In this chapter I argue for the importance of a specific kind of abstract analytical theorizing that differs from many other types of sociological theorizing in its focus on social mechanisms. A social mechanism, as defined here, is a constellation of entities and activities that are linked to one another in such a way that they regularly bring about a particular type of outcome. We explain an observed phenomenon by referring to the social mechanism by which such a phenomenon is regularly brought about.

The purpose of the chapter is threefold. First, I very briefly discuss different notions of theorizing within contemporary sociology. Sociology is a fragmented discipline and theorizing means different things to different sociologists. Second, I consider at some length what ‘explanation’ entails. Sociologists differ not only on what they consider to be the objective of theorizing, but also on the meanings they attach to the concept of ‘explanation.’ Third, I discuss in some detail the main thrust of the mechanism-based approach developed here.

As this brief synopsis indicates, in this chapter I touch upon a range of difficult problems related to causation and explanation. It goes without saying that a discussion of such vast and difficult problems runs the risk of being perceived as shallow. But the purpose of the chapter must be kept in mind. My intention is not to contribute to the theory of explanations or to the philosophy of the social sciences. The purpose is to outline an explanatory framework that I believe to be of crucial importance for sociology, and in doing this I draw upon some of the relevant philosophical literature. This literature is of considerable importance for sociology because it identifies general criteria that can assist us in distinguishing between adequate and inadequate explanations. By being rigorous about what an adequate explanation should look like, we are likely to arrive at better theories and research than otherwise. Sociology is likely to benefit from a strict ‘Ulyssian’ strategy of precommitting itself to certain explanatory standards, or so I suggest here.
Varieties of sociological theorizing

In most scientific fields and in much of the philosophy of science literature, ‘theories’ are seen as abstract codifications of knowledge that allow us to explain and predict events and processes, and a ‘theorist’ is a person who contributes to the development of such theories. But this is not always the case in sociology. Some sociological theorists see it as their main mission to voice and give expression to deeply felt sentiments and concerns in society (e.g., Beck and Ritter 1992). Others see their role as interpreters of societal trends and conditions (e.g., Bauman 2001; Castells 2000), and yet others view social theory as a predominantly normative endeavour, the main purpose of which is to criticize and/or suggest alternatives to existing social, cultural and economic orders (e.g., Habermas 1987).

The fragmentation of the discipline is readily seen in recent ‘handbooks’ of sociological and social theory, such as those by B. S. Turner (1996), Sica (1998) and Ritzer and Smart (2001). The editors of these volumes seem to equate theory with everything from general reflections on modernity to exegetic digressions on the founding fathers of the discipline. Whatever value these various non-explanatory traditions in social theory may have, they are not the concern of this book. In my view, too much social theory currently falls within these non-explanatory traditions, and it therefore appears more essential to try to develop and strengthen the traditional canon of explanatory theorizing.

What, then, does it mean to ‘explain’ something? In answering this question, let me start with stating what an explanation is not. First of all, explanations are not descriptions. As emphasized by Sen (1980), the choices and judgements involved in producing good descriptions

---

1 The former editor of Sociological Theory, Craig Calhoun, made a similar observation when trying to characterize the submissions he received to the journal. He noted that the submissions all too often were ‘summaries of what dead people said (with no indication of why living ones should care or how the revered ancestor’s work would advance contemporary analytic projects)’ and ‘criticisms of what other people have said that dead people said (with no more indication of why we should care than that those criticized are famous)’ (Calhoun 1996: 1).

2 A telling sign of the current state of the discipline, at least in Britain, is the official website of the British Sociological Association (http://www.britsoc.co.uk). The history of the discipline there is approvingly described as a movement from explanations to reflections. According to the BSA, in the nineteenth century there was a view of the discipline as offering ‘explanations of the collective entities and relationships of human beings’. This by now outmoded view has been replaced by a ‘modern’ reflexive view: ‘From this original purpose . . . sociology has moved on to more reflexive attempts to understand how society works.’
resemble those faced when one makes predictions or proposes explanations, but descriptions differ in important respects from explanations. A description of something tells us how it is constituted, or how it varies over time or between different groups or social settings, but it does not say *why* it looks, changes or varies as it does. This also holds true for a common form of ‘thick’ descriptive work, where one ‘interprets’ a phenomenon by describing it with a vocabulary borrowed from some specific sociological tradition such as symbolic interactionism (see Charon 2001 for various examples). The exact boundaries between descriptions and explanations are not always clear-cut, but it should be noted that ‘to explain’ and ‘to describe’, although interrelated, provide answers to different types of questions.

