Coda—The past and future of analytical sociology[1]

Peter Hedström

1. INTRODUCTION

Twenty-five years now have passed since the Royal Swedish Academy of Sciences conference took place in Stockholm, leading to *Social Mechanisms: An Analytical Approach to Social Theory* (Hedström and Swedberg 1998), the book that, according to many, marked the beginning of analytical sociology as we perceive it today. This calls for some reflections on how the field has developed. Since the field has grown considerably during these years, my reflections will only cover a small subset of the issues that analytical sociologists are concerned with, and my set of priorities for the future may not necessarily be shared by everyone.

The chapter is organized as follows: First, I briefly discuss the birth of analytical sociology, what motivated its emergence, and what its most essential characteristics are. Second, and in more detail, I consider a set of topics that I believe to be of crucial importance for the future of analytical sociology. I pay particular attention to the role of action-based explanations of macro-level change. While the link between micro behaviors and macro dynamics is at the very core of analytical sociology, the importance many analytical sociologists attach to intentional explanations of behavior, combined with the lack of information on the individuals' intentions, has led to a proliferation of just-so stories. This sort of storytelling, based on the analyst's own preferred action theory, contradicts one of the main tenets of analytical sociology—that explanations should refer to the causal processes known to be responsible for the outcomes to be explained—and I suggest that we should avoid intentional explanations in all but the rare cases where we have access to reliable information on the mental states of the acting individuals. Instead we should concentrate our explanatory efforts on interaction structures and the macro patterns they give rise to.

2. THE CORE OF ANALYTICAL SOCIOLOGY

Since I already have written extensively on the core characteristics of analytical sociology (e.g. Hedström and Bearman 2009), I will keep this section brief. When writing *Dissecting the Social* (Hedström 2005), my ambition was to start from first principles: to identify what characterizes an ideal explanation, and then, from this ideal, derive how one ought to go about explaining social phenomena and how one ought to design empirical research in order for it to be as relevant as possible for the explanatory endeavor.

Sociological theory and research of the day had an ambivalent and poorly articulated relationship to explanations, particularly to the kind of mechanism- and process-oriented explanations that came to be adopted by analytical sociologists. Social theory was more oriented toward

[1] I would like to thank Marc Keuschnigg and Gianluca Manzo for their excellent comments on an earlier draft. The research has been supported by a grant from the Swedish Research Council (DNR 445-2013-7681).

normative critique and conceptual elaborations than toward the development of explanatory theories, and empirical research was more concerned with describing social developments than to test and empirically identify the cogs and wheels of important social processes. Quantitative empirical research, by and large, was dominated by a rather superficial variable-centered form of analysis where, to use Abbott's (1998) apt expression, variables and not actors did the acting (see also Goldthorpe 2021 for an interesting discussion of this state of affairs).

The explanatory approach advocated by many analytical sociologists was a form of a mechanism approach with the following core characteristics (see also Manzo 2014):

1. The *typical explanandum* is some sort of macro-outcome or pattern. Macro outcomes can be observed at different scales, ranging from the small group to society at large, and outcomes of interest include traditional sociological concerns such as inequalities, segregation, and diffusion patterns, as well as, say, the growth dynamics of organizational populations or the structural properties of social networks.
2. An acceptable explanation of such outcomes details the *social processes* through which the outcome was brought about. Simply referring to a particular event as the cause of the outcome is not sufficient; the process through which the initial event brings about the outcome must also be specified.[2]
3. Individuals are the *basic entities* of these kinds of processes because, although variously constrained, it is their behavior that drives the processes forward.
4. Existing *relationships* between the individuals also are important because they often influence the individuals' behavior and thereby also the ways in which the social process unfolds.
5. The kind of explanation we seek takes the form of a clear, precise, and empirically supported *model* of a social process, which is such that it steers the group of individuals into eventually bringing about the type of outcome to be explained.

The importance of using model-based explanations derives from the fact that micro-macro dynamics tend to be highly complex. Particularly when many individuals are involved in them, it often is impossible to know whether a proposed explanation is at all relevant in the sense of being able to bring about the type of macro-outcome to be explained.

Because of its flexibility and its ability to also represent complex social processes, agent-based simulation modelling emerged as a central tool for analytical sociologists. It is a simple-to-use formalism that allows researchers to simulate various processes and thereby examine whether a proposed mechanism or process can generate the type of outcome to be explained (see Flache and de Matos Fernandes, this volume). Although many applications of agent-based modelling leave a lot to be desired from an explanatory point of view, the methodology as such is of crucial importance for analytical sociology (Hedström and Manzo 2015).

---

[2] In some of my previous publications I have put more emphasis on social *mechanisms* than on social *processes*. This change in emphasis, in part, reflects changed priorities. An explanation that refers to a mechanism without spelling out its implied dynamics is not of much use for explaining a macro outcome, while information on a social process, even when poorly articulated and understood, can provide important insights useful for our explanatory efforts. More generally, though, whether we emphasize one rather than another is rather inconsequential because, as Manzo (2007: 5–6) expressed it, a "process is nothing more than the dynamic aspect of one (or several) mechanism(s): it is what the mechanism can trigger".

The type of model I have in mind here when I refer to model-based explanations also is some sort of agent-based simulation model, but clearly other types of models could be envisioned such as purely mathematical models or even exclusively verbal models. But for a verbal account to count as a model, in this respect, it must be so clearly and precisely stated that it directly and without any ambiguities can be translated into an agent-based simulation model. However, a serious drawback with purely verbal models is that they are not well-equipped for getting a handle on complex processes involving large numbers of interacting individuals.

In order to make these abstract explanatory principles more concrete, let us use as an example that currently is of utmost concern throughout the world, the COVID-19 pandemic. If we were to explain why the pandemic developed as it did in a certain area, a simplistic causal-effect approach relating some attributes of the area in question to the state of the pandemic would not suffice. Instead, we would have to follow the scheme outlined and clearly specify the core ingredients of the dynamic *process* through which the virus spreads in the population. Although the virus as such is the key to the epidemic, how the epidemic develops over time to a large extent is the result of a social process. The virus would have died in its infancy had it not been for individuals interacting with one another, and the size of the epidemic and its development over time is, to a large extent, driven by the interaction patterns between the individuals.

