J. 2003. "A Neurobiological Theory of Meaning in Perception. Part I: Information and Meaning in Nonconvergent and Nonlocal Brain Dynamics." *International Journal of Bifurcation and Chaos* 13 (9): 2493–2511.
Friedman, Jeffrey, ed. 1996. *The Rational Choice Controversy: Economic Models of Politics Reconsidered*. New Haven, Conn.: Yale University Press.
Gaddis, John Lewis. 2002. *The Landscape of History: How Historians Map the Past*. New York: Oxford University Press.
Gannon, Thomas M., and Elizabeth A. Freidheim. 1982. "'Structuralism' or Structure: A Comment on Mayhew." *Social Forces* 60 (3): 877–82.
Geertz, Clifford. 1975. "Common Sense as a Cultural System." *Antioch Review* 33 (1): 5–26.
Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press.
Gelman, Andrew, and Guido Imbens. 2013. "Why Ask Why? Forward Causal Inference and Reverse Causal Questions." NBER Working paper no. 19614 (November). National Bureau of Economic Research, Cambridge, Mass.
Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. New York: Norton.
Gigerenzer, Gerd, Peter M. Todd, and ABC Research Group. 1999. *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.
Gilbert, Daniel. 2006. *Stumbling on Happiness*. New York: Alfred A. Knopf.
Gilbert, Daniel T., and Patrick S. Malone. 1995. "The Correspondence Bias." *Psychological Bulletin* 117 (1): 21.
Gilovich, Thomas, Dale Griffin, and Daniel Kahneman, eds. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.
Gintis, H. 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton, N.J.: Princeton University Press.
Goel, Sharad, Roby Muhamad, and Duncan Watts. 2009. "Social Search in Small-World Experiments." Pp. 701–10 in *Proceedings of the 18th International Conference on the World Wide Web*. New York: ACM. doi:10.1145/1526709.1526804.
Goldman, Alvin I. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford: Oxford University Press.
Goldthorpe, John H. 1998. "Rational Action Theory for Sociology." *British Journal of Sociology* 49 (2): 167–92.
Gordon, Robert M. 1986. "Folk Psychology as Simulation." *Mind and Language* 1 (2): 158–71.
Granovetter, M. S. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83 (6): 1420–43.
Green, Donald P., and Ian Shapiro. 1994. *Pathologies of Rational Choice Theory*. New Haven, Conn.: Yale University Press.
———. 2005. "Revisiting the Pathologies of Rational Choice." Pp. 51–99 in *The Flight from Reality in the Human Sciences*, edited by Ian Shapiro. Princeton, N.J.: Princeton University Press.
Gross, Neil. 2009. "A Pragmatist Theory of Social Mechanisms." *American Sociological Review* 74 (3): 358–79.
Harrison, Glenn W., and John A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42 (4): 1009–55.
Harsanyi, John C. 1969. "Rational Choice Models of Political Behavior vs. Functionalist and Conformist Theories." *World Politics* 21 (4): 513–38.
Hedström, Peter. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
———. 2006. "Experimental Macro Sociology: Predicting the Next Best Seller." *Science* 311 (5762): 786–87.

Hedström, Peter, and Charlotta Stern. 2008. "Rational Choice and Sociology." In *The New Palgrave Dictionary of Economics*, edited by Steven Durlauf and Lawrence E. Blume. New York: Palgrave Macmillan.
Hedström, Peter, and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36:49–67.
Hempel, Carl G., and Paul Oppenheim. 1948. "Studies in the Logic of Explanation." *Philosophy of Science* 15 (2): 135.
Hodgson, Geoffrey M. 2007. "Institutions and Individuals: Interaction and Evolution." *Organization Studies* 28 (1): 95–116.
Hollis, Martin. 1977. *Models of Man: Philosophical Thoughts on Social Action*. New York: Cambridge University Press.
Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2 (8): e124.
Jacob, Brian A., and Steven D. Levitt. 2003. "Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating." *Quarterly Journal of Economics* 118 (3): 843–77.
James, William. 1909. *Pragmatism*. New York: Longmans, Green.
Kalugin, Vladimir. 2006. "Donald Davidson." *Internet Encyclopedia of Philosophy*. http://www.iep.utm.edu/davidson/.
Kirman, Alan D. 1992. "Whom or What Does the Representative Individual Represent?" *Journal of Economic Perspectives* 6 (2): 117–36.
Kiser, Edgar, and Michael Hechter. 1998. "The Debate on Historical Sociology: Rational Choice Theory and Its Critics." *American Journal of Sociology* 104 (3): 785–816.
Kohavi, Ron. 1995. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." Pp. 1137–45 in *The International Joint Conference on Artificial Intelligence*, vol. 14, no. 2. San Francisco: Morgan Kaufmann.
Kopell, N. 1988. "Toward a Theory of Central Pattern Generators." Pp. 369–413 in *Neural Control of Rhythmic Movement in Vertebrates*, edited by S. Rossignol, A. H. Cohen, and S. Grillner. New York: John Wiley.
Lakatos, Imre. 1980. *The Methodology of Scientific Research Programmes*. Vol. 1, *Philosophical Papers*. Cambridge: Cambridge University Press.
Lazarsfeld, Paul F. 1949. "The American Solider: An Expository Review." *Public Opinion Quarterly* 13 (3): 377–404.
Lazer, David, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, and Myron Gutmann. 2009. "Life in the Network: The Coming Age of Computational Social Science." *Science* 323 (5915): 721.
Leamer, Edward E. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73 (1): 31–43.
Levitt, Steven D., and Stephen J. Dubner. 2006. *Freakanomics: A Rogue Economist Explores the Hidden Side of Everything*. London: Penguin.
Lieberson, Stanley, and Freda B. Lynn. 2002. "Barking Up the Wrong Branch: Scientific Alternatives to the Current Model of Sociological Science." *Annual Review of Sociology* 28 (1): 1–19.
Lindenberg, S. 1989. "Social Production Functions, Deficits, and Social Revolutions: Prerevolutionary France and Russia." *Rationality and Society* 1 (1): 51–77.
Lukes, Steven. 1968. "Methodological Individualism Reconsidered." *British Journal of Sociology* 19 (2): 119–29.
Macy, Michael W. 1993. "Backward-Looking Social Control." *American Sociological Review* 58:819–36.
Macy, Michael W., and Andreas Flache. 2002. "Learning Dynamics in Social Dilemmas." *Proceedings of the National Academy of Science, U.S.A.* 99:7229–36.

