Studying mechanisms to strengthen causal inferences in quantitative research[1]

Peter Hedström
Nuffield College
University of Oxford

1. Introduction.

Social scientists and philosophers of science have paid considerable attention to the logic and use of so-called mechanism-based explanations. In this chapter I focus on the role of mechanisms in explanatory social-science research, and I highlight various ways by which the study of mechanisms can make quantitative research more useful for causal inference.

Let me emphasize at the outset that as the terms are being used here, causal inference is not the same as statistical inference. The two types of inference are similar in that they both use "localized" information to draw conclusions about more general phenomena; however the types of phenomena about which one seeks to generalize are not the same and the types of information used also often differ. In statistical inference, one typically uses information obtained from a limited number of observations—usually based on a random sample—to draw conclusions about the likely value of some *parameter* in the population at large such as a regression coefficient or a standard deviation. In

[1] This paper is to be published in J. M. Box-Steffensmeier, H. E. Brady and D. Collier (eds.) *The Oxford Handbook of Political Methodology*. Oxford: Oxford University Press. It draws upon and further develops some of the core themes in Hedström (2005) and Hedström and Swedberg (1996; 1998).

causal inference, as the term is being used here, the information being used is not necessarily confined to a specific sample, and the entity one seeks to generalize about is the *process* by which something has been brought about and the *mechanisms* governing this process. As discussed below, knowing the relevant mechanisms is often important for statistical inference, and statistical estimates and inferences are important for understanding the process by which something has been brought about, but statistical inference and causal inference are different kinds of activities.

The type of "mechanism approach" discussed in this chapter also differs in another important respect from more traditional quantitative approaches. The focus is not on relationships between variables, but on actors, their relationships, and the intended and unintended outcomes of their actions. Properties of actors and/or their social environments often influence the outcomes of individuals' actions. These properties as well as the action outcomes can be measured and represented in the form of variables, but the causality does not operate at the variable level. As discussed in some detail below, since actors, by themselves or in consort with other actors, are the agents of change, causal processes should typically be specified at the actor level.

The chapter is organized as follows. First I describe what I mean by mechanisms and mechanism-based explanations. Then, I focus on three aspects of the role of mechanisms in causal and statistical inference: (i) how an understanding of the mechanisms at work can improve statistical inference by guiding the specification of the statistical models to be estimated; (ii) how mechanisms can

strengthen causal inferences by improving our understanding of why individuals do what they do; and (iii) how mechanism-based models can strengthen causal inferences by showing why, acting as they do, individuals bring about the social outcomes they do.[2]

2. Mechanisms and mechanism-based explanations.

The core idea behind the mechanism approach is that we explain by specifying the mechanisms by which that which we wish to explain was brought about. As Elster (1989: 3-4) has expressed it: "To explain an event is to give an account of why it happened. Usually… this takes the form of citing an earlier event as the cause of the event we want to explain…. [But] to cite the cause is not enough: the causal mechanism must also be provided, or at least suggested."[3]

As observed by Mahoney (2001), however, there is no consensus on what is meant by a "mechanism." In fact, at least at a superficial level, there appears to be almost an overabundance of definitions. Philosophers and social scientists have defined the mechanism concept in numerous different ways (e.g. Bhaskar 1978; e.g. Bunge 1996;

---

[2] In this chapter, "social" outcomes/phenomena and "macro-level" outcomes/phenomena are used as synonymous terms. They refer to collective properties that are not definable by reference to any single member of the collectivity. Important examples of such properties include: (1) typical actions, beliefs, or desires among the members of the collectivity; (2) distributions and aggregate patterns such as spatial distributions and inequalities; (3) topologies of networks that describe relationships among the members of the collectivity; and (4) formal and informal rules that constrain the actions of the members of the collectivity. For a similar definition, see Carlsson (1968).

[3] Space limitations do not allow for a discussion of other notions of what "explaining" is all about. See Hedström (2005) as well as the chapter by Brady in this volume for a discussion of the alternative notions that appear most relevant for the social sciences.

Elster 1999; Gambetta 1998; Glennan 1996; Hedström and Swedberg 1998; Karlsson 1958; Little 1991; Mahoney 2001; Mayntz 2004; Mcadam, Tarrow and Tilly 2001; Pawson 2000; Salmon 1984; Schelling 1998). Table 1 describes some of the most frequently cited definitions.

Author	*Definition*	*References*
Bunge	A mechanism is a process in a concrete system which is capable of bringing about or preventing some change in the system.	Bunge (1997; 2004)
Craver	Mechanisms are entities and activities organized such that they are productive of regular changes from start to finish.	Craver (2001), Machamer, Darden and Craver (2000)
Elster (I)	A mechanism explains by opening up the black box and showing the cogs and wheels of the internal machinery. A mechanism provides a continuous and contiguous chain of causal or intentional links between the explanans and the explanandum.	Elster (Elster 1983; Elster 1989)
Elster (II)	Mechanisms are frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions.	Elster (1998; 1999)
Hedström & Swedberg	A social mechanism is a precise, abstract, and action-based explanation which shows how the occurrence of a triggering event regularly generates the type of outcome to be explained.	Hedström and Swedberg (1996; 1998)
Little	A causal mechanism is a series of events governed by law-like regularities that lead from the explanans to the explanandum.	Little (1991)
Stinchcombe	A mechanism is a piece of scientific reasoning which gives knowledge about a component of another, ordinarily higher-level theory.	Stinchcombe (1991)

Table 1. Alternative mechanism definitions.