In order to explain why the epidemic developed as it did, we need to specify an empirically supported model of the process believed to have been at work. This model must represent the relevant group of individuals, the patterns of interaction between them, the risk that an infected individual transmits the virus to a susceptible individual, recovery rates, and so on. Once a model is at hand, we can examine its explanatory relevance by comparing the macro outcomes it generates with those observed in reality. If it passes this relevance test, we can use the model to address more specific queries about what is likely to have happened to the epidemic had the entities, relations, or activities that govern the process been subject to some sort of surgical intervention that did not affect any other part of the causal structure, e.g. interventions affecting the interaction patterns between the individuals such as travel restrictions or isolating or vaccinating high-contact individuals (for an interesting analysis of this latter point, see Manzo and van de Rijt 2020).

The explanatory strategy would have been exactly the same had we instead sought to explain, say, the ethnic segregation of a city, the emergence of status hierarchies, or downloads at Spotify. That is, we must specify an empirically supported model of the process through which a group of interacting individuals gradually bring about the outcome to be explained. However, as Centola has called attention to with his distinction between simple and complex contagion (see Centola, this volume), the process through which a behavior or an idea spreads from one individual to another often is more complex than the processes through which a virus spreads, but the explanatory principles are the same (see also Goldberg and Stein 2018).

What this explanatory scheme calls attention to is the importance of basing explanations of macro outcomes on empirically plausible and process-oriented micro foundations. Without such foundations we are likely to misconceive the complex causal structure of the social world, and we are likely to offer incorrect or at least vague and shallow explanations for why we observe what we observe. My sense is that the importance of using this approach has become gradually more recognized within the discipline (e.g. Bearman et al. 2004; Bruch and Mare 2006; Goldberg and Stein 2018). This positive development is not only due to the examples set

by analytical sociologists, but also reflects the increased availability of data on large populations of interacting individuals and new methods suitable for analyzing such data (see Spaiser, this volume, Jarvis et al. (forthcoming), Keuschnigg et al. 2018).

3. THEORETICAL DEVELOPMENTS

The social mechanism approach, as originally perceived, not only called for empirically plausible and process-oriented models; it was also a call for a different kind of sociological theory. It insisted that we should abandon the idea of general theory, be it of the Parsonian kind of the past or the rational choice kind of the more recent past, and instead adopt a view of sociological theory as a toolbox of social mechanisms.

This vision of theory does not require that the theories in the toolbox eventually be organized into a grand unified theory. Nor does it require that the theories in the toolbox are true or false in any absolute sense of the word (Hedström and Ylikoski 2010). Instead, these theories have the character of what Coleman (1964) referred to as "sometimes-true theories". A theory of this kind, which details the logic of a certain type of social process, should not be evaluated in terms of whether it is true or false in general, but whether it applies to the specific situation at hand. In many instances several different kinds of processes could have brought about the outcome to be explained, and "the problem in any given application is to know *which* of the many processes is followed in this case" (Coleman 1964: 516–7) because the explanation must be done with reference to the type of process that indeed was at work in the empirical case to be explained.

This type of theory has much in common with Merton's (1967) idea of sociological theories of the middle range, something that Merton himself also strongly emphasized at the 1996 mechanism conference in Stockholm. Merton originally introduced the idea of middle-range theory in a comment on a paper by Parsons. In this comment he objected to Parsons's belief in the possibility of creating a sociological *theory*, in the singular, and the corollary belief that it would be possible to develop a theory that would be general enough to treat *all phenomena* of interest to sociology. "Sociology will advance in the degree that the major concern is with developing theories adequate to limited ranges of phenomena and it will be hampered if attention is centered on theory in the large" (Merton 1948: 165–6).

As discussed in Hedström and Udehn (2009), a theory of the middle range is a type of theory that clearly and precisely specifies a type of process that can bring about a particular type of outcome. The prototypical example is Merton's (1948) theory of the self-fulfilling prophecy, which details a process through which individuals, in interaction with one another, gradually make an initially false belief about the world come true. The theory is general in the sense that it applies to many different types of beliefs and behaviors, but the explanation is partial in the sense that it does not seek to explain token outcomes in all their minute details. This type of theory instead seeks to bring out the heart of the matter by detailing a *type* of process that can explain the *type* of outcome that the token outcome is an example of.

While there has been considerable methodological progress in how sociologists go about analyzing and explaining macro-outcomes, the theoretical toolbox has not progressed in the same manner. At the close of the 1996 Stockholm conference on social mechanisms, Thomas Schelling suggested that the next important step for us should be to systematize sociological theory by creating better order in the theoretical toolbox. To do this, he suggested, we should

initiate a book project that arranged and categorized the already known social mechanisms and identified those areas of importance where the existing toolbox was shallow or empty. Unfortunately, that never happened, and the lack of progress in this area is perhaps most clearly seen in that we still use Merton's old theories to illustrate what kind of mechanism theories we have in mind. With a risk of sounding unappreciative, which I certainly am not, although the very best examples of analytical sociology, such as the winners of the Robert K. Merton best article awards, clearly articulate the type of processes that have brought about the outcomes they seek to explain (see Manzo, this volume), they have not produced *theories* of the same compelling simplicity and elegance as Merton's.

It is perhaps to demand too much of a field that it should produce theories on par with those of Merton. After all, the entire history of sociology has not produced all that many of them. I do believe, however, that a more rapid theoretical progress would have been possible, but that it has been held back by the disciplinary culture in which we are embedded. In this culture, as emphasized by Kroneberg and Tutic (this volume), theorizing tends to be equated with proposing concepts, heuristics, and frameworks, and theoretical progress with the development of increasingly differentiated and complex conceptualizations of actions and practices. This idea of what theory and theorizing is all about stands in sharp contrast to the analytical-sociology view of *theory as explanation*, and to the importance it attaches to clear and precise specifications of different types of social processes.