Madigan, David, Paul E. Stang, Jesse A. Berlin, Martijn Schuemie, J. Marc Overhage, Marc A. Suchard, Bill Dumouchel, Abraham G. Hartzema, and Patrick B. Ryan. 2014. "A Systematic Statistical Approach to Evaluating Evidence from Observational Studies." *Annual Review of Statistics and Its Application* 1:11–39.
Mahoney, James. 2000. "Strategies of Causal Inference in Small-N Analysis." *Sociological Methods and Research* 28 (4): 387–424.
Malpas, Jeff. 1996. "Donald Davidson." *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/friends/entries/davidson/.
Manis, Jerome G. 1972. "Common Sense Sociology and Analytic Sociology." *Sociological Focus* 5 (3): 1–15.
Manski, Charles F. 2007. *Identification for Prediction and Decision*. Cambridge, Mass.: Harvard University Press.
Marcus, Gary. 2008. *Kluge: The Haphazard Construction of the Human Mind*. New York: Houghton Mifflin.
Marini, Margaret M., and Burton Singer. 1988. "Causality in the Social Sciences." *Sociological Methodology* 18:347–409.
Mason, Winter, Siddharth Suri, and Duncan J. Watts. 2014. "Long-Run Learning in Games of Cooperation." In *Proceedings of the 15th ACM Conference on Economics and Computation*. New York: ACM.
Mathisen, James A. 1989. "A Further Look at 'Common Sense' in Introductory Sociology." *Teaching Sociology* 17 (3): 307–15.
Mayhew, Bruce H. 1980. "Structuralism versus Individualism. Part 1: Shadowboxing in the Dark." *Social Forces* 59 (2): 335–75.
McFadden, Daniel. 1999. "Rationality for Economists?" *Journal of Risk and Uncertainty* 19 (1–3): 73–105.
Merton, Robert K. 1968. "Social Theory and Social Structure." In *Social Theory and Social Structure*. New York: Free Press.
Milgram, Stanley. 1969. *Obedience to Authority*. New York: Harper & Row.
Mill, J. S. 1875. *A System of Logic*. 9th ed. London: Longmans, Green.
Mitchell, Gregory. 2004. "Case Studies, Counterfactuals, and Causal Explanations." *University of Pennsylvania Law Review* 152 (5): 1517–608.
Morgan, Stephen L., and Christopher Winship. 2007. *Counterfactuals and Causal Inference: Methods and Principles for Social Research*. Cambridge: Cambridge University Press.
Muchnik, Lev, Sinan Aral, and Sean J. Taylor. 2013. "Social Influence Bias: A Randomized Experiment." *Science* 341 (6146): 647–51.
North, A. C., D. J. Hargreaves, and J. McKendrick. 1997. "In-Store Music Affects Product Choice." *Nature* 390:132.
Olson, Mansur. 1965. *The Logic of Collective Action*. New York: Schocken.
Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. "Discrimination in a Low-Wage Labor Market: A Field Experiment." *American Sociological Review* 74 (5): 777–99.
Parsons, Talcott, and Edward A. Shils. 1951. *Toward a General Theory of Action*. Cambridge, Mass.: Harvard University Press.
Provost, Foster, and Tom Fawcett. 2013. *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. Sebastapol, Calif.: O'Reilly Media.
Quadagno, Jill, and Stan J. Knapp. 1992. "Have Historical Sociologists Forsaken Theory? Thoughts on the History/Theory Relationship." *Sociological Methods and Research* 20 (4): 481–507.
Rosenfeld, Sophia A. 2011. *Common Sense*. Cambridge, Mass.: Harvard University Press.
Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688.