These definitions differ a great deal from one another. Some refer to causal mechanisms in general, while others refer exclusively to mechanisms of relevance for the social sciences; some are rather precise while others are broad and general; some refer to concrete existing entities while others refer to models or reasoning about such entities. However, despite all of these differences, they share an important common denominator. Underlying them all is an emphasis on making intelligible the regularities being observed by specifying in detail how they were brought about.

As I have argued in Hedström (2005), the currently most satisfactory conceptual analysis of the mechanism concept is found in Machamer, Darden and Craver (2000). The spirit of their approach is very similar to the "cogs-and-wheels" approach of Elster I and the "social mechanism" approach of Hedström and Swedberg. Following these leads, mechanisms can be said to consist of *entities* (with their properties) and the *activities* that these entities engage in, either by themselves or in concert with other entities. These activities bring about change, and the type of change brought about depends upon the properties of the entities and the way in which the entities are linked to one another. A mechanism, thus defined, refers to a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome, and we explain an observed outcome by referring to the mechanism by which such outcomes are regularly brought about.

As mentioned above, actors and actions play a privileged role in social-science explanations because actors are the entities and actions are

the activities that bring about change. Needless to say, a focus on actors and actions does not mean that extra-individual entities such as "cultures," "institutions," or "social structures" are unimportant; only that actors are needed in order for these types of entities to have any effects. Let us take social structure as an example, and thereby also highlight the fact that social science explanations often refer to mechanisms nested within other mechanisms (see Stinchcombe 1991). Following the social-network tradition in sociology (such as White, Boorman and Breiger 1976), a social structure here is defined as a pattern of interactions among a set of individuals. As shown in numerous studies, these patterns of interactions are likely to influence the outcomes that the individuals embedded in the structure are likely to bring about (see Strang and Soule 1998 for a review of some of the relevant literature). That is to say, the same entities (individual actors) strung together in different relational patterns can regularly be expected to bring about different types outcomes. In this sense, different types of relational configurations can be said to constitute different mechanisms (see also Mcadam, Tarrow and Tilly 2001).

The way in which "structural" mechanisms like these influence aggregate outcomes can be seen most clearly in the study of infectious diseases. As Watts and Strogatz (1998) have shown, even very small changes in the structure of a network can decisively influence the probability that an epidemic will take off. Demonstrating such effects is more difficult in the social sciences because the transmission mechanisms are more complex, but the logic of the processes is similar. Whether the outcome to be explained refers to the size of an epidemic or the extent to which a certain belief is held in the population at large, we explain it (in part) by referring to the

mechanism (that is, in this case, to a specific type of network structure) that regularly brings about such outcomes.

Networks, however, cannot produce outcomes in and of themselves. If they are to have any effect, they must influence the ways in which individuals act, and in order to understand how this influence operates, we must specify the mechanisms involved. That is to say, we must explicate the mechanisms that explain the actions of individuals, and which are nested within these "structural" mechanisms. These types of action-related mechanisms may also be characterized in terms of their entities (and their properties) and the ways in which the entities are linked to one another. The core entities are different, however, and now include entities such as beliefs, desires, and opportunities of the actors. But the explanatory logic is the same: we explain an observed phenomenon, in this case an individual action, by referring to the mechanism (that is, the constellation of beliefs, desires, opportunities, etc) by which such actions are regularly brought about.

Why is it so important that we identify the mechanisms that appear to generate the outcomes we observe, whether they be the actions of individuals or collective outcomes that result from the collective or sum-total of the actions of numerous individuals? For one thing, identifying the details of the mechanisms tend to produce explanations that are more precise and intelligible. In other words, we can only really understand and explain what we observe by referring to the mechanisms involved.

Another important reason is that focusing on mechanisms tends to reduce theoretical fragmentation. For example, we may imagine

numerous theories (of voting, social movements, etc.) that are all based on the same set of mechanisms of action and interaction. By focusing on mechanisms we may avoid unnecessary proliferation of theoretical concepts. We may also be able to place in relief structural similarities between processes that at first glance seem completely dissimilar.

Finally, an understanding of the mechanism involved in an outcome is what permits us to conclude that we are dealing with a genuine causal relationship and not simply a correlation. As Glennan (1996:65) has emphasized, "two events are causally connected when and only when there is a mechanism connecting them." Without the ability to identify such a mechanism we cannot conclude with any certainty that an observed regularity is indicative of a genuine causal relationship.