Being embedded in a somewhat foreign theoretical culture has many unfortunate consequences. One is that those who have the analytical ability to fully understand the complexities involved in these types of social processes typically do not view themselves as theorists and do not see it as their role to contribute to the theoretical toolbox. This is highly unfortunate because if we leave theory development to those who currently define themselves as theorists we will see little progress in the type of theory we are aiming for. In my vision of the future, analytical sociologists also routinely publish in journals like *Sociological Theory* and there they bring to the fore the general theoretical core of the matters they study in the form of theoretical models of social processes that bring about different types of macro outcomes.

Another consequence of being embedded in a somewhat foreign theoretical culture is that it forces us to engage in debates that often are entirely unproductive from our point of view. Two influential social theory traditions, that also are featured in this volume, critical realism and pragmatist theory, can be used to illustrate my point. Although there are some surface similarities between analytical sociology and these traditions, such as the concern with mechanisms, the analytical sociology agenda in other respects is so far removed from their agendas that the prospect for a productive dialogue is minimal.

In the case of critical realism, as emphasized by Di Iorio and León-Medina (this volume), most of these difficulties can be traced to their rather obscure ontological commitments. One aspect of this is their firm commitment to Bhaskar's (1978) view of reality as composed of three separate domains, the empirical, the actual, and the real. Since they assume that mechanisms operate at the empirically non-observable domain of the real, they exclude by definition the possibility of constructing the kind of empirically fine-tuned mechanism models that analytical sociology aims for. Another ontological assumption that hampers the prospect for a fruitful dialogue is their assumption that social structures have causal powers that are autonomous of the individuals who are embedded in them. As emphasized by Di Iorio and León-Medina, this assumption introduces a "mystery gap" between micro behaviors and macro outcomes that closes the door to any serious dialogue with analytical sociology about

generative social processes. Macro outcomes often are more difficult to explain than isolated micro events, particularly when numerous interacting individuals are part of the process, but the fact that something is complex and difficult to explain does not mean that it is inexplicable or that emergent social structures are autonomous from their micro bases.

Pragmatist theory is not at all as marred by obscure ontological commitments as critical realism is, and the work of pragmatists such as Gross (2009) and Norton (this volume) who discuss the relationship between pragmatist theory and analytical sociology, is clear and eminently understandable. In his article on a pragmatist theory of social mechanism, Gross (2009: 375) argues as follows:

> Sociology should aim to identify the main social mechanisms by which cause and effect relationships in the social world that are of moral, political, or intellectual importance come about. This entails breaking complex social phenomena into their component parts to see how aggregations or chains of actors employing habits to resolve problem situations bring about systematic effects.

Although most analytical sociologists may not feature "habits" and "problem situations" as centrally as Gross does in a general programmatic statement, I assume that most would find his statement perfectly agreeable. That habits of the mind as well as habitual behavior can be important for explaining macro outcomes seems entirely uncontroversial, particularly when allowing for moments of "creativity" at which such habitual patterns are altered. It is more unclear what explanatory purchase there is in framing the analysis in this pragmatist cloak. Simply introducing the concepts and the perspective will not expand our theoretical toolbox or offer any new insights into complex social processes.

I am open to the possibility that pragmatist ideas may produce important additions to our theoretical toolbox, but neither Gross nor Norton make any serious move in that direction. Their apparent lack of interest in clearly and precisely detailing how different types of macro outcomes are brought about is rather typical for mainstream social theorists. As suggested by Kroneberg and Tutic (this volume), most social theorists are exclusively concerned with developing *conceptualizations* of actions, practices, etc., and in that respect they differ considerably from analytical sociology. The view of sociological theory as a toolbox of sometimes-true theories implies a much more pragmatic and flexible attitude toward the conceptualization of actions and practices. There is no general conceptual scheme that all analyses must make use of. Instead, a proposed mechanism is judged exclusively in terms of whether it is clearly and precisely articulated, whether it is empirically plausible, and whether it leads to new or more precise models of important social processes.

To what extent has analytical sociology benefited from dialogues with these traditions? Not much, I'm afraid. For a field to have a fertile impact on the analytical agenda, it must at least share our commitments to clarity and precision, to empirical realism and to generative explanations of social change. If none of these requirements are met, dialogues are likely to be distractive rather than constructive. For these reasons I am highly skeptical of the idea of developing an "analytical sociology in the pragmatist mode" as suggested by Norton (this volume).

What is needed in order to show that a pragmatist move has something to offer is not further conceptual elaborations but tangible evidence, preferably in the form of clearly articulated models of different types of social processes. In order for pragmatist statements to be taken seriously they must do something similar to what Goldberg and Stein (2018) did for our understanding of the diffusion of culture. That is, to include some of their core ideas in a formalized model of a social process, and tangibly show that this will enrich our

understanding of the ways in which macro outcomes are brought about (see also Manzo 2021 this volume).

As I see it, analytical sociology gains much of its strength from being open in conceptual matters, from not being tied down by any general social theory, and for being wholeheartedly focused on explaining important macro outcomes in the sense of showing how the outcomes were brought about. I fear that a closer dialogue with traditional social theorists will simply lead to endless discussions about purely conceptual matters. The philosophy of science community is likely to offer much more productive dialogues as far as meta-theoretical matters are concerned, since they tend to share our commitment to clarity and precision, and interactions with empirically oriented sociologists—experimental, quantitative, and qualitative—are likely to be much more conducive for the development of new middle-range theories than dialogues with social theorists.

