Explaining Social Change: An Analytical Approach*

Peter Hedström

University of Oxford. Nuffield College
peter.hedstrom@nuffield.oxford.ac.uk

Abstract

Analytical sociology is a middle-range approach to sociological theorizing which seeks to develop precise and realistic action-based explanations of various social phenomena. At the heart of the analytical approach is the idea that acceptable explanations must detail the mechanisms through which the phenomena to be explained were brought about. In this article I discuss the most important elements of this approach and I try to illustrate the various components involved in an explanation of social change. The analytical approach adopted here seeks to closely integrate mechanism-based theories of action and interaction, and agent-based computational modelling. The article is organized in the following way: first I present the guiding ideas behind the mechanisms approach. Second I discuss various mechanisms of action and interaction to illustrate what it is all about. Third I use a so-called agent-based model for illustrating how one can link individual-level mechanisms to social outcomes. Fourth I show how important it is to empirically calibrate these types of models. And fifth, and finally, I round it all up with some concluding remarks.

Key words: social change, social mechanisms, analytical sociology, social interaction, social simulation.

Resumen. *Explicando el cambio social: un enfoque analítico*

La sociología analítica es una aproximación a la teorización social «de alcance intermedio» que persigue desarrollar explicaciones precisas, realistas y basadas en la acción acerca de diversos fenómenos sociales. En el núcleo del enfoque analítico se encuentra la idea de que las explicaciones aceptables deben detallar los mecanismos a través de los cuales se produjeron los fenómenos a explicar. En este artículo discuto los elementos más importantes de este enfoque e intento ilustrar los diversos componentes que integran la explicación del cambio social. El enfoque analítico adoptado aquí busca integrar estrechamente las teorías de la acción y la interacción basadas en mecanismos con los modelos computacionales basados en agentes. El artículo se organiza de la siguiente manera: primero, presento las ideas-guía que subyacen al enfoque de los mecanismos. Segundo, discuto varios mecanismos de acción e interacción con objeto de ilustrar cómo funciona el enfoque. Tercero, utilizo un modelo basado en la agencia para ilustrar cómo pueden vincularse los mecanismos de nivel individual con los

resultados sociales. Cuarto, muestro lo importante que resulta el calibrar empíricamente este tipo de modelos. Y, quinto y último, redondeo todo lo anterior con unas observaciones finales.

Palabras clave: cambio social, mecanismos sociales, sociología analítica, interacción social, simulación social.

Contents

Introduction

The history of analytical sociology can be traced back to the works of Max Weber and Alexis de Tocqueville, and the analytical agenda was further developed by prominent mid-twentieth-century sociologists such as Robert Merton. Only in recent decades, however, has a clearly articulated analytical sociology started to emerge. This type of sociology seeks to explain complex social processes by carefully dissecting them and then bringing into focus their most important constituent components. Analytical sociology is a middle-range approach to sociological theorizing which seeks to develop precise and realistic action-based explanations of various social phenomena.

At the heart of the analytical approach is the idea that acceptable explanations must detail the mechanisms through which the phenomena to be explained were brought about. This means that one must clearly explicate the relevant mechanisms, and then seek to derive the social phenomena that groups of actors, acting on the basis of such mechanisms, are likely to bring about. In this article I discuss the most important elements of this approach[1]. The article is organized in the following way:

— First I present the guiding ideas behind the mechanisms approach.
— Second I discuss various mechanisms of action and interaction to illustrate what it is all about.
— Third I use a so-called agent-based model for illustrating how one can link individual-level mechanisms to social outcomes.
— Fourth I show how important it is to empirically calibrate these types of models.
— And fifth, and finally, I round it all up with some concluding remarks.

1. Needless to say, given the constraints imposed by the article format, the discussion will be rather brief. For a detailed account, see Hedström (2005).

What are the guiding ideas behind the mechanisms approach?

One way of identifying the distinguishing characteristics of the mechanism approach is to compare it with other explanatory approaches (see Table 1). The most important alternatives to the mechanism approach are the covering-law approach of Hempel and other philosophers (e.g., 1965), and the statistical approach of many quantitatively oriented social scientists (for a discussion of the logic of statistical explanations, see e.g., King, Keohane and Verba 1994; Lazarsfeld 1955; Salmon 1971).

While the covering-law approach takes the position that an acceptable explanation consists in subsuming the event to be explained under a general causal law, the statistical approach, explicitly or implicitly, sets an equal sign between explanation and predictive accuracy; a variable is said to be explanatory if it is statistically relevant for the event to be explained. In contrast to these approaches, the core idea behind the mechanism approach is that we deductively explain a social phenomenon by referring to a basic constellation of *entities* and *activities* that are spatially and temporally organized in such a way that they regularly bring about the type of phenomena we seek to explain (see Hedström and Swedberg 1998; Machamer, Darden and Craver 2000). We explain an observed phenomenon by referring to the social mechanism by which such phenomena are regularly brought about.

From the view point of sociology, Hempel's approach is of limited relevance because, as far as we know, there exist no strict Hempelian laws in sociology and there are good reasons to suspect that such laws do not exist. The statistical approach, as practiced by most quantitative sociologists, is wanting for another reason. As argued by Boudon (1979), Coleman (1986), Goldthorpe (2000), and others, statistical analyses summarize patterns in data, they do not explain them. From the mechanism perspective, correlations and constant conjunctions do not explain but are observational phenomena that needs to be explained by reference to the mechanisms that brought them into existence.

Table 1. Main types of explanations.