3. Mechanisms and quantitative research.

Survey analysis and the statistical techniques needed for analyzing survey data are invaluable tools. However, as many observers have noted, the use of such data and methods also has contributed to a variable-centered form of analysis that pays scant attention to the processes that are likely to explain the outcomes observed. Coleman (1986) aptly described this type of variable-focused research as a form of "individualistic behaviorism." The guiding principle behind such research is the idea that individual behavior of whatever sort can be explained by various individual and environmental "determinants" such as age, gender, class, and ethnicity. The hope is that by regressing the outcome variable of interest on a set of variables like these one can

identify the causes of the behavior in question. According to Coleman, this type of "causal" explanation of behavior represented a considerable change in the type of explanatory approach used by most social scientists: "One way of describing this change is to say that statistical association between variables has largely replaced meaningful connection between events as the basic tool of description and analysis" (Coleman 1986:1327-8).

David Freedman has discussed the statistical foundations of this type of causal modeling approach in some detail (e.g. Freedman 1987; 1991; 1999). According to Freedman, the belief of some social scientists in the possibility of estimating causal models is counterproductive. The basic statistical tools of causal modelers, Freedman argues, are based upon a network of highly restrictive stochastic assumptions that are rarely, if ever, met. The basic thrust of Freedman's argument is that social scientists need to think more about underlying social processes, and to look more closely at data unencumbered by the distorting lens of conventional and, according to Freedman, largely irrelevant stochastic models: "In my opinion, the confusion between descriptive and structural models pervades the social-science scholarly literature of the past 20 years, and has distorted the research agenda of a generation." (Freedman 1992: 123). As Freedman has emphasized, at best, such statistical analyses can provide compact descriptive summaries of data, but they cannot in themselves provide causal explanations. Causal inferences should always ride on the strength of the argument, and not on "the magic of least squares" (Freedman 1992).

So where does this leave us? I do not wish to suggest that quantitative empirical research is of minor importance for the social sciences. Quite the contrary. But in order for statistical analyses to provide reliable and useful information, the statistical model must represent reasonably well the process by which the outcome of interest was brought about. In order to arrive at such models, as Sørensen (1998) has argued, we need a different division of labor between social scientists and statisticians than the one that currently predominates. The proper division should be one in which social-science theory suggests a mathematical model of a causal process and statistics provides the tools to estimate this model, not, as is common today, that statistics provides models that social scientists use as ad hoc models of social processes (see also Cox 1992; Goldthorpe 2000). One important way in which the study and identification of mechanisms can help to strengthen quantitative research is by providing the basic building blocks of such subject-specific generative models. I will return to this after giving some concrete examples of the types of mechanisms that are likely to be relevant.

Another important way in which a focus on mechanisms can help to strengthen quantitative research is by allowing the identification of models that otherwise could not be identified. Winship and Harding (2005) have shown how this can be done in general, and they illustrate the principles involved with reference to so-called age, period, cohort (APC) models. As is well known, linear APC models cannot be identified because they are exact linear functions of one another, i.e., Age = Period – Cohort. Winship and Harding's core idea is that identification can be achieved by extending the models to include variables that represent the mechanisms by which age, period,

and cohort are assumed to influence the outcome to be explained. They have shown how Pearl's (2000) so-called front-door criteria can be used to deal with situations in which there is a perfect linear dependence between some subset of the independent variables. The inclusion of variables representing the relevant mechanisms amounts to expanding the APC model into multiple APC models and this makes it possible to identify the parameters of interest in the original APC model.

4. Mechanisms and the explanation of individual actions.

In addition to being a guide to the types of statistical models that are possible and meaningful to estimate, a focus on mechanisms strengthens causal inferences by helping us to understand why individuals do what they do. Even if we are exclusively interested in large-scale social and political processes, understanding why individuals do what they do is of fundamental importance because it is individuals' actions that make society "tick."

To illustrate what a mechanism-based explanation of action is all about, I will take my point of departure in the so-called DBO theory. According to this theory, desires (D), beliefs (B), and opportunities (O) are the primary entities upon which the analysis of action and interaction is based. That is to say, we can understand why individuals do what they do if we perceive of their behavior as being endowed with meaning, that is, that there is an intention that explains why they do what they do (see Elster 1983; von Wright 1971).

Beliefs and desires are mental events that can be said to cause an action in the sense of providing reasons for the action. A particular combination of desires and beliefs constitutes a "compelling reason" for performing an action. They have a motivational force that allows us to understand, and in this respect to explain the action (see von Wright 1989).

Individuals do not act in isolation from one another, however. In order to explain why they do what they do, we must also seek to understand how their beliefs, desires, and opportunities are formed in interaction with other individuals. Simply assuming that beliefs and desires are fixed and unaffected by the actions of others may be plausible in some very specific situations, but it would be an untenable assumption in the general case. Ignoring or misrepresenting the ways in which individuals interact and influence one another often leads to flawed predictions about the social outcomes that a group of individuals are likely to bring about (see e.g., Holland 1998 for a general discussion). Therefore, we must try to specify mechanisms by which the actions of some may come to influence the beliefs, desires, opportunities, and actions of others, and in the dyadic case we can describe the interaction between two actors as in Figure 1.