4. DO WE NEED INTENTIONAL EXPLANATIONS?

Individuals' actions are of central importance for analytical sociology since the social processes we are particularly concerned with are those that steer a group of individuals into acting in such a way that they collectively bring about the type of outcome to be explained.[3] The fact that individuals' actions are part of the process often has been understood as necessitating two different types of explanations (e.g. Elster 1983): First, we must intentionally explain why individuals do what they do and, second, we must causally explain why they bring about the macro outcomes they do when they act as they do.

In *Dissecting the Social* I also emphasized the importance of firmly anchoring all explanations of social outcomes in intentionally explained behavior. I argued that "intentional explanations are important for sociological theory because… they make the act 'understandable' in the Weberian sense of the term.… [I]n order for a theory [of the social] to be explanatory it must consider the reasons why individuals act as they do."

I am now considerably more skeptical to the feasibility and desirability of such an approach, and this is for three interrelated reasons:

1. In order to assess the correctness of a proposed intentional explanation, we must have access to information on the mental states of those whose behavior we wish to explain, but such information is scarcely ever available.
2. Our theories of action are not nearly powerful enough to serve as a substitute for the missing empirical information on the mental states of the individuals.
3. If we nevertheless were to use Elster's double-explanatory approach, we would commit what Elster (2009) himself has referred to as the fallacy of "excessive ambitions", and this is something we should avoid at all costs.

Let me expand a bit on each of these three points.

Following Davidson (1980), mental states such as beliefs and desires can be said to cause our actions in the sense of providing reasons for them, and knowing an individual's mental

[3] Please note that I use the terms *action* and *behavior* as synonyms, and as referring to the behavior of an individual, whether or not that behavior is guided by an intention.

states allows us to understand why the individual did what they did. An intentional explanation of an individual's behavior refers to the specific constellation of mental states that motivated the behavior in question.

Intentional explanations are appealing because they provide subjective understanding. They are difficult to use appropriately, however, since we scarcely ever know the reason that motivated the behavior of another individual. I think it is fair to say that sociologists have not fully appreciated this problematic aspect of intentional explanations. Merton's (1936: 897) discussion of how to impute reasons to individuals is a case in point:

> Ultimately, the final test [of the imputation] is this: does the juxtaposition of the overt action, our general knowledge of the actor(s) and the specific situation and the inferred or avowed purpose 'make sense,' is there between these, as Weber puts it, a 'verständlicher Sinnzusammenhang?' If the analyst self-consciously subjects these elements to such probing, there is substantial probability that his conclusion in respect to purpose is not too far afield in the majority of instances. (p. 897)

This way of seeking to understand other individuals' behavior by placing us in their shoes is common practice in everyday life. The procedure also may be sufficiently reliable for some social-scientific work, particularly if we only need to impute reasons to a handful of persons whose life histories and outlooks on the world are very similar to our own. What Merton and likeminded analysts have failed to appreciate, however, is the considerable heterogeneity that exists in individuals' motivations. Some may do X because of belief B, while others may do X because of belief C or desire D, and the same individual may do X for entirely different reasons at different points of time.

Particularly in our kind of research where the focus is on the macro outcomes that *numerous* interacting individuals gradually bring about, this verstehen-based procedure is likely to be highly error prone. We can know how the individuals act, in the current as well as in the past, we can know the socio-demographic properties of the individuals themselves as well as the social settings in which they are embedded, and we can know their relationships to others and how the others around them acted in the past, but normally we will not have access to what goes on in their minds and how their mental states change as the social process unfolds. For this reason, we cannot know the specific constellation of mental states that made the individuals' act as they did, and this would seem to preclude the use of intentional explanations in our kind of research. Although there exist ways of eliciting individuals' reasons and emotions in lab settings, or via survey instruments, such procedures are not easy to adapt to large-scale observational studies. This is particularly so if the proposed explanation hinges on specific changes in the individuals' mental states during the process being analyzed.[4]

Schematically the situation may be described as follows, where **A** is the action of an individual, **M** is the individual's relevant mental states at the time of acting, and **S** are various observable social characteristics of the individual and the individual's social environment likely to affect the individual's mental states and actions:

$$\mathbf{S} \rightarrow \mathbf{M} \rightarrow \mathbf{A}$$

[4] Since we cannot elicit individuals' mental states on the fly, the only option would seem to be to use time-stamped digital data, when available, as an indicator on the individuals' reasons and emotions at the time when they acted. However, such data rarely is available, and often does not map properly onto our theoretical constructs.

An intentional explanation is focused on the **M→A** part of the scheme. As with any other type of explanation, an intentional explanation can be correct or incorrect. The explanation is correct if it refers to the mental states that actually motivated the individual to behave as she/he did, and it is incorrect if it refers to the wrong set of mental states. Since we do not know **M**, and since many different constellations of **M** can motivate the same **A**, the prospect for developing correct intentional explanations seems meager indeed. The inaccessibility of **M** also casts doubt onto sociological research programs that hinge on details about **M** and the **M→A** link, such as Esser's frame-selection approach (e.g. Esser 2009).

If we had a sufficiently strong theory of action, we could partly circumvent this problem by using this theory to impute the missing **M→A** component. This was the position taken by Popper (1961, 1994) when proposing his form of *situational analysis*. Inspired by economic theorists and their methodology of constructing explanations of economic behavior by imputing a motive to the individuals, and then derive what their optimal choices would be in the circumstances in which they found themselves, Popper argued that a similar approach would also be adequate for most kinds of social explanations (Hoover 2017; Hedström et al. 1998). Popper seems to have believed that this assumption was theoretically innocent and an almost empty principle that he nevertheless found useful because it allowed us to "pack or cram our whole theoretical effort, our whole explanatory theory, into an analysis of the situation" (1961: 359).