	Covering-law explanations	Statistical explanations	Mechanism explanations
Explanatory principle	To subsume under a causal law	To identify statistically relevant factors	To specify a social mechanism
Key explanatory factors, entities and/or activities	No restrictions, except that the factor must exhibit a law-like relation to the event to be explained	No restrictions, except that the factor must be statistically relevant for the event to be explained	Actors, actions, and the way in which these are temporally and spatially organized

Mechanisms of action and interaction

In sociology the basic *entities* of a mechanism always tend to be actors, and the basic *activities* tend to be the actions of these actors. Through their actions actors make society «tick,» and without their actions social processes would come to a halt. Theories of action are therefore of fundamental importance for explanatory sociological theories, but how should we go about conceptualizing action and interaction? In my view, the most attractive alternative is the so-called DBO theory. According to this theory, desires (D), beliefs (B), and opportunities (O) are the primary theoretical terms upon which the analysis of action and interaction is based. That is to say, the desires, beliefs, and opportunities of an actor are seen as the proximate causes of the actor's action (see Figure 1).

We can understand why actors do what they do if we perceive of their behaviour as being endowed with meaning, that is, that there is an intention explaining why they do what they do (see Elster 1983a; von Wright 1971). As seen in Figure 1, these intentions in turn can be understood in terms of the desires, beliefs, and opportunities of the actor. A *desire* here is defined as a wish or want for something to happen (or not to happen). A *belief* is defined as a proposition about the world held to be true. And, *opportunities*, as the term is used here, is the «menu» of action alternatives available to the actor, that is, the actual set of action alternatives that exists independently of the actor's beliefs about them (for discussions of various aspects of the DBO theory, see Davidson 1980; Elster 1989; Hahn 1973; Lewis 1994).

Beliefs and desires thus 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, explain the action (see von Wright 1989).

The causal efficacy of beliefs, desires, and opportunities, can be illustrated by the following set of everyday examples focusing on the reasons why Mr. Smith did not bring an umbrella today. There are three ideal-typical explanations:

— *Belief-based explanation*: Mr. Smith desires not to get wet and he had an umbrella that he could have brought, but by mistake he read yesterday's
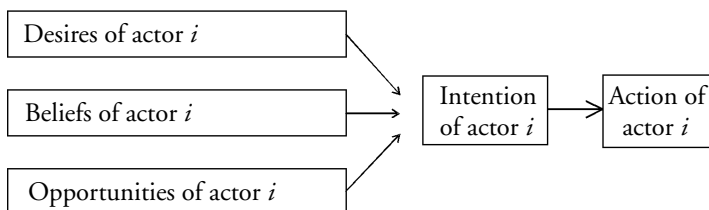

Figure 1. Core components of the DBO theory.

weather column which made him believe that it would not rain today. Therefore he did not bring an umbrella today.

— *Desire-based explanation*: Mr. Smith believed that it would rain today and he had an umbrella that he could have brought, but Mr. Smith has somewhat unusual desires: Walking in heavy rain always makes him feel like Gene Kelly in *Singin' in the Rain*, and feeling like Gene Kelly is something he really desires. Therefore he did not bring an umbrella today.

— *Opportunity-based explanation*: Mr. Smith believed that it would rain today and he had a strong desire not to get wet, but when he was leaving for work in the morning he found that his son had taken his umbrella and there were no other umbrella in the house. Therefore he did not bring an umbrella today.

In order to explain why we observe what we observe, we also must seek to understand how beliefs, desires, and opportunities are formed in interactions with others. 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. Therefore, we must problematize and try to specify the mechanisms through which the actions of some actors may come to influence the beliefs, desires, opportunities, and actions of others[2].

There are numerous ways to conceptualize social-interaction processes, but from the perspective of the DBO theory it appears essential to distinguish between three broad types of social interactions: (1) desire-mediated; (2) belief-mediated; and (3) opportunity-mediated interactions. In the dyadic case we can describe the interaction between two actors as in the Figure 2[3].

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 behaviour) of Actor *i* can influence the desires, the beliefs, or the opportunities of Actor *j* and thereby the actions of *j*[4].

Using the basic concepts of the DBO theory —desires, beliefs, opportunities, actions, and relations— more complex «molecular» mechanisms can be defined. These molecular mechanisms differ from one another in terms of how these basic entities and activities are temporally and spatially organized.

2. I will have little to say about the specifics of these beliefs and desires —whether, for example, one individual desires or believes *p* rather than *q*. The type of interaction mechanisms that I focus on are of a more general kind and deal with the centripetal forces that tend to make interacting individuals coalesce around a certain *p* or a certain *q* whatever that *p* or *q* may be.

3. To simplify the presentation I have removed the «intention box» from the figure, but it is still assumed that intentions intervene between desires, beliefs, and opportunities, on the one hand, and actions, on the other.

4. It is important to note that in many situations, «Actor *i*» may not be a single actor but a small group of actors with whom *j* interacts, or a «generalized other» representing typical actions as perceived by Actor *j*.
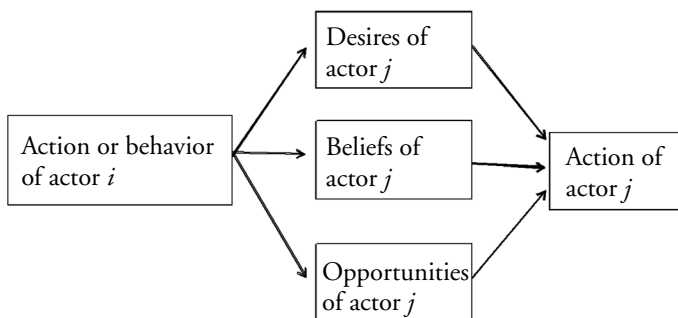
Figure 2. Dyadic interaction between Actor i and Actor j according to the DBO theory.

Some examples can be seen in Figure 3. The letters D, B, O, and A stand for desires, beliefs, opportunities, and actions, and the letters *i*, *j*, and *k* identify different actors.

The first pattern of entities and activities exemplifies *wishful thinking*. As the term is used here, wishful thinking denotes a causal connection from an actor's desires to his/her beliefs that makes the actor believe what (s)he desires to be the case (see Davidson 1980). The second type of mechanism, the *sour-grapes* syndrome exemplifies the opposite causal direction. That is to say, it is a causal connection from an actor's beliefs to his/her desires, which makes the actor desire only what (s)he believes (s)he can get (see Elster 1983b).