Second, explanations differ from *typologies* and *taxonomies*. Much sociological theory consists of typologies. We have class typologies (e.g., Wright 1997), welfare-state typologies (e.g., Esping-Andersen 1990), and typologies of historical trajectories (e.g., Therborn 1995), to mention a few. Such typologies can be extremely useful in that they can create order out of otherwise perplexing chaos. But typologies are not explanations. They are classificatory devices that allow us to attach labels to different phenomena in an orderly fashion, but they do not tell us why we observe the phenomena we observe. Thus, as the terms are used here, we can have a typology of explanations and we can have a typology based on explanations, but we cannot have an explanatory typology.

A basic characteristic of all explanations is that they provide plausible causal accounts for why events happen, why something changes over time, or why states or events co-vary in time or space. At least three types of explanations can be identified in the literature that differ in terms of the types of answers they consider appropriate to such why-questions.

These three types of explanations are (1) covering-law explanations,
Before discussing these different types of explanations, we must deal with one terminological issue. Most philosophers of science insist that causes and effects must be events, while sociologists and other social scientists also refer to social states and various individual attributes as potential causes and effects. In many respects the former terminology appears more precise and appropriate for the simple reason that causes bring things about, and what is brought about (that is, the effect) cannot have been there before, and therefore what is brought about must be a change or an event. Similarly, it is difficult to see how change can be brought about except by another change, which suggests that causes are events too. Nevertheless, I use standard sociological terminology and refer to states and other non-events as potential causes and effects as well. Substantively this does not seem to be of much import. From an event-causation perspective, one would not view such entities as proper causes, but one could see them as conditions that make it possible for one event to cause another (Lombard 1990). Similarly, non-events could, as effects, be viewed as the aggregate (or otherwise combined) outcome of a series of events.

Unless otherwise noted, I will use ‘mechanisms’, ‘causal mechanisms’ and ‘social mechanisms’ as synonymous terms. Also, although my choice of wording may be a bit ambiguous here and there, a ‘mechanism’ refers to the real empirical entities and activities that bring about phenomena. These should be distinguished from theories or models of mechanisms, since we otherwise are likely to commit what Whitehead (1930: 52) called ‘the fallacy of misplaced concreteness’. Most of the discussions in this book are concerned with the models of or theories about real mechanisms.
Covering-law explanations

One of the most influential notions of what an appropriate explanation should look like is the ‘covering-law model’ most closely associated with the work of Carl Hempel (e.g., Hempel 1965). Assume that we have an event, $e$, that we seek to explain. In order to provide an appropriate answer to the question ‘Why did $e$ happen?’ we subsume the event under a general law. That is, we explain $e$ by pointing to one or several general laws and the conditions that make these laws applicable to the specific case.\footnote{Although Hempel specified in detail the logic of this type of explanation, the basic idea behind the covering-law model has been around for a long time and was formulated in the following way by John Stuart Mill: ‘An individual fact is said to be explained, by pointing to its cause, that is, by stating the law or laws of causation, of which its production is an instance’ (1874: 332). See also Braithwaite (1953) for another influential work with a view similar to that of Hempel.}

To illustrate the logic of his proposal Hempel often used the example of an automobile radiator cracking during a freezing-cold night. The general laws cited in the explanation would need to refer to how the pressure exerted by water varies with changes in temperature and volume, and the initial conditions referred to would be conditions such as the temperature during the night and the bursting pressure of the radiator. A proper explanation has been proposed if, and only if, the sentence describing the cracking of the radiator can be logically deduced from the sentences stating the laws and the initial conditions. From a covering-law perspective, the appropriate answer to the question ‘Why do we observe phenomenon $X$?’ is that $X$ was expected given the existence of certain causal laws.

To the extent that general laws of the kind ‘All $A$ are $B$’ exist, Hempel’s Hempel’s proposal seems highly attractive. If $B$ is a property describing society $x$, a perfectly reasonable answer to the question ‘Why is $x$ a $B$?’ would be that society $x$ is an ‘$A$’ and ‘All $A$ are $B$’. But although they are attractive in principle, such explanations are of limited relevance to the social sciences because we do not yet know of any general laws of the ‘All $A$ are $B$’ kind, and human agency seems to render such laws highly implausible in the social and the cultural sciences.