Similar considerations are likely to explain why so many social scientists were attracted by rational choice theory. Although we cannot, but in rare circumstances, know what motivates individuals to do what they do, rational choice theory provided a coherent narrative that could be used as a substitute for this missing information. This use of rational-choice theory was perhaps most clearly seen in the work of Gary Becker and similarly minded theorists who sought to develop a new foundation for the social sciences, not by painstaking empirical research into the true motives of individuals' behavior, but by proposing general and model-based narratives that provided rational-choice–based interpretations of various kinds of behavior (e.g. Becker 1976, 1981; Coleman 1990). What Becker and his followers showed is that it is possible to come up with coherent rational narratives that fit even the most puzzling and seemingly irrational kinds of behavior, particularly if the narratives are free to ignore known empirical facts about the behavior of individuals.[5] Inadvertently, Becker et al. thereby also demonstrated that being able to tell a coherent narrative about a possible causal process is far from sufficient for an acceptable explanation to be at hand. In order for a narrative to be explanatory, it has to be firmly anchored in known facts about the acting individuals and their social settings.[6]

---

[5] During my years at the University of Chicago, the famous Becker–Coleman seminar was at its height, and this seminar in many respects was an ideal type of this kind of rational-narrative approach. The typical seminar format was that an invited speaker presented some research and proposed an explanation, not seldom of a non-rational-choice kind. Very quickly the regular seminar participants interrupted the speaker and proposed numerous, often very ingenious alternative rational choice explanations of the findings. The seminars were intellectually extremely stimulating, but obviously they had more in common with purely intellectual games of Herman Hesse's type of glass-bead game than with an empirically grounded social science proper.

[6] Another serious problem with these kinds of post hoc accounts based on little or no empirical data is their lack of predictive power. While, as has been discussed, there is more to a good explanation than being able to predict if a proposed explanation fails miserably in predicting the kind of phenomena it is supposed to explain, it is a sure sign that something is seriously wrong with the proposed explanation (see Watts 2014 for an interesting discussion of this point).

Although these kinds of just-so stories where the theorist makes up the acting individuals's motives may provide a sense of understanding, this understanding is illusionary as long as the story is factually incorrect. If we nevertheless push ahead and act as if the micro foundations were correct, we commit what Elster (2009) has referred to as the fallacy of "excessive ambitions". He introduced this notion in a rather damaging critique of Bates et al.'s (1998) attempt to explain complex historical developments in rational-choice terms (Elster 2000), and he argued correctly that this sort of explanatory hubris not only leads to poor and misleading explanations but that an acceptance of such principles is likely to hinder the development of serious social science research.

Rather than seeking excessively precise fictions, we should aim to develop theories applicable to the real-world settings that we are analyzing. Such a modest and realist strategy characterizes some of the best theoretical work in the discipline. Take Merton (1968) and his theory of the Mathew effect as an example. This theory specifies the details of a cumulative advantage process in which success breeds further success, in academia as well as elsewhere (see also Lynn and Espy, this volume). Merton never specified any precise model of the decision calculus, nor did he propose any elaborate **M**-story. Doing so would be possible, of course, but if the resulting model did not properly describe what transpired in the case at hand, the formalization would render the account non-explanatory. The situation is similar with Schelling's (1971) classic segregation model that has been an important source of inspiration for analytical sociology. What matters for the aggregate segregation patterns that emerge from the process that Schelling analyzes is *how* individuals react to the behavior and properties of their neighbors, not *why* (in the intentional or mentalistic sense) they react as they do. That is, the segregation dynamics will be the same irrespective of the reasons the individuals do what they do. The crucial aspect that gives the process its unique properties is not related to the **M** or the **M→A** component, but is exclusively driven by the interaction patterns as such, for instance, the extent to which the **A** of one individual influences the **A** of another individual.[7] In both of these cases, more details clearly could be specified and might also be desirable, but not excessive precision that simply amounts to precisely stating and assuming to be true what is untrue, and which thereby would deprive the theory of its explanatory power.

What this discussion suggests is that it is important to make a clear distinction between explanations of actions, on the one hand, and intentional explanations (or action theories founded on such explanations), on the other. In my view, the **M→A** information problem is of such significance that we should avoid intentional explanations in all but the utterly rare cases where we have access to dependable information on **M**. Since the essential **M**-information usually is inaccessible to us, we can neither test the assumed details of the **M→A** link nor

[7] One clarification is called for. If the actors in the Schelling model move when 20 per cent of the neighbors are of the other kind rather than 60 per cent, this will affect the segregation patterns that emerge. One interpretation of this behavioral difference could be that the former group has stronger segregationist preferences than the other, but as long as we do not have any independent information on their **M**-states, this is only a guess. This behavioral difference between the groups could, equally well, arise if there were no differences in their segregationist preferences. It could, for example, arise if the 60 per cent group owned its housing and, therefore, was concerned with how the inflow of the other kind of actors may come to affect their property values, while the other group lived in rental properties and, therefore, was entirely unconcerned about property values. All we know for sure is that we have two groups that differ in their thresholds for leaving their neighborhoods, and that this difference in how they react to others affects the aggregate dynamics. We do not know anything about **M** and the **M→A** link, and my point is that we should avoid making our explanations hinge on causal narratives that we do not have firm empirical support for.

fine-tune our theoretical accounts on the basis of detailed empirical analyses of the **M→A** link. As a result, proposals about the **M→A** link will be unreliable and explanations founded upon them are likely to be highly error prone.

As the examples of Merton and Schelling illustrate, the fact that intentional explanations are rarely possible to use is not necessarily of much importance for us. The reason for this is that most of the social processes that we are concerned with are not dependent on motivational details, but are driven by **A**-based interaction patterns as such. Thus, in many situations, what is essential for explaining macro outcomes is not to have a model of the individuals' deliberations, but to have an accurate model of social interactions and their effects.