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

The fourth mechanism, *rational-imitation*, is the case where one actor's action influences the beliefs and subsequent actions of others. For example: The number of guests at a restaurant is likely to influence other individuals' choice of restaurant because the number of guests is a signal about the quality of the restaurant likely to influence the beliefs and actions of others (see Hedström 1998).

The fifth mechanism, *vacancy chains*, is a pattern where actions of some 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 in US clergy. An important feature of job mobility within organizations, cap-

Entities and activities:	Temporal and spatial pattern:	Type of mechanism:
Mental states, opportunities, and actions of a single individual	$Dj \rightarrow Aj$, $Dj \rightarrow Bj$, $Bj \rightarrow Aj$	Wishful thinking. See e.g. Davidson (1980)
– « –	$Bj \rightarrow Dj$, $Dj \rightarrow Aj$, $Bj \rightarrow Aj$	Sour-grapes syndrome. See e.g. Elster (1983b)
Mental states, opportunities, and actions of two or more individuals	$Ai \rightarrow Dj$, $Dj \rightarrow Aj$, $Bj \rightarrow Aj$	Dissonance-driven desire formation. See e.g. Festinger (1957)
– » –	Dj, $Ai \rightarrow Bj$, $Bj \rightarrow Aj$	Rational-imitation. See e.g. Hedström (1998)
– « –	Ai, Dj, Aj, $Oj \rightarrow Bj$, Ok	Vacancy chain. See e.g. White (1970)
– » –	Ai, Dj, Aj, Bj, Dk, Ak, Bk	Self-fulfilling prophecy. See e.g. Merton (1968).
– « –	Di, Ai, Bi, Dj, Aj, Bj, $D'i$, $A'i$, Bi	The »Old Regime» pattern. See Tocqueville (1998)

Figure 3. Examples of some action- and interaction-related mechanisms.

tured in White's analysis, is that individuals' opportunities are constrained by the number of vacant jobs. Vacancies are created either when individuals leave their organizations or when new positions are created. When an individual fills a vacancy, a new vacancy is created in that person's old job, and this represents a mobility opportunity to others. One of these people will get the job and the vacancy will disappear, but a new vacancy has now been created in this person's old job. Individuals and vacancies thus move in different directions, and the mobility process is governed by these chains of opportunity.

The sixth mechanism, *the self-fulfilling prophecy*, is a sequential concatenation of several rational-imitation mechanisms. Merton (1968) focused on the case in which an initially false belief evokes behaviour that eventually makes the false belief come true. The example he used is a run on a bank. Once a rumour 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 strengthen the beliefs of others that the bank is in financial difficulties, partly because the withdrawals may actually hurt the financial standing of the bank, but more importantly because 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 seventh and final mechanism in Figure 3, *the «Old Regime» pattern*, is a sequential concatenation of rational-imitation and dissonance-driven desire-formation mechanisms ($D'i \rightarrow Ai \rightarrow Bj \rightarrow Aj \rightarrow Di \rightarrow Ai$ where $D'i \neq Di$). For opportunistic reasons one actor decides to do something (s)he does not genuinely desire. The action is observed by others and the rational-imitation mechanism makes them follow suit. Eventually this feeds back on the first actor: The actions of others produce dissonance and a change in the desires of the first actor which makes him or her genuinely desire what (s)he initially only pretended to do. A mechanism like this 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)

Why, then, is it so important to specify the mechanisms that are supposed to have generated observed outcomes? From the perspective of sociological theory, one important reason for insisting on a detailed specification of mechanisms is that it tends to produce precise and intelligible explanations without glaring black boxes (see Boudon 1998). Another important reason is that a focus on mechanisms tends to reduce theoretical fragmentation. For example, we may have numerous different theories (of crime, social movements, or whatnot), that are all based on the same set of mechanisms of action and interaction. Focusing on the mechanisms as such avoids unnecessary proliferation of theoretical concepts and may help in bringing out structural similarities between seemingly disparate processes. Finally, it is the knowledge that the type of outcome we seek to explain regularly is brought about by the entities and activities referred to in the mechanism that gives us reason to believe that

there indeed is a genuine causal relationship between a proposed cause and its effect, and not simply a spurious correlation.

Analyzing the link between the individual and the social

To understand why actors do what they do is not sufficient, however; we must also address the question of why, acting as they do, they bring about the social outcomes they do. Sociology is not a discipline concerned with explaining the actions of single individuals. The focus on actions is merely an intermediate step in an explanatory strategy that seeks to understand change at a social level.

As the term is used here, the «social» refers to collective properties that are not definable for a single member of the collectivity (see also Carlsson 1968). Important examples of such phenomena include:

— typical actions, beliefs, desires etc. among the members of the collectivity;
— distributions and aggregate patterns such as spatial distributions and inequalities;
— topologies of networks that describe relationships between the members of the collectivity; and
— informal rules or social norms that constrain the actions of the members of the collectivity.

Social outcomes like these are *emergent* phenomena. With emergent phenomena I am not referring to any mystic holistic entities with their own causal powers, but simply to social phenomena that are brought about as an unintended by-product of the actions of interacting individuals. Social outcomes, like other emergent phenomena, are difficult to anticipate because the outcome depends to a high degree on how the individual parts are interrelated (see e.g., Holland 1998). Small and seemingly unimportant changes in the way in which actors are interrelated can have profound consequences for the social outcomes that are likely to emerge because the links between the actors influence the extent to which a belief, desire, or social practice spreads through a population (see e.g., Watts and Strogatz 1998 for some striking examples). For this reason, social outcomes cannot simply be «read off» from the properties of the individuals that generate them. In order to explain the social phenomena we observe, we need to develop generative models that show how large numbers of actors, in interaction with one another, over time bring about different types of social outcomes.