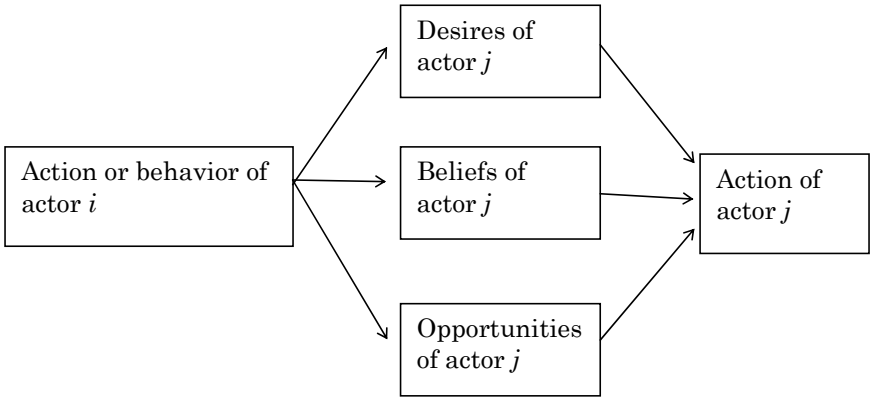


Figure 1. Dyadic interaction between actor *i* and actor *j* according to the DBO theory

To the extent that the action of one actor, here referred to as actor *i*, influences the action of another, actor *j*, this influence must be mediated via the action opportunities or mental states of actor *j*. In terms of the DBO theory, the action (or behavior) of *i* can influence the desires, the beliefs, or the opportunities of actor j and thereby the actions of *j*.[4]

Using these basic ideas, we may define more complex "molecular" mechanisms. These molecular mechanisms differ from one another in terms of how these basic entities and activities are linked. Some examples can be seen in Figure 2. The letters D, B, O, and A stand for desires, beliefs, opportunities, and actions, and the letters *i*, *j*, and *k* identify different actors.

[4] "Actor *i*" may not be a single actor but could be a small group of actors with whom *j* interacts, or a "generalized other" representing a typical actor as perceived by actor *j*.

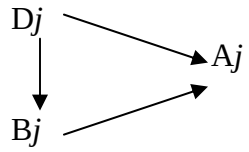
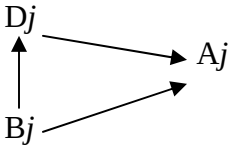
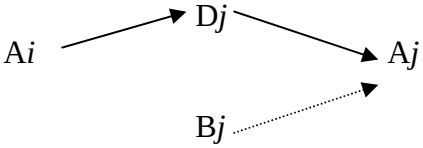
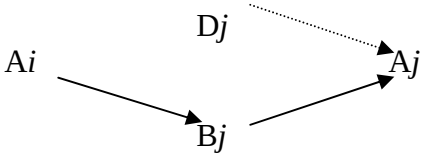
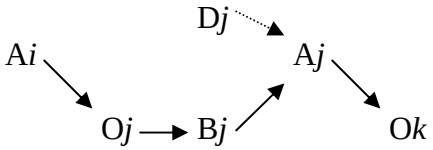
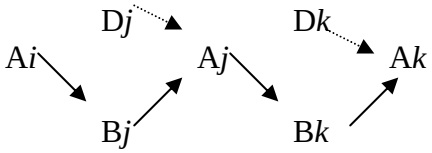
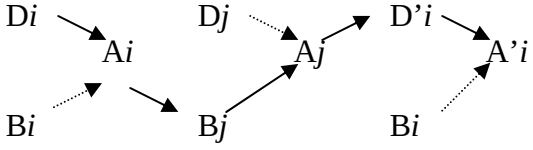
Entities and activities:	Structural pattern:	Type of mechanism:
Mental states, opportunities, and actions of a single individual		Wishful thinking. See e.g. Davidson (1980)
– " –		Sour-grapes syndrome. See e.g. Elster (1983)
Mental states, opportunities, and actions of two or more individuals		Dissonance-driven desire formation. See e.g. Festinger (1957)
– " –		Rational-imitation. See e.g. Hedström (1998)
– " –		Vacancy chain. See e.g. White (1970)
– " –		Self-fulfilling prophecy. See e.g. Merton (1968).
– " –		The "Old Regime" pattern. See Tocqueville (1998)

Figure 2. Examples of action- and interaction-related mechanisms.

Taking the mechanisms charted above in order, the first pattern of entities and activities exemplifies *wishful thinking*. As used here, the term wishful thinking denotes a connection between an actor's desires and his/her beliefs that makes the actor believe that what (s)he desires is in fact the case (see Davidson 1980). The sour-grapes syndrome, the second type of mechanism, exemplifies the opposite pattern. Here, there is a connection between an actor's beliefs to his/her desires such that the actor only desires what (s)he believes (s)he can get (see Elster 1983).

Dissonance-driven desire formation, is a mechanism in which actions of others lead to a change in the focal actor's desires and thus to a change in his/her actions. Festinger's (1957) notion of cognitive dissonance is a classic example. For example, if I desire *p* but the people I engage with do not, this may cause strong dissonance, particularly if the desire is important to me *and* I value my relationship with these people. One way to eliminate the dissonance would be to persuade them to value *p*. Another, often easier, way would be to "persuade" oneself that *p* was not really as desirable as one initially thought.