The ‘law-like’ relationships that it might be possible to establish in the social sciences are instead of a probabilistic nature, and Hempel proposed a different explanatory model in such situations. The differences in the explanatory logic between these two models can be described in the following way:
1. All $A$ are $B$.
   \[ \begin{align*} &x \text{ is } A. \\ &x \text{ is } B. \end{align*} \]

2. Most $A$ are $B$.
   \[ \begin{align*} &x \text{ is } A. \quad [p] \\ &x \text{ is likely to be } B. \end{align*} \]

The single line in (1) represents a deductive relation between the premises and the conclusion. The double line in (2) represents an inductive relation, and $p$ states the conditional probability of the conclusion given the premises. Hempel referred to (1) as the ‘deductive-nomological’ model, and to (2) as the ‘inductive-probabilistic’ model.

Once again, Hempel’s proposal seems reasonable as far as it goes. The problem is that it does not seem to go very far. First of all, as pointed out by Nagel (1961), Salmon (1971) and others, the covering-law model does not seem to describe adequately the defining characteristics of what are generally considered to be acceptable scientific explanations. First, there exist generally accepted scientific explanations – not the least in the social sciences – that would not be considered acceptable if we were to follow Hempel’s model because of the difficulty of specifying any relevant laws. Second, there exist statements that fulfil all of Hempel’s logical requirements but which nevertheless are not explanatory. The following ‘explanation’ is a case in point. If we wanted to explain the fact that Peter did not become pregnant, the following line of reasoning would appear to be acceptable from Hempel’s perspective (adopted from Salmon 1971):

\[
\begin{align*}
\text{No one who regularly takes birth-control pills becomes pregnant.} \\
\text{Peter regularly takes birth-control pills.} \\
\text{(Therefore) Peter did not become pregnant.}
\end{align*}
\]

The fact to be explained can be logically deduced from the premises – both of which can be assumed to be true – but the explanation is nevertheless incorrect because it refers to the wrong causal mechanism.

Furthermore, and related to the latter type of objection, Hempel’s model is not sufficiently restrictive in the sense that it does not rule out obviously superficial explanations. Hempel’s form of explanation entails applying a law to a specific situation. The insights offered by this exercise depend on the depth and robustness of the ‘law’. If this ‘law’ is only a statistical association, which is the norm in the social and cultural sciences (according to Hempel as well), the specific explanation will offer no more insight than the statistical association itself and will usually
only suggest that an event is likely to happen but give no clue as to why this is likely to be the case.

Consider the following example. It would be possible to statistically estimate the parameters of an equation describing the relationship between the intake of, say, strychnine and the risk of dying. If the statistical model had the correct functional form and took into account relevant factors such as body weight, we could describe the result as a ‘probabilistic law’ of the dose–response relationship, and we could use this ‘law’ as one of the premises when explaining why individual x died after ingesting a specific dose of strychnine: given individual x’s intake of strychnine and our ‘probabilistic law’, we could have predicted what would happen. In this sense we would also have explained the death of x because this outcome was expected given the initial conditions and the ‘probabilistic law’.

Such an explanation seems wanting, however. When posing such questions in a scientific context we normally expect answers that not only state that the event was likely because this is what has happened in the past, we also want to know why this is so. Below I discuss in some detail the important role played by causal mechanisms in providing such answers, but it already seems clear that what is required is some form of mechanism that provides an intelligible link between the causal factor and the event to be explained. By pointing to how strychnine typically inhibits the respiratory centre of the brain and to the biochemical processes typically responsible for such paralysis, we provide a mechanism that allows us not only to predict what is likely to happen but also to explain why (Bunge 1967). For these reasons, I am inclined to agree with von Wright that it is better ‘not to say that the inductive-probabilistic model [of Hempel] explains what happens, but to say only that it justifies certain expectations and predictions’ (von Wright 1971: 14).

The covering-law model has encouraged and legitimized a type of theorizing that I do not think has been entirely conducive to the development of a rigorous body of explanatory theory. The problem is not with the mode of theorizing as such. In fact, the precision and clarity of those endorsing this mode of theorizing often greatly surpass those of other theorists. The problem is rather the aforementioned lack of restrictions on the content of the propositions. Let me examine one semi-classic example, Peter Blau’s theory of organizational differentiation (Blau 1970).