Until fairly recently the type of generative model used for addressing the link between the individual and the social was some sort of stylized mathematical model (see Edling 2002 for an overview). One fundamental problem with many of these models is that they force the analyst to introduce knowingly false assumptions because otherwise the model cannot be solved, and a mathematical model that cannot be solved is not of much use. Such models can be extremely elegant and sometimes they represent remarkable intellectual achievements, but it is questionable whether they can be said to explain any-

thing in the real world. As emphasized by Elster (e.g. 1989), genuine explanations of social phenomena always must account for what happens, as it happens, not as it could have happened in a fictional world very different from our own.

Agent-based modelling is important as a tool for analyzing these links between the individual and the social because of its flexibility. Unlike traditional mathematical models they do not force the analyst to base the analysis on knowingly false assumptions. The basic idea behind agent-based modelling simply is to create a set of virtual actors that act and interact according to known principles, and then use computer simulations to examine the social outcomes they bring about (for an overview of the relevant literature on agent-based modelling, see Macy and Willer 2002). Coleman's (1986) so-called micro-macro graph can be used for describing the role of mechanisms and agent-based modelling in explaining social phenomena (see Figure 4).

As mentioned above, as sociologists we are typically not concerned with explaining the actions or behaviours of single individuals. The focus is on larger-scale *social* phenomena characterizing groups of actors, collectivities, or societies at large; that is to say, the focus of interest in on the upper level of the Coleman graph. But the properties of these social phenomena and changes in them over time must always be explained with reference to individuals' actions, since it is individuals, not social entities, which are endowed with causal powers. Simply correlating social phenomena with one another (arrow 4) leads to superficial causal accounts since such exercises do not explicate the generative mechanisms believed to be at work. In order to explain we must try to show how social phenomena influence individuals' beliefs, desires, opportunities etc (arrow 1), how these mental states and opportunities influence individuals' actions (arrow 2), and how actors, through their actions, bring about the social phenomena we seek to explain (arrow 3).

Some of the mechanisms discussed above were concerned with how individuals' desires, beliefs, and opportunities are influenced by the social contexts in which the actors are embedded (arrow 1). Other mechanisms were con-
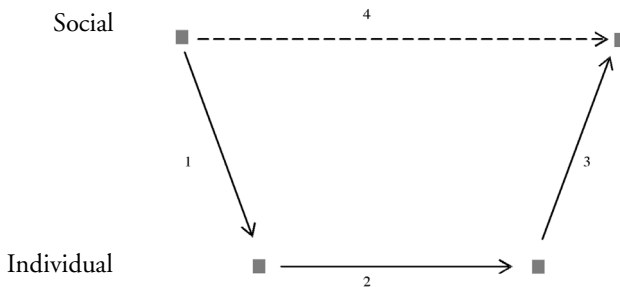

Figure 4. Coleman's micro-macro graph.

cerned with how desires, beliefs, and opportunities, in interaction with one another, influence actions (arrow 2). Agent-based modelling is used for analysing the link between individual actions and social outcomes (arrow 3). That is to say, agent-based modelling is a tool we use for assessing the social outcomes that groups of interacting actors are likely to bring about given the logics of action and interaction expressed in the social mechanisms. We need tools like these because the processes we focus on are so complex that we cannot establish the social outcomes that are likely to be brought about unless we use such generative models.

How would then an agent-based model founded on the DBO theory look like? As discussed above, the cause of an action can be seen as a constellation of desires, beliefs, and opportunities in the light of which the action appears reasonable. If we simplify the notion of desires and beliefs in such a way that they can either be said to be or not to be at hand, the possible patterns of desires, beliefs, opportunities, and actions can be described as in Table 2.

An entry of 1 here indicates the presence of the relevant desire, belief, opportunity, or action, and an entry of 0 indicates its absence. The third pattern, for instance, represents a situation where an actor desires a certain outcome and has the opportunity to perform the relevant action, but does not believe that the action will bring about the desired outcome, and therefore decides not to act.

Of these eight possible DBO-patterns, only the first one will bring about an action, because only in this situation does the actor have the opportunity to act in a way that (s)he believes will bring about the desired outcome.

In order to simplify the presentation, I will focus exclusively on the first four patterns in the table, i.e., I will assume that all actors have the opportunity to act. The agent-based simulation can be characterized in the following way. It focuses on the desires, beliefs, and actions of 2,500 actors that are situated on a lattice with 50 rows and 50 columns. At each point in time the relevant properties of an actor can be described in terms of a desire-belief-action triplet, $<D,B,A>$. If the first entry of the triplet is equal to one, the actor is said to have a «positive» desire, and if the second entry is equal to one the actor

Table 2. DBO-patterns and associated courses of action.

Pattern	Desire	Belief	Opportunity	Action
(1)	1	1	1	1
(2)	0	1	1	0
(3)	1	0	1	0
(4)	0	0	1	0
(5)	1	1	0	0
(6)	0	1	0	0
(7)	1	0	0	0
(8)	0	0	0	0
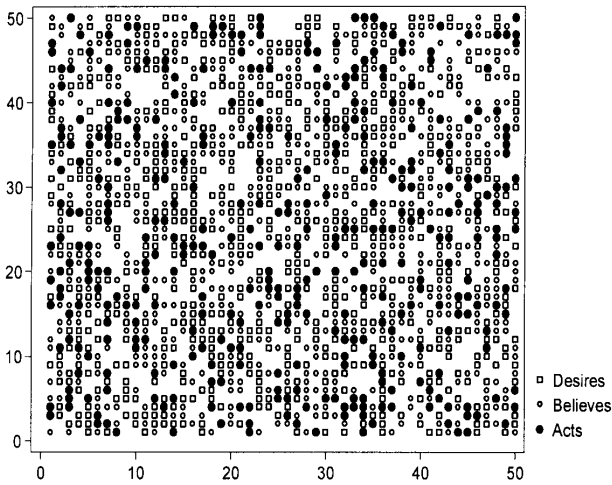
Figure 5. Initial patterns of beliefs, desires and actions in a population of 2,500 actors. Each cell describes the current state of an actor's DBA-triplet.

is said to have a «positive» belief[5]. If these entries are both equal to one, then the third entry will also become equal to one because actors act when they believe that the action will bring about the desired outcome.