Rational imitation may be said to occur when one actor's action influences the beliefs and subsequent actions of others. For example, since it is commonly acknowledged that how crowded a restaurant is says something about the dining experience, a large crowd is likely to

affect the beliefs and actions of potential diners (see Hedström 1998)—unless, of course, they simply want to be alone.

In *vacancy chains*, the actions of some actors create new opportunities and changes in the actions of others. A classic example is Harrison White's (1970) analysis of the vacancy-driven mobility pattern of clergy in the United States. Vacancies occur either when individuals leave their organizations or when new positions are created. When a person fills a vacancy, his/her old position becomes vacant and represents a mobility opportunity to someone else. One of these people will get the job, which will fill that vacancy, but another vacancy is created in this person's old job. This process permits individuals and vacancies to move in different directions, and the mobility process is governed by these chains of opportunity.

The *self-fulfilling prophecy* is a sequential chain of several rational-imitation mechanisms. Merton (1968) focused on the case in which an initially false belief evokes behavior that eventually makes the false belief come true. The example he used was a run on a bank. Once a rumor of insolvency gets started, some depositors are likely to withdraw their savings, acting on the principle that it is better to be safe than sorry. Their withdrawals may hurt the bank's financial viability. But even more importantly, the act of withdrawal in itself *signals* to others that something might be wrong with the bank. This produces even more withdrawals, which further strengthens the belief, and so on. By this mechanism, even an initially sound bank may go bankrupt if enough depositors withdraw their money in the (initially) false belief that the bank is insolvent.

The *"Old Regime" pattern* is a sequential concatenation of rational-imitation and dissonance-driven desire-formation mechanisms (D*i* → A*i* → B*j* → A*j* → D'*i* → A*i* where D'*i* ≠ D*i*). For opportunistic reasons one actor decides to do something (s)he does not genuinely desire. Others see the action and the rational-imitation mechanism makes them follow suit. Eventually this feeds back to the first actor such that the actions taken by others produce dissonance and a change in the desires of the first actor. This causes him or her to desire in fact what (s)he initially only pretended to want. Such a mechanism was used by Tocqueville to explain the rapid secularization that took place in France at the end of the eighteenth century:

> Those who retained their belief in the doctrines of the Church became afraid of being alone in their allegiance and, dreading isolation more than the stigma of heresy, professed to share the sentiment of the majority. So what was in reality the opinion of only a part (though a large one) of the nation came to be regarded as the will of all and for this reason seemed irresistible even to those who had given it this false appearance. (Tocqueville 1998:155)

By detailing such mechanisms we arrive at a better understanding of why individuals do what they do. We also have taken an important first step towards understanding how the large-scale social and political outcomes that most of us are particularly interested in explaining are brought about. But to understand what kind of macro-level outcomes micro-processes like these are likely to generate is far

from trivial. In the next section, I will briefly discuss some useful approaches for addressing this link between micro and macro.

5. Mechanisms and the explanation of macro-level outcomes.

As emphasized by Coleman (1986), one of the main hurdles to the development of explanatory theory has been the difficulty of linking individual actions to the social or macro-level outcomes they are likely to bring about. Small and seemingly unimportant changes in the type of mechanisms discussed in the previous section can have considerable consequences for the aggregate social outcomes that a group of individuals is likely to bring about. For this reason, social outcomes cannot simply be "read off" from the properties of the actors that generate them. Even in very small groups in which actors act on the basis of known action logics and where the interaction patterns are fully understood, we often fail to anticipate the social outcomes likely to be brought about (see Schelling 1978 for a range of illuminating examples). Anticipating and explaining the link between the individual and the social is simply too complex for us to handle without the use of some formal analytical tools.

Until fairly recently, the most readily available formalism for addressing the link between the individual and the social was some sort of mathematical model, not seldom an equilibrium model imported from economics. One fundamental problem with many of these models is that they force the analyst to introduce intentionally false assumptions because otherwise the model cannot be solved, and a mathematical model that cannot be solved is not of much use. When analyzing the link between micro and macro, it is important not to fall

into the trap of fictionalism or instrumentalism. Although explanatory accounts are always and by necessity descriptively incomplete, relying upon descriptively false accounts must be avoided because they will give incorrect answers to the question of why we observe what we observe.[5] We may, for instance, be able to tell a story that shows how a group of atomistic and rational individuals could have brought about the social outcome we seek to explain, but such an account would not explain why we observed what we observed unless the individuals acted and interacted as postulated by the theory (or at least approximately so). Basing the analysis on descriptively false assumptions about the logic of action and interaction may allow us to formulate elegant and parsimonious models, but if such logics are found only in hypothetical worlds much different from our own, explanations based on such assumptions will be fictional in our world. In order to explain we must refer to mechanisms known to be operative in the real-world settings that we are analyzing (Elster 1989).

In many instances, agent-based modeling is a more attractive type of formalism than traditional mathematical models for addressing the link between the individual and the social. Agent-based modeling uses computer simulations to assess the social outcomes that groups of virtual actors are likely to bring about, and what distinguishes agent-

[5] The difference between descriptively false and descriptively incomplete statements can be defined as follows. If we have a set A = {a b c d} and we assume that A = {e f}, our assumption would be descriptively false, while if we assume that A = {c d}, our assumption would be descriptively incomplete. In the former case we ascribe to A characteristics which it does not have, while in the latter case we assume A to be what it is only in part, that is, we accentuate certain aspects by ignoring others. See Parsons (1937) and Sen (1980) for further discussions of the explanatory importance of these distinctions.