The lack of information on **M**, combined with the premium attached to intentional explanations within the discipline, has led to a proliferation of just-so stories with little or no empirical support. This practice comes into conflict with what I consider one of the main pillars of analytical sociology; that our explanations should always refer to processes actually responsible for the outcomes that we seek to explain. If we do not have information on **M** it seems obvious to me that we should not pretend to know what kind of **M→A** link is at work, but instead frame the explanation exclusively in **A** and **S** terms. But equally obvious, of course, if we happen to have access to relevant **M** information for the individuals involved in the process, we should make full use of it in our explanation.[8]

5. THE STRUCTURE OF SOCIAL INTERACTION

For all the reasons mentioned, unless we have access to reliable **M** information, we are well advised to focus our theoretical and explanatory efforts on the analysis of the social situation and leave theorizing about individuals' minds to those who have specialized knowledge in such matters; i.e. psychologists and cognitive scientists (see also Hedström and Ylikoski 2014). That is to say, we should cater to our own strengths and focus, theoretically as well as empirically, on how the social situations in which individuals are embedded affect their behavior. Since other individuals are at the core of these social situations, this implies detailed analyses of social interactions, the ways in which they are structured, and how they affect individuals' behavior and the macro outcomes the individuals collectively bring about.

Michael Mäs's chapter on interactions (this volume) calls attention to Lopez-Pintado and Watts' (2008) notion of *influence-response functions*, and he argues that such functions can

[8] The position taken here differs in important ways from that taken in my previous writings on the so-called DBO model (Hedström 2005). In our everyday life we make constant use of the DBO scheme or similar folk-psychological ideas to make sense of the behavior of others. The scheme is flexible and well adapted for such everyday needs, but it does not work as well as a theory of action. This is most clearly seen in its failure to predict the future behavior of individuals. As mentioned, although there is more to a good explanation than predictive success, if a theory of action fails miserably in predicting the behavior it is supposed to explain, it is a sure sign that something is seriously wrong with the theory (Watts 2014). It is this tension between the psychological appeal of intentional emphatic explanations, on the one hand, and the realist dictum that proper explanations must refer to the correct causal structure, on the other, that has led me to the current position. In Hedström (2005) such considerations led me to express great disbelief in rational choice theories, and in Hedström and Ylikoski (2014), on similar grounds, we questioned the very idea of basing explanations on sociological theories of action. In this chapter, I take the final step and argue for going fully naturalistic; i.e. to give up on the privileged role of intentional explanations, and instead adopt an explanatory account of individual behavior that is directly influenced by our empirical research.

serve to coalesce many seemingly disparate ideas related to the structure of social interaction. Influence-response functions may indeed provide a unifying lens for analytical sociology because they refer to social interactions as such, and they show how an individual's behavior is affected by the behavior of others and, thereby indirectly, how the structure of social interaction is likely to affect the macro outcomes the individuals bring about. The core idea behind such functions has been around for a long time, but that does not diminish its importance.[9]

So, what is an influence-response function precisely? Let me explain with reference to a study of the spatial diffusion of Swedish social movements at the turn of the nineteenth century (Hedström 1994). In this study I took my departure in Granovetter's (1978) threshold model and used influence-response ideas to extend it so that it also could consider the networks in which the individuals were embedded. Central to the influence-response approach is that we have a *n* by *n* matrix **S**, which describes the structure of social interaction between *n* individuals. In my spatial diffusion study, the entries of **S** were inverted spatial distances between the actors, but more generally **S** refers to the social network the individuals are embedded in, or to the extent to which individuals pay attention to other individuals' behavior. In addition to this social-interaction matrix, we have a *n* by *1* column vector, **B**, which identifies those who had adopted the behavior in question at the previous point in time; the influence-response function is given by the *n* by *1* column vector **I**:[10]

$$\mathbf{I} = \mathbf{S} \times \mathbf{B}$$

The entries of **I** constitute the exposure part of the influence-response function and by relating this exposure to each individual's behavior at the next point in time we have a function that shows how individuals respond to the influences emanating from others. As emphasized, individual heterogeneities are common and important for the collective outcomes the acting individuals are likely to generate. Some individuals, for example, are more susceptible to social influence than others. Within an influence-response framework, such heterogeneities can be handled in terms of thresholds in Granovetter's (1978) sense of the term and, as shown in Hedström (1994), individually specific thresholds can be empirically estimated with standard statistical techniques.

I used Schelling's segregation model to bring home the point that the macro outcomes that interacting individuals bring about often are exclusively driven by *how* they react to each other,

---

[9] As far as I know, the core idea was introduced into the social sciences by the social physicist Stewart (1941, 1947) in the 1940s (see also Duncan 1957), and similar ideas are central to Latané's (1981) social-impact theory (see also Nowak et al. 1990), Centola's (2018) work of the spread of behavior, Friedkin's work on network-based social influence (e.g. Friedkin and Johnsen 1999), Land et al.'s (1991) analysis of church attendance, and my own work on contagious collectivities (Hedström 1994; Hedström et al. 2000), just to mention a few.

[10] Assume, for example, that we study a minimal social system comprised of only three individuals and that

$$\text{their } \mathbf{S} \text{ matrix is equal to } \begin{bmatrix} 0 & .9 & .8 \\ .1 & 0 & .2 \\ .4 & .4 & 0 \end{bmatrix}$$

These entries suggest that the first individual, in row 1, pays a lot of attention to the other two individuals, that the second individual, in row 2, pays little attention to the others, and that the third individual, in row 3, occupies a position in between the other two. If the first two individuals, but not the third individual, exhibited the behavior in question at the previous time period, their behavior is given by this

$$\mathbf{B} \text{ vector } \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \text{ and the } \mathbf{I} \text{ vector then is equal to } \begin{bmatrix} .9 \\ .1 \\ .8 \end{bmatrix}$$

and not by *why* (in the intentional sense) they react as they do. What Pintado and Watts and Mäs show is that this conclusion can be generalized to wide classes of behavior.

Pintado and Watts focused on three types of influence-response functions: (1) monotonically increasing functions where the probability of an individual adopting a behavior increases monotonically with the number of others who already have adopted it; (2) monotonically decreasing functions where the probability of an individual adopting a behavior decreases monotonically with the number of others who already have adopted it; and (3) non-monotonic functions where the probability of adoption increases (decreases) up until a certain number of others have adopted the behavior after which it decreases (increases). They show that substantively very different kinds of behaviors can have very similar influence-response functions and, to the extent that this is the case, they will give rise to similar macroscopic patterns.