We start from a «state of nature» in which the actors' beliefs and desires exhibit no social patterning whatsoever —they are entirely random. A typical initial pattern of desires, beliefs, and actions then looks like the one in Figure 5.

Squares identify actors with positive desires. Circles identify actors with positive beliefs. And, black dots identify actors with positive desires *and* positive beliefs. They are the ones who act because they believe that the action will bring about the desired outcome. The white areas of the graph consist of actors who neither believe in the efficacy of the action nor desire the result, and therefore they do not act. In the figure 40 percent of the actors have positive desires, 40 percent have positive beliefs, and accordingly about 16 percent act because they are the ones who have positive beliefs *and* desires.

A social structure is introduced into the analysis by assuming that each actor directly interacts with the four neighbours described in Figure 6.

If a majority of these four neighbours have a different belief than the focal actor, the focal actor's belief will change. Otherwise it will remain the same. The desires of the actors evolve according to the same logic. Thus, there will be two parallel contagion processes at work, one operating on the beliefs of

5. A positive belief thus means that the actor believes that the action is a good, efficient, and/or appropriate mean of attaining the desired result.
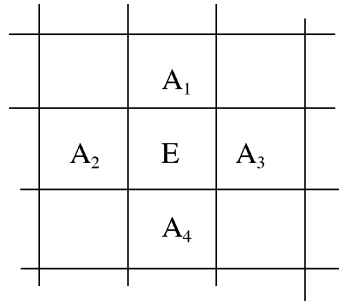
Figure 6. The structure of social interaction between Ego and Alters.
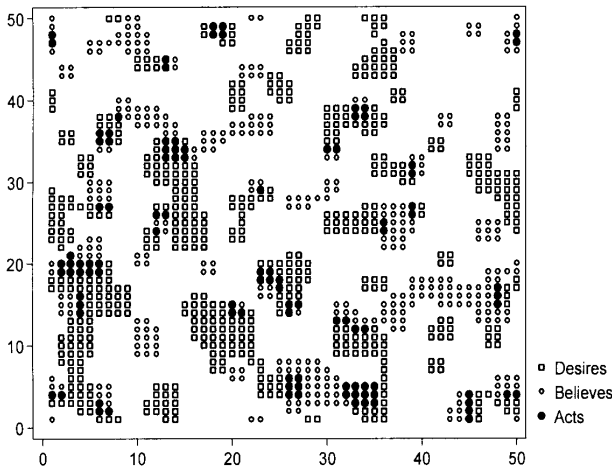

Figure 7. Typical patterns of beliefs, desires and actions in a population of 2,500 actors who socially interact with four neighbours.

the actors and the other on their desires, and actions are the joint outcome of these two processes[6].

Although the point of departure is an entirely random state-of-nature pattern, the interaction process quickly leads to a lock-in on a highly clustered and segregated pattern. Figure 7 is a typical example of the type of pattern that emerges; in this case the actors have interacted for 40 rounds.

6. The lattice used here is a so-called torus, i.e., a lattice which is wrapped around itself in such a way that actors positioned at the borders of the lattice have neighbours on the other side of the lattice. Hence all actors have the same number of neighbours. It is assumed that all actors update their desires and beliefs at the same time.

Running a large number of simulations like these confirm that this pattern indeed is a typical one in the sense that it contains islands of desires and islands of beliefs that occasionally overlap and then lead to actions. Since there are no intrinsic differences between the actors located in different regions of this social space, this example shows that social-interaction processes, in and of themselves, can explain social differentiation, that is to say, the tendency of different groups to spontaneously organize themselves into social clusters with different beliefs, desires, and/or actions.

Previously I discussed the logic of two intra-individual mechanisms, wishful thinking and the sour-grape syndrome. Wishful thinking, as noted above, denotes a causal connection from an actor's desires to his/her beliefs that makes the actor believe what (s)he desires to be the case. In terms of DBA triplets, this means that a *<1,0,0>* triplet will be transformed into a *<1,1,0>* triplet. But since actors act when they believe that an action will bring about a desired outcome, this pattern will be further transformed into a *<1,1,1>* triplet. The sour-grapes syndrome, as noted above, is a causal connection from an actor's beliefs to his/her desires, which makes the actor desire only what (s)he believes (s)he can get. In terms of DBA triplets, this means that a *<1,0,0>* pattern will always be transformed into a *<0,0,0>* pattern. Figure 8 shows how these two mechanisms influence social outcomes like these.

The computational models that generated these results were the same as before, with the exception that 20 percent of the actors were exposed to wishful thinking and 20 percent to the sour-grapes syndrome. While the previous analyses revealed that approximately 5 percent of the actors acted when they had interacted and influenced one another for forty rounds, the introduction of some «wishful thinkers» and some «sour grapes» into the population, changes these proportions to 10 and 4 percent respectively.

As compared to the effect of the wishful-thinking mechanism, the simulation results reported in Figure 8 reveal that the sour-grapes mechanism has a rather marginal impact on the extent to which the actors act. At first sight it may seem surprising that the sour-grapes mechanism has any effects whatsoever on action, since the change from *<1,0,0>* to *<0,0,0>* does not represent any change in action —the third entry of the triplet is equal to zero in both cases. The change in the actor's desires brought about by the sour-grapes mechanism may, however, influence the desires of those with whom the actor interacts and thereby alter *their* actions. In this respect there is a subtle but important difference between the wishful-thinking and the sour-grapes mechanisms. While the former mechanism has a direct effect on the focal actor's actions and an indirect effect on the actions of others, the latter mechanism has only an indirect effect on the actions of others.