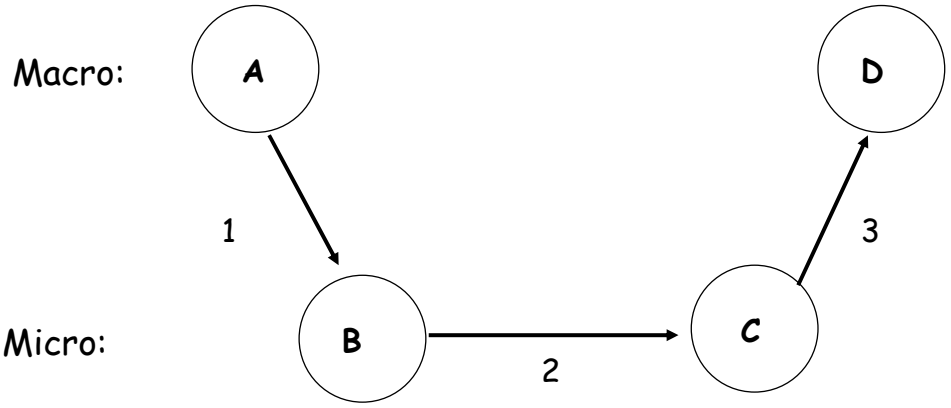
based analyses from other simulation approaches is that they are actor-based, and that they explain social phenomena from the bottom up. That is to say, agent-based models make predictions about social outcomes, given different assumptions about the properties, actions, and interactions among the actors (see Axelrod 1997; Epstein and Axtell 1996; Macy and Willer 2002; and the chapter by Page in the current volume for a range of examples). Such models may lack the elegance and beauty of mathematical models, but they often have more explanatory power because they do not force the analyst to base the analysis on intentionally false assumptions.

Despite the possibilities for bridging the gap between theoretical models and empirical reality that agent-based modeling offer, many agent-based modelers view themselves as pure theorists and pay little or no attention to empirical matters. Macy and Willer, for example, describe agent-based modeling as "a new tool for theoretical research" (2002: 161) and they argue that the core idea behind agent-based modeling is "to perform highly abstract thought experiments that explore plausible mechanisms that may underlie observed patterns" (2002: 147). Similarly, Axelrod argues that agent-based models do not "…aim to provide an accurate representation of a particular empirical application. Instead, the goal of agent-based modeling is to enrich our understanding of fundamental processes that may appear in a variety of applications" (Axelrod 1997: 25).

Although pure thought experiments are of obvious importance, when we are to explain a specific outcome, we must make sure that the mechanisms being modeled have indeed been in operation in this particular case. Furthermore, we must make sure that the more fine-

grained detail of the mechanisms seem to correspond to what we know about the real-world case being studied. Unless there is such correspondence between model and reality, the analysis will only offer an as-if story of little or no explanatory value.

In Hedström (2005), Yvonne Åberg and I illustrated how one can narrow the gap between model and reality by forging a tighter link between quantitative research and agent-based modeling. Coleman's (1986) so-called micro-macro graph can be used to illustrate the principles involved (see Figure 3).



A: Actions of others or other relevant environmental conditions

B: Individual reasons or other properties influencing their actions.

C: Individual action.

D: Social outcome.

Figure 3. Macro-micro linkages.

Most of us are engaged in explaining social or macro-level outcomes (**D** in Figure 3), and usually we explain such outcomes with references to other social or macro-level phenomena (**A**). But simply relating the two to one another, statistically or otherwise, would lead to a rather superficial explanation since the mechanisms that inform us about how and why they are related cannot be found at this aggregate level. We must instead seek to show how individuals' properties and orientations to action are influenced by the social environments in which they are embedded (link 1).[6] Thereafter, we must seek to show how these properties and orientations to action influence how they act (link 2), and how these actions bring about the social outcomes we seek to explain (link 3).

The essence of the approach advocated by Åberg and me is to use large-scale quantitative data to analyze the details of the first two links in the Coleman graph, and then to incorporate the results of these analyses into an agent-based model that formally represents the core mechanisms believed to be at work (see also Bearman, Moody and Stovel 2004 for a similar approach linking quantitative research and agent-based modeling). Although this type of approach differs in certain respects from the modal type of agent-based model, the logic of the analysis remains the same. That is to say, it is the actions of and interactions between the agents that generate the social patterns that emerge, and by altering various aspects of the simulation set-up one ascertains what effects these changes may have on the outcomes.

[6] See Wikström (2007) for a detailed discussion of how such "situational mechanisms" can influence how individuals act.