Instead of investing time and effort into developing stories about an unobserved **M→A** component, how it may vary between different individuals, and how it may change during the course of a social process, there are strong reasons to focus on **S** and **A** instead, and adopt and further develop the influence-response approach. This approach steers our focus to those matters on which we can speak clearly, and which clearly matters for the macro outcomes we seek to explain. All key aspects of the influence-response functions can be studied empirically, and the form of the functions can be estimated with the kind of empirical data that contemporary sociologists have access to (see Hedström et al. 2000; Keuschnigg et al. 2018).[11]

6. WHAT'S NEXT? SOME CONCLUDING REMARKS

As a way of conclusion, let me concisely summarize the main message of the chapter in terms of four particularly important priorities for the future:

1. Invest more time and effort into developing the toolbox of middle-range theories.
2. Avoid intentional explanations of behavior, unless reliable data on the individuals' mental states is available.
3. Pay less attention to mainstream social theory and more attention to fields that share our commitments to clarity, empirical realism, and generative explanations of social change.
4. Embrace the computational social science revolution.

As noted, the development of our theoretical toolbox has not kept pace with the development of our applied work. In the spirit of Thomas Schelling's message at the 1996 social mechanism conference in Stockholm, we need to systematize sociological theory by creating better order in our theoretical toolbox, and by identifying important areas where the existing toolbox is shallow or empty. Many parts of the theoretical toolbox would benefit from close attention, but in the light of the discussion it appears particularly important to systematize and deepen our understanding of influence-response functions. Since relevant work on such functions is scattered into many different areas, one important task would be to

[11] But once again, I am not arguing in favor of a behavioristic approach that denies the existence of mental states or their possible causal efficacy. What I am arguing against is just-so storytelling about the **M→A** component. If we have access to **M** information, we should make use of it and analyze the role that different aspects of the **M→A** component played in bringing about the macro outcome we seek to explain.

coalesce this literature, show its common denominators, and identify what we know about the mapping between types of social situations and types of influence-response functions. In addition, in order for influence-response functions to take a central role in analytical sociology, extensive work is needed in order for us to fully understand the types of macro patterns that different types of influence-response functions give rise to. Using the terminology introduced, we need to know how the macro patterns that emerge depend on how the entries of **S** are structured, the initial entries of **B**, and the form of the function relating **I** to **B**. Similarly, we must understand what role individual heterogeneity may play in these processes. Such heterogeneities can arise in the **S** matrix in terms of clustering, but also in the link between **I** and **B**, since some individuals may be become more likely to adopt the behavior in question when many others have done so, while others may shy away from behaviors that have become "too common", and it is unclear how different mixtures of such groups may affect the collective outcomes they jointly bring about.

Let me end by saying a few words about the emerging field of computational social science (CSS) and its importance for analytical sociology. The reason that I have not discussed CSS is not that these developments are of marginal importance, but that we have already discussed their importance elsewhere (see Keuschnigg et al. 2018; Jarvis et al. (forthcoming)). As emphasized there, CSS has the potential to accomplish for sociology what the introduction of econometrics did for economics in the past half century, i.e. to provide relevant data and data-analytical tools needed for rigorously addressing the core questions of the discipline. As pointed out in several of the chapters of this volume, the kind of survey data that has dominated the discipline for several decades, and the statistical tools used for analyzing such data, are poorly equipped to handle analyses of networked social systems and their dynamics. During the last decade, computationally intensive statistical methods appropriate for handling data on large-scale populations of interacting individuals have emerged at a fast rate, as well as new and improved computational methods for analyzing large corpuses of text. These kinds of developments are important since they enable us to reduce the gap between empirical and theoretical work, and thereby allow empirical research to have more bearing on sociological theory than what has been possible in the past.

These developments are important not only for observational but also for experimental studies. Salganik et al.'s (2006) web-based experiments on artificial cultural markets represent a new type of macro-sociological experiments (Hedström 2006). Their experimental design breaks with older experimental traditions by randomizing participants into separate social systems that each serve as a unit of analysis. The focus is on the macro outcomes that the individuals in the different social systems collectively bring about, and for this reason this type of macro experiment is likely to be of considerable importance for the future of analytical sociology.

REFERENCES

Abbott, A. (1998), "The casual devolution", *Sociological Methods and Research*, **27**: 148–81.

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R. Weingast (1998), *Analytical Narratives*, Princeton, NJ: 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.

Becker, G.S. (1976), *The Economic Approach to Human Behavior*, Chicago, IL: University of Chicago Press.