The results presented so far thus show that social interaction processes as well as the details of the intra-individual mechanisms can have a profound impact on the social patterns that emerge. What I have tried to illustrate with these analyses is how a combination of mechanism-based theories of action and interaction, and agent-based computational models makes it possible to ana-
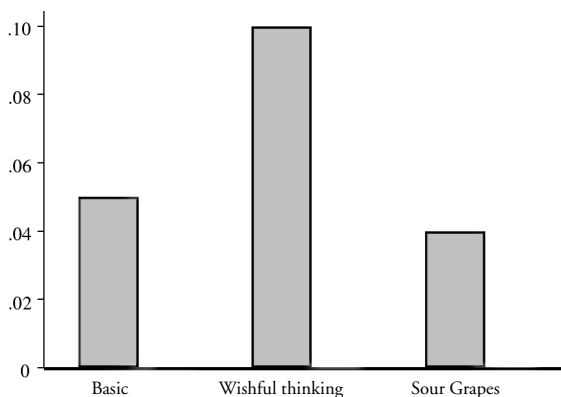
Figure 8. Summary of simulation results. Each bar is an average based on 500 simulations.

lyze, in a theoretically grounded fashion, how individual desires, beliefs, and actions are the product as well as the producer of large-scale social change.

The importance of empirical calibration

Let me then move from the purely theoretical and virtual to the at least somewhat more real. How can we use models like these to explain *specific* empirical events or processes? In order to convince someone that a particular type of mechanism explains an observed social pattern, it is not sufficient to demonstrate that a particular mechanism *can* bring about the pattern. Many different mechanisms can generate the same outcome so the fact that we observe a social outcome compatible with a specific mechanism is not particularly strong evidence in its support. We also need to show or make probable that it indeed was this particular mechanism that was at work.

Furthermore, societies are not closed systems. We must always allow for the possibility that various events, external and unrelated to the processes focused upon in the analysis, influence the outcomes we seek to explain. Observed outcomes typically are the result of many different processes operating in parallel, and it certainly would be useful to know whether the processes we focus upon also operate as we assume them to do when we control for the importance of other types of processes and events. In order to accomplish these tasks we need to reduce the gap between the theoretical model and the empirical reality, and we can do this by empirically calibrating the model.

To clarify what I have in mind, I will use data on youth unemployment in the Stockholm metropolitan area to calibrate these models. As shown in Åberg, Hedström and Kolm (2003), social-interaction processes are highly important for explaining unemployment durations. The outcome focused upon here

—exit from unemployment— may not be the ideal outcome for calibrating these types of models because an individual's possibilities of leaving unemployment also are the result of other individuals' actions, not the least the actions of employers. For the theoretical and methodological points I wish to make here, however, this type of data is perfectly fine.

The data set contains information on all 20- to 24-year-olds who ever lived in the Stockholm metropolitan area during the period from January 1993 to December 1999. Since the process being analyzed here unfolds over time, and since the outcome event of interest —leaving unemployment— is a discrete event, the statistical model being used is a so-called discrete-time event history model (see Allison 1982). This type of model essentially is a regular logistic regression model where the unit of analysis has been changed from persons to 'person-weeks'. All in all 87,924 individuals experienced some unemployment during this time period, and they contributed a total of 2,463,079 person-weeks. These person-weeks are our units of analysis. In addition to the lagged unemployment level among neighbourhood peers, the model includes a number of covariates, such as the number of job vacancies in the region, gender, age, education, and ethnicity[7]. The results are described in Table 3.

The variable of interest here is the unemployment level among neighbourhood peers, and the extent to which this variable is related to the focal individual's propensity of leaving unemployment. For an average individual these results imply the social-interaction effects described in Figure 9. The graphs show that the probability of an unemployed individual leaving unemployment is considerably influenced by the unemployment level among neighbourhood peers, also when all the covariates of Table 3 are taken into account.

We can use these results to empirically calibrate the type of agent-based model used above. This allows us to examine the social outcomes likely to be brought about when we base the analysis on a more plausible model of the extent to which the agents influence one another. If we assume, as before, that the actors' action opportunities are such that they can be in only one of two mutually exclusive states, the first regression model of Table 3 says that the probability that an actor will change state/action at a specific point in time is given by the expression:

$$p_{jt} = \frac{1}{1 + e^{2.085 + 4.086 \times U_{jt\text{-}1}}} \qquad (1)$$

where $U_{jt\text{-}1}$ equals the proportion of the neighbourhood peers who were in the same state or acted in the same way as the focal actor at the previous point in

---

7.   The Stockholm metropolitan area is divided into 699 so-called SAMS areas and these serve as my definition of a neighbourhood. The SAMS areas have been defined so as to contain socially homogeneous residential areas. See Hedström (2005) for a detailed description of the data.

Table 3. Logistic event-history regression model of the propensity of leaving unemployment: Regression coefficients, with z statistics in parentheses.

	Model 1	**Model 2**
Unemployment level among neighbourhood peers (at the end of the preceding week)	-4.086 (82.99)	-2.087 (-33.59)
Woman		0.132 (25.98)
Age		-0.023 (-9.74)
Vocational training		0.027 (3.83)
High school education		0.137 (20.16)
College education		0.206 (21.71)
Immigrant from eastern Europe or former Soviet Union		-0.138 (-8.40)
Immigrant from Middle East or Africa		-0.192 (-18.84)
Immigrant from the rest of the world		-0.014 (-1.53)
Less than 3 years in Sweden		-0.455 (-27.64)
3 – 5 years in Sweden		-0.044 (-3.28)
Married		-0.034 (-2.76)
No. of children		-0.055 (-6.16)
Previous unemployment experiences (no. of weeks/10)		-0.019 (-19.12)
Number of unemployed per vacant job (at the beginning of the month)/100		-0.034 (-0.24)
Length of current unemployment spell (no. of weeks/10)		0.319 (65.40)
Square of the length of current unemployment spell		-0.045 (-51.26)
Constant	-2.085 (363.35)	-2.145 (33.69)
Annual and monthly dummy variables included	No	Yes
Log likelihood	-644312.97	-627468.57
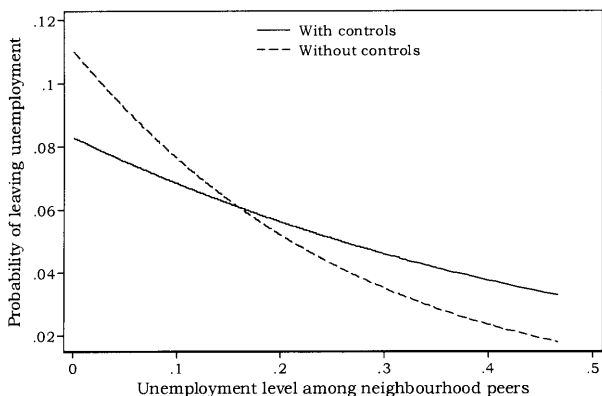
Figure 9. Estimated strength of the social interaction effect for an average unemployed individual.