Space limitations do not allow for a detailed description of this approach, but in all brevity it seeks to closely integrate mechanism-based theories and empirical research, and the way in which this is done can be summarized in the following manner:

1. Start by developing a stylized agent-based model that explicates the logic of the mechanism assumed to be operative. Simulate the model in order to make sure that the model can generate the type of social outcome to be explained. If it can, we have a mechanism-based explanation of the outcome, although the explanation has not yet been empirically verified.
2. For empirical verification, use relevant data to examine the most important bits and pieces of the causal machinery in order to verify that the mechanism actually works as postulated.
3. Make sure that the model generates the type of outcome to be explained even when the model has been modified in the light of (2) and after controls for likely confounders have been introduced.

Only when the explanatory account has passed all of these three stages can we claim to have an empirically verified mechanism-based explanation of a social outcome.

This type of empirically calibrated agent-based modeling can accomplish two important tasks. First, it enables a test of the agent-based model by examining the extent to which the model brings about the social outcome to be explained, even for realistic parameter values. Second, and more important in this context, it strengthens causal inferences in quantitative research. By embedding the results from a statistical analysis in an agent-based model that represents the core mechanisms believed to be at work, it becomes possible to derive

the social outcomes that individuals would bring about were they to act and interact as the results of the empirical analysis suggest that they do.

6. Concluding remarks.

In this chapter I have focused on the logic of mechanism-based explanations and on how the study of mechanisms can strengthen causal inference. A mechanism, as here defined, refers to a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome, and we explain an observed outcome by referring to the mechanism by which such outcomes are regularly brought about. I briefly discussed three ways through which a focus on mechanisms can strengthen causal inferences: (1) it can strengthen causal inferences by helping us to decide on what kind of statistical models are at all meaningful to estimate; (2) it can strengthen causal inferences by helping us to understand why individuals do what they do; and (3) it can strengthen causal inferences by helping us to understand why, acting as they do, they bring about the outcomes they do.

Over the last few years there has been a surge of interest in mechanism-based explanations, in political science as well as in sociology. Most of this work has been important and valuable in that it has sought to clarify the distinctiveness of the approach and to apply it empirically. But some of it has been somewhat problematic in that it threatens to strip the approach of all its distinctive features. If a mechanism-based approach simply becomes synonymous with an approach that is attentive to potential causes or to intervening

variables, as some recent contributions seem to suggest, adopting a mechanism-based vocabulary simply contributes to an unnecessary proliferation of theoretical concepts. In this chapter I have sought to clearly articulate the guiding principles behind the mechanism approach. This approach is abstract, realistic, and precise, and it explains specific political or social phenomena on the basis of explicitly formulated theories of action and interaction.

References.

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

Bearman, P. S., J. Moody, and K. Stovel. 2004. "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks." *American Journal of Sociology* 110:44-91.

Bhaskar, R. 1978. *A Realist Theory of Science*. Hassocks: Harvester Press.

Bunge, M. A. 1996. *Finding Philosophy in Social Science*. New Haven: Yale University Press.

—. 1997. "Mechanism and Explanation." *Philosophy of the Social Sciences* 27:410-465.

—. 2004. "How Does It Work? The Search for Explanatory Mechanisms." *Philosophy of the Social Sciences* 34:182-210.

Carlsson, G. 1968. "Change, Growth, and Irreversibility." *American Journal of Sociology* 73:706-714.

Coleman, J. S. 1986. "Social Theory, Social Research, and a Theory of Action." *American Journal of Sociology* 91:1309-1335.

Cox, D. R. 1992. "Causality. Some Statistical Aspects." *Journal of the Royal Statistical Society Series A -Statistics in Society* 155:291-301.

Craver, C.F. 2001. "Role Functions, Mechanisms, and Hierarchy." *Philosophy of Science* 68:53-74.

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Clarendon Press.

Elster, J. 1983. *Explaining Technical Change: A Case Study in the Philosophy of Science*. Cambridge: Cambridge University Press.

—. 1983. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

—. 1989. *Nuts and Bolts for the Social Sciences*. Cambridge: Cambridge University Press.

—. 1998. "A Plea for Mechanisms." Pp. 45-73 in *Social Mechanisms: An Analytical Approach to Social Theory*, edited by P. Hedstrom and R. Swedberg. Cambridge: Cambridge University Press.

—. 1999. *Alchemies of the Mind: Rationality and the Emotions*. Cambridge: Cambridge University Press.