Becker, G.S. (1981), *A Treatise on the Family*, Cambridge, MA: Harvard University Press.
Bhaskar, R. (1978), *A Realist Theory of Science*, Hassocks, Sussex: Harvester Press.
Bruch, E.E. and R.D. Mare (2006), "Neighborhood choice and neighborhood change", *American Journal of Sociology*, **112**: 667–709.
Centola, D. (2018), *How Behavior Spreads*, Princeton, NJ: Princeton University Press.
Coleman, J.S. (1964), *Introduction to Mathematical Sociology*, New York: Free Press of Glencoe.
Coleman, J.S. (1990), *Foundations of Social Theory*, Cambridge, MA: Harvard University Press.
Davidson, D. (1980), *Essays on Actions and Events*, Oxford, UK: Clarendon Press.
Duncan, O.D. (1957), "The measurement of population distribution", *Population Studies*, **11** (1): 27–45.
Elster, J. (1983), *Explaining Technical Change: A Case Study in the Philosophy of Science*, Cambridge, UK: Cambridge University Press.
Elster, J. (2000), "Rational choice history: A case of excessive ambition", *American Political Science Review*, **94** (3): 685–95.
Elster, J. (2009), "Excessive ambitions", *Capitalism and Society*, **4** (2).
Esser, H. (2009), "Rationality and Commitment: The Model of Frame Selection and the Explanation of Normative Action", pp. 207–30, in M. Cherkaoui and P. Hamilton (eds), *Raymond Boudon: A Life in Sociology*, Oxford: The Bardwell Press.
Friedkin, N.E. and E.C. Johnsen (1999), "Social influence networks and opinion change", *Advances in Group Processes*, **16**: 1–29.
Goldberg, A. and S.K. Stein (2018), "Beyond social contagion: Associative diffusion and the emergence of cultural variation", *American Sociological Review*, **83** (5): 897–932.
Goldthorpe, J.H. (2021), *Pioneers of Sociological Science: Statistical Foundations and the Theory of Action*, Cambridge: Cambridge University Press.
Granovetter, M.S. (1978), "Threshold models of collective behavior", *American Journal of Sociology*, **83**: 1420–43.
Gross, N. (2009), "A pragmatist theory of social mechanisms", *American Sociological Review*, **74** (3): 358–79.
Hedström, P. (1994), "Contagious collectivities: On the spatial diffusion of Swedish trade-unions, 1890–1940", *American Journal of Sociology*, **99**: 1157–79.
Hedström, P. (2005), *Dissecting the Social: On the Principles of Analytical Sociology*, Cambridge, UK: Cambridge University Press.
Hedström, P. (2006), "Experimental macro sociology: Predicting the next best seller", *Science*, **311** (5762): 786–7.
Hedström, P. and P. Bearman (eds) (2009), *The Oxford Handbook of Analytical Sociology*, Oxford, UK: Oxford University Press.
Hedström, P. and G. Manzo (2015), "Recent trends in agent-based computational research: A brief introduction", *Sociological Methods and Research,* **44** (2): 179–85.
Hedström, P., R. Sandell and C. Stern (2000), "Mesolevel networks and the diffusion of social movements: The case of the Swedish Social Democratic Party", *American Journal of Sociology*, **106** (1): 145–72.
Hedström, P. and R. Swedberg (eds) (1998), *Social Mechanisms: An Analytical Approach to Social Theory*, Cambridge, UK: Cambridge University Press.
Hedström, P., R. Swedberg and L. Udehn (1998), "Popper's situational analysis and contemporary sociology", *Philosophy of the Social Sciences*, **28** (3): 339–64.
Hedström, P. and P. Ylikoski (2010), "Causal mechanisms in the social sciences", *Annual Review of Sociology*, **36**: 49–67.
Hedström, P. and P. Ylikoski (2014), "Analytical Sociology and Rational Choice Theory", in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley.
Hoover, K.D. (2017), "Situational Analysis", pp. 182–90, in L. McIntyre and A. Rosenberg (eds), *The Routledge Companion to Philosophy of Social Science*, London: Routledge.
Jarvis, B., M. Keuschnigg and P. Hedström (forthcoming), "Analytical sociology amidst a computational social science revolution", in U. Engel et al. (eds), *Handbook of Computational Social Science*, Routledge.
Keuschnigg, M., N. Lovsjö, and P. Hedström, (2018), "Analytical sociology and computational social science", *Journal of Computational Social Science,* **1** (1): 3–14.
Land, K.C., G. Deane and J.R. Blau (1991), "Religious pluralism and church membership: A spatial diffusion model", *American Sociological Review*, **56** (2): 237–49.
Latané, B. (1981), "The psychology of social impact", *American Psychologist*, **36** (4): 343–56.
Lopez-Pintado, D. and D.J. Watts (2008), "Social influence, binary decisions and collective dynamics", *Rationality & Society*, **20** (4): 399–443.
Manzo, G. (2007), "Comment on Andrew Abbott", *Sociologica*, **2**. doi: 10.2383/24752.
Manzo, G. (2014), "Data, generative models, and mechanisms: More on the principles of analytical sociology", pp. 1–52, in G. Manzo (ed.), *Analytical Sociology: Actions and Networks*, Chichester, UK: Wiley.
Manzo, G. and A. van de Rijt (2020), "Halting SARS-CoV-2 by targeting high-contact individuals", *Journal of Artificial Societies and Social Simulation*, **23** (4).
Merton, R.K. (1936), "The unanticipated consequences of purposive social action", *American Sociological Review*, **1**: 894–904.

Merton, R.K. (1948), "The Self-Fulfilling Prophecy", pp. 475–90, in Robert K. Merton (ed.), *Social Theory and Social Structure*, New York: The Free Press.
Merton, R.K. (1967), "On sociological theories of the middle range", pp. 39–72, in Robert K. Merton (ed.), *On Theoretical Sociology*, New York: The Free Press.
Merton, R.K. (1968), "The Matthew effect in science: The reward and communication systems of science are considered", *Science*, **159** (3810): 56–63.
Nowak, A., J. Szamrej and B. Latané (1990), "From private attitude to public opinion: A dynamic theory of social impact", *Psychological Review*, **97** (3): 362–76.
Popper, K.R. (1961), *The Poverty of Historicism*, London: Routledge.
Popper, K.R. (1994), "Models, instruments, and truth: The status of the rationality principle in the Social Sciences", pp. 154–84, in K.R. Popper (ed.), *The Myth of the Framework: In Defence of Science and Rationality*, London: Routledge.
Salganik, M.J., P.S. Dodds and D.J. Watts (2006), "Experimental study of inequality and unpredictability in an artificial cultural market", *Science*, **311** (5762): 854–6.
Schelling, T.C. (1971), "Dynamic models of segregation", *Journal of Mathematical Sociology*, **1**: 143–86.
Stewart, J.Q. (1941), "An inverse distance variation for certain social influences", *Science*, **93** (2404): 89–90.
Stewart, J.Q. (1947), "Empirical mathematical rules concerning the distribution and equilibrium of population", *Geographical Review*, **37** (3): 461–85.
Watts, D.J. (2014), "Common sense and sociological explanations", *American Journal of Sociology*, **120** (2): 313–51.