time[8]. The equation says that, the larger the proportion of the neighbourhood peers that acted in the same way as the focal actor, the less likely it was that the actor would change action.

In order to examine the social patterns that emerge when agents' actions are decided on the basis of this rule, we proceed in the same manner as before. We assume that 2,500 actors are placed on a lattice (torus) with fifty rows and fifty columns. We start with an entirely random action pattern and then we examine the social patterns that emerge when the agents interact and influence one another. One important difference between these analyses and those presented above is that we now focus on the actions as such and not on the underlying beliefs and desires of the actors. It would have been desirable to include beliefs and desires in the analysis but we do not have any empirical information about them.

A typical initial action pattern looks like the upper-left graph of Figure 10. Black areas identify actors who acted in one way (call their action a B-action), and white areas identify those who acted in the other way (call their action a W-action). In the simulation reported in Figure 10, 40 percent of the actors performed a B-action and 60 percent performed a W-action at the outset of the analysis.

Once again, we take social interactions into account by assuming that the actors are directly influenced by their four immediate neighbours (see Figure 6). These neighbours influence the focal actor's probability of changing his/her

8. The results of these analyses can be interpreted as either referring to the states in which the actors are or in terms of their actions. To simplify the presentation, hereinafter the results are presented in action terms.
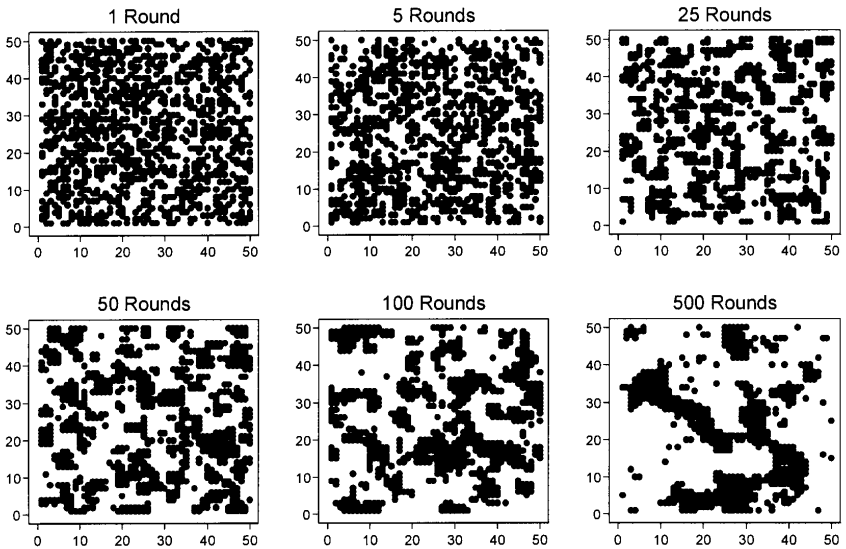
Figure 10. Typical action patterns in a population of 2,500 actors who socially interact with four neighbours on the basis of an empirically calibrated action rule.

action in the manner described by Equation 1. The social patterns that emerge under these conditions typically look like those in Figure 10. Although we start with an entirely unstructured social pattern, a highly segregated pattern emerges rather quickly. Already when the actors have interacted and influenced one another over five rounds, segregated patterns start to emerge. As the interaction process proceeds, the extent of clustering and segregation increases.

Obviously, some of the social-interaction effects assumed in these analyses are not proper interaction effects but are due to individual heterogeneity. If we did not control for such differences, we would overestimate the extent to which individuals are influenced by others. This was the reason that I included a set of control variables in the second regression model in Table 3.Using the social-interaction effect estimate from the second regression model to calibrate the agent-based model and holding all the other variables constant at their mean values, brings about the type of social outcomes described in Figure 11.

Also in this case there are clear segregation and clustering tendencies, but they are not of the same magnitude as in the previous set of analyses, and the social patterns that emerge look rather different. These analyses hence show that social-interaction processes can bring about highly segregated social patterns also when the agents act on the basis of plausible assumptions about the
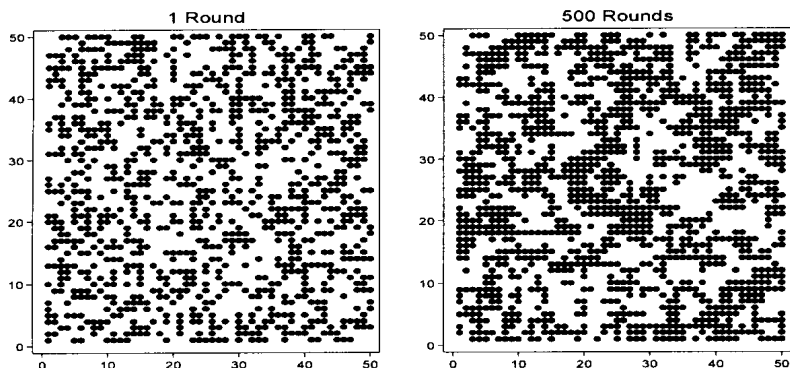
Figure 11. Typical action patterns in a population of 2,500 actors who socially interact with four neighbours on the basis of an empirically calibrated action rule with controls for confounders.

strength of social interaction[9]. In addition they show how sensitive the social-level predictions are to the way in which social-interaction effects are estimated. Taking other events and processes into account considerably influences the social outcomes predicted by the model.