Epstein, J. M., and R. Axtell. 1996. *Growing Artificial Societies: Social Science from the Bottom Up*. Washington, D.C.: Brookings Institution Press.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Stanford, California.: Stanford University Press.
Freedman, D. 1987. "As Others See Us: A Case Study in Path Analysis." *Journal of Educational Statistics* 12:101-128.
—. 1991. "Statistical Models and Shoe Leather." *Sociological Methodology*:291-258.
—. 1992. "A Rejoinder on Models, Metaphors, and Fables." Pp. 108-25 in *The Role of Models in Nonexperimental Social Science: Two Debates*, edited by J.P. Schaffer. Washington, DC: American Educational Research Association.
—. 1999. "From Association to Causation: Some Remarks on the History of Statistics." *Statistical Science* 14:243-258.
Gambetta, D. 1998. "Concatenations of Mechanisms." in *Social Mechanisms: An Analytical Approach to Social Theory*, edited by P. Hedström and R. Swedberg. Cambridge: Cambridge University Press.
Glennan, S. S. 1996. "Mechanisms and the Nature of Causation." *Erkenntnis* 44:49-71.
Goldthorpe, J. H. 2000. *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.
Hedström, P. 1998. "Rational Imitation." Pp. 306-27 in *Social Mechanisms: An Analytical Approach to Social Theory*, edited by P. Hedström and R. Swedberg. Cambridge: Cambridge University Press.
—. 2005. *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.
Hedström, P., and R. Swedberg. 1996. "Social Mechanisms." *Acta Sociologica* 39:281-308.
—. 1998. "Social Mechanisms: An introductory essay." Pp. 1-31 in *Social Mechanisms: An Analytical Approach to Social Theory*, edited by P. Hedström and R. Swedberg. Cambridge: Cambridge University Press.
Holland, J. H. 1998. *Emergence: From Chaos to Order*. Cambridge, Mass.: Perseus Books.
Karlsson, G. 1958. *Social Mechanisms: Studies in Sociological Theory*. Stockholm: Almqvist & Wiksell.
Little, D. 1991. *Varieties of Social Explanation: An Introduction to the Philosophy of Social Science*. Boulder: Westview Press.
Machamer, P., L. Darden, and C.F. Craver. 2000. "Thinking About Mechanisms." *Philosophy of Science* 67:1-25.
Macy, M. W., and R. Willer. 2002. "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28:143-166.

Mahoney, J. 2001. "Beyond Correlational Analysis: Recent Innovations in Theory and Method." *Sociological Forum* 16:575-593.
Mayntz, R. 2004. "Mechanisms in the Analysis of Social Macro-Phenomena." *Philosophy of the Social Sciences* 34:237-259.
Mcadam, D., S. Tarrow, and C. Tilly. 2001. *Dynamics of Contention*. Cambridge: Cambridge University Press.
Merton, R. K. 1968. "The Self-Fulfilling Prophecy." Pp. 475-490 in *Social Theory and Social Structure*, edited by Robert K. Merton. New York: The Free Press.
Parsons, T. 1937. *The Structure of Social Action*. New York: The Free Press.
Pawson, R. 2000. "Middle-Range Realism." *Archives Europeennes de Sociologie* XLI:283-325.
Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.
Salmon, W. C. 1984. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.
Schelling, T. C. 1978. *Micromotives and Macrobehavior*. New York: W.W. Norton.
—. 1998. "Social Mechanisms and Social Dynamics." Pp. 32-44 in *Social Mechanisms: An Analytical Approach to Social Theory*, edited by P. Hedström and R. Swedberg. Cambridge: Cambridge University Press.
Sen, A. 1980. "Description as Choice." *Oxford Economic Papers* 32:353-369.
Sørensen, A. B. 1998. "Theoretical Mechanisms and the Empirical Study of Social Processes." Pp. 238-66 in *Social Mechanisms: An Analytical Approach to Social Theory.*, edited by P. Hedström, and Swedberg, R. Cambridge: Cambridge University Press.
Stinchcombe, A. L. 1991. "The Conditions of Fruitfulness of Theorizing About Mechanisms in Social-Science." *Philosophy of the Social Sciences* 21:367-388.
Strang, D., and S. A. Soule. 1998. "Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills." *Annual Review of Sociology* 24:265-290.
Tocqueville, A. de. 1998. *The Old Regime and the Revolution*. New York: Anchor Books.
von Wright, G. H. 1971. *Explanation and Understanding*. Ithaca, N.Y.: Cornell University Press.
—. 1989. "A Reply to My Critics." in *The Philosophy of George Henrik von Wright*, edited by P.A. Schlipp and L. E. Hahn. La Salle: Open Court.
Watts, D. J., and S. H. Strogatz. 1998. "Collective Dynamics of 'Small World' Networks." *Nature* 393:440-442.

White, H., S.A. Boorman, and R.L. Breiger. 1976. "Social Structure from Multiple Networks. Blockmodels of Roels and Positions." *American Journal of Sociology* 81:730-780.
White, H.C. 1970. *Chains of Opportunity: System Models of Mobility in Organizations.* Cambridge: Harvard University Press.
Wikström, P-O. 2007. "Individuals, Settings, and Acts of Crime: Situational Mechanisms and the Explanation of Crime." Pp. Chapter 3 in *The Explanation of Crime. Contexts, Mechanisms and Development. Cambridge*, edited by Wikström P-O and R.J. Sampson. Cambridge: Cambridge University Press.
Winship, C., and D.J. Harding. 2005. "A Mechanism Based Strategy for the Identification of Age-Period-Cohort Models." *Sociological Methodology and Research (forthcoming)*.

3. Three variants of PT 4. Causal inference in PT 5. Studying causal mechanisms? 6. When can PT be used, and not used? Department of political science.Â Single case research method that can be used to make withinâ€case inferences about presence/absence of causal mechanisms. Department of political science. 3.Â â€˜the differences between the quantitative and qualitative traditions are only stylistic and are methodologically and substantively unimportant. All good research can be understood â€“ indeed, is best understood â€“ to derive from the same underlying logic of inference.â€™ (King, Keohane and Verba, 1994: 4). Department of political science.