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311 (5762): 854–56.
Sarle, Warren S. 1995. "Stopped Training and Other Remedies for Overfitting." Pp. 352–60 in *Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics*. Fairfax Station, Va.: Interface Foundation of North America.
Schacter, Daniel L. 2001. *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Boston: Houghton Mifflin.
Schelling, Thomas C. 1978. *Micromotives and Macrobehavior*. New York: Norton.
Schrodt, Philip A. 2013. "Seven Deadly Sins of Contemporary Quantitative Political Analysis." *Journal of Peace Research* 51:139–44.
Skocpol, Theda. 1979. *States and Social Revolutions: A Comparative Analysis of France, Russia, and China*. New York: Cambridge University Press.
Scriven, Michael. 1959. "Explanation and Prediction in Evolutionary Theory." *Science* 130 (3374): 477–82.
Sewell, William H. 1996. "Historical Events as Transformations of Structures: Inventing Revolution at the Bastille." *Theory and Society* 25 (6): 841–81.
Sloman, Steven. 2009. *Causal Models: How People Think about the World and Its Alternatives*. Oxford: Oxford University Press.
Sobel, Michael E. 2000. "Causal Inference in the Social Sciences." *Journal of the American Statistical Association* 95 (450): 647–51.
Somers, Margaret R. 1998. "'We're No Angels': Realism, Rational Choice, and Relationality in Social Science." *American Journal of Sociology* 104 (3): 722–84.
Sorensen, Alan T. 2007. "Bestseller Lists and Product Variety." *Journal of Industrial Economics* 55 (4): 715–38.
Stark, R. 1996. *The Rise of Christianity: A Sociologist Reconsiders History*. Princeton, N.J.: Princeton University Press.
Stouffer, Samuel A. 1947. "Sociology and Common Sense: Discussion." *American Sociological Review* 12 (1): 11–12.
Strogatz, Steven H. 2003. *Sync: The Emerging Science of Spontaneous Order*. New York: Theia.
Sunstein, Cass R. 2003. "What's Available? Social Influences and Behavioral Economics." *Northwestern University Law Review* 97 (3): 1295–314.
Taylor, Carl C. 1947. "Sociology and Common Sense." *American Sociological Review* 12 (1): 1–9.
Tetlock, Philip. 2005. *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, N.J.: Princeton University Press.
Tilly, Charles. 1984. *Big Structures, Large Processes, Huge Comparisons*. New York: Russell Sage.
Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124–31.
Ugander, Johan, Brian Karrer, Lars Backstrom, and Cameron Marlow. 2011. "The Anatomy of the Facebook Social Graph." arXiv preprint arXiv:1111.4503.
van de Rijt, Arnout, Soong Moon Kang, Michael Restivo, and Akshay Patil. 2014. "Field Experiments of Success-Breeds-Success Dynamics." *Proceedings of the National Academy of Sciences* 111 (19): 6934–39.
Walt, Stephen M. 1999. "Rigor or Rigor Mortis? Rational Choice and Security Studies." *International Security* 23 (4): 5–48.
Watts, D. J., and S. H. Strogatz. 1998. "Collective Dynamics of 'Small-World' Networks." *Nature* 393 (6684): 440–42.
Weber, Max. 1968. *Economy and Society: An Outline of Interpretive Sociology*. New York: Bedminster Press.

Whitford, Josh. 2002. "Pragmatism and the Untenable Dualism of Means and Ends: Why Rational Choice Theory Does Not Deserve Paradigmatic Privilege." *Theory and Society* 31 (3): 325–63.

Wimmer, Heinz, and Josef Perner. 1983. "Beliefs about Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception." *Cognition* 13 (1): 103–28.

Winfree, Art T. 2000. *The Geometry of Biological Time*. New York: Springer.

Wippler, R., and S. Lindenberg. 1987. "Collective Phenomena and Rational Choice." Chap. 5 in *The Micro-Macro Link*, edited by Jeffrey Alexander, Bernard Giesen, Richard Münch, and Neil J. Smelser. Berkeley and Los Angeles: University of California Press.

Woodward, James. 2005. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Young, Cristobal. 2009. "Model Uncertainty in Sociological Research: An Application to Religion and Economic Growth." *American Sociological Review* 74 (3): 380–97.

Zelditch, Morris. 1969. "Can You Really Study an Army in the Laboratory?" Pp. 528–39 in *Complex Organizations*, 2d ed. Edited by Amitai Etzioni. New York: Holt, Rinehart, & Winston.