Performing analyses like these allow us to not only state that a social-interaction process *might* have been of importance, but they allow us to state with some confidence that such processes actually were at work and they give some indication of how important they were for the social outcome we seek to explain. Being able to make such claims, I believe, is essential for the future of theoretical micro-macro modelling in an empirically oriented discipline like sociology.

Concluding remarks

In this paper I have tried to illustrate the various components involved in an explanation of social change. The analytical approach adopted here seeks to closely integrate mechanism-based theories of action and interaction, and agent-based computational modelling. The core of the approach can be summarized in the following way:

9.   Bruch and Mare (2004) found that the use of plausible probabilistic decision rules in a traditional Schelling (1971) model of residential segregation did not generate the highly segregated patterns that Schelling models normally generate. Bruch and Mare discussed whether their finding could be generalized to social-influence processes more generally. These results suggest that they cannot.

— Start with explicating the logic of the action and interaction mechanisms assumed to be operative. Develop an agent-based model that emulates the operation of these mechanisms. Simulate the model in order to examine *generative sufficiency* (Epstein and Axtell 1996), that is, make sure that the model *can* generate the type of social outcome to be explained. If the model exhibits generative sufficiency, we have a mechanism-based explanation of the outcome, but the explanation has not yet been empirically verified.
— For empirical verification, use relevant data to examine the most important bits and pieces of the causal machinery in order to verify that the mechanisms actually work as postulated.
— Examine generative sufficiency when the agent-based model has been modified in the light of (2) and after controls for likely confounders have been introduced.

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

In this article I used five basic theoretical concepts —desires, beliefs, opportunities, actions, and relations— and I sought to express more complex molecular mechanism in terms of combinations of these basic entities and activities. Obviously the basic vocabulary may need to be extended, but it is important to try to avoid the current embarrassment of the riches as far as sociological concepts are concerned. We must try to find a basic theoretical vocabulary, consisting of a small number of fundamental concepts in which complex processes can be expressed and analyzed, and this vocabulary must be firmly anchored in our theories of action and interaction.

References

ÅBERG, Y.; HEDSTRÖM, P.; KOLM, A.-S. (2003). «Social Interactions and Unemployment». Working Paper 18. Uppsada University: Department of Economics.

ALLISON, P.D. (1982). «Discrete-Time Methods for the Analysis of Event Histories». *Sociological Methodology 1982*: 61-98.

BOUDON, R. (1979). «Generating Models as a Research Strategy». In: ROSSI, P.H. (ed.). *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld.* New York: The Free Press, p. 51-64.

— (1998). «Social Mechanisms without Black Boxes». In: SWEDBERG, R. (ed.). *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press, p. 172-203.

BRUCH, E.E.; MARE, R.D. (2004). «Neighborhood Choice and Neighborhood Change». *California Center for Population Research. On-Line Working Paper Series* n. 007-04.

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.

DAVIDSON, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.

EDLING, C. (2002). «Mathematics in Sociology». *Annual Review of Sociology* 28: 197-220.

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

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

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

EPSTEIN, J.M.; Axtell, R. (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.

GOLDTHORPE, J.H. (2000). *On Sociology: Numbers, Narratives, and the Integration of Research and Theory*. Oxford: Oxford University Press.

HAHN, R.A. (1973). «Understanding Beliefs: An Essay on The Methodology of the Statement and Analysis of Belief Systems». *Current Anthropology* 14: 207-229.

HEDSTRÖM, P. (1998). «Rational Imitation». In: SWEDBERG, R. (ed.). *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press, p. 306-27.

— (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge: Cambridge University Press.

HEDSTRÖM, P.; SWEDBERG, R. (1998). «Social Mechanisms: An introductory essay». In: SWEDBERG, R. (ed.). *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press, p. 1-31.

HEMPEL, C.G. (1965). *Aspects of Scientific Explanation*. New York: The Free Press.

HOLLAND, J.H. (1998). *Emergence: From Chaos to Order*. Cambridge, Mass.: Perseus Books.

KING, G.; KEOHANE, R.; VERBA, S. (1994). *Designing Social Inquiry: Scientific Inference in Qualitative Reseach*. Princeton: Princeton University Press.

LAZARSFELD, P. (1955). «Interpretation of Statistical Relations as a Research Operation». In: LAZARSFELD, P.; ROSENBERG, M. (eds.). *The Language of Social Research*. New York: Free Press, p. 115-25.

LEWIS, D.K. (1994). «Lewis, David: Reduction of Mind». In: GUTTENPLAN, Samuel (ed.). *A Companion to the Philosophy of Mind*. Oxford: Blackwell Publishers, p. 412-431.

MACHAMER, P.; DARDEN, L.; CRAVER, C.F. (2000). «Thinking About Mechanisms». *Philosophy of Science* 67: 1-25.

MACY, M.W.; WILLER, R. (2002). «From Factors to Actors: Computational Sociology and Agent-Based Modeling». *Annual Review of Sociology* 28: 143-166.

MERTON, R.K. (1968). «The Self-Fulfilling Prophecy». In: MERTON, Robert K. (ed.). *Social Theory and Social Structure*. New York: The Free Press, p. 475-490.

SALMON, W.C. (1971). *Statistical Explanation and Statistical Relevance*. Pittsburgh: University of Pittsburgh Press.

SCHELLING, T.C. (1971). «Dynamic Models of Segregation». *Journal of Mathematical Sociology* 1: 143-186.

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: HAHN, L.E. (ed.). *The Philosophy of George Henrik von Wright.* La Salle: Open Court.

WATTS, D.J.; STROGATZ, S.H. (1998). «Collective Dynamics of "Small World" Networks». *Nature* 393: 440-442.

WHITE, H.C. (1970). *Chains of Oppurtunity: System Models of Mobility in Organizations.* Cambridge: Harvard University Press.