Blau’s theory of organizational differentiation, as noted by Calhoun, Meyer and Scott (1990), was in many respects a direct precursor of his later so-called macro-sociological theory of social structure (e.g., Blau...
and it exemplifies his approach to theorizing even more clearly than the latter theory. With reference to the work of Hempel and Braithwaite, Blau argued that general propositions of the aforementioned kind are at the heart of all explanations: ‘Inasmuch as the generalizations [that is, the general propositions] subsume many empirically demonstrated propositions, that is, logically imply them, they explain these regularities’ (Blau 1970: 202).

Blau’s general ‘law-like’ propositions were the following:

1. Increasing size generates structural differentiation in organizations along various dimensions at decelerating rates.
2. Structural differentiation in organizations enlarges the administrative component.

Expressed more plainly, these propositions say that when organizations grow in size the elements of their internal composition – for example, the number of different job tasks and the number of departments – also tend to increase, but the increase will gradually level off the larger the organization gets. This ‘structural differentiation’, in turn, tends to increase the proportion of administrative personnel in the organization.

On the basis of these general propositions Blau then deduced a set of ‘lower-level’ propositions that follow from the simultaneous operation of these two law-like propositions. The details need not concern us here, however, since we are exclusively interested in Blau’s theory as an example of a type of sociological theorizing that has been much influenced by the writings of Hempel and other logical positivists. According to Blau, theories are systems of general propositions, and explanations are arrived at by subsuming the events to be explained under the law-like regularities expressed in these propositions.

It is also of interest to note that theorists in this tradition often seem to consider it a strength if their explanations make no reference to actions or to ‘psychological’ phenomena such as beliefs and desires. In the words of Blau (1970: 203): ‘The theory centers attention on the social forces that govern the interrelations among differentiated elements in a formal structure and ignores the psychological forces that govern individual behavior. Formal structures exhibit regularities that can be studied in their own right without investigating the motives of the individuals in organizations.’ In similar vein, Donald Black (1979: 149–50) summarized his approach to ‘pure sociology’ in the following way: ‘[Pure sociology] has nothing to say about how people experience themselves, their freedom of choice, or the causes of their actions . . . It is a way to predict
and explain the behavior of social life, and that is all’ (see also Black 1976; Black 2000; Mayhew 1980; Mayhew 1981).

The type of approach advocated by Blau and Black follows rather naturally if one adopts the covering-law model. However, a basic flaw with this mode of theorizing, as I see it, is that it excludes from the explanation exactly those processes that would have allowed us to understand why social entities exhibit the regularities they do. In this respect, theories like these are wanting for the same reason that the correlation-based explanation of the relationship between the intake of strychnine and the risk of dying is wanting: they are both black-box explanations that exclude from focus those processes that would allow us to understand why a specific causal factor is likely to be of explanatory relevance.8 What are lacking in the approach of Blau, Black and other methodological ‘holists’ are the basic entities and activities that generate these correlations. The most reasonable ontological hypothesis we can formulate in order to make sense of the social world as we know it is that it is individuals in interaction with others that generate the social regularities we observe. Hence, social interaction processes are the parallels to the biochemical processes in the strychnine example. Both constitute the intelligible links between the explanatory factors and the events to be explained that are required to answer the type of why-questions normally posed by explanatory sciences, and it is these types of processes that are missing in Blau’s type of theory.9 Even if one suspects that there are ‘nomic’ law-like social regularities, it seems more reasonable, as von Wright (1989: 838) once pointed out, to try to understand first why this is so before we accord explanatory force to the ‘law’ that it is so. For these reasons, and as is discussed in more detail below, we are likely to arrive at better and more precise sociological theories if they are based upon explicit theories of action.

Blau defended his methodological position with a line of argumentation common to many of those advocating similar holistic or pseudo holistic approaches: ‘Social structures (and indeed all structures composed of subunits) have emergent properties that cannot be understood on the basis of the properties of the subunits’ (Blau 1986: ix). Blau never explained why he believed this to be the case, that is, whether his belief was based on ontological or methodological considerations. Therefore

---

8 As noted by Goldstone (1998), an explanatory logic similar to the one criticized here is also at the heart of much of contemporary comparative-historical work.

9 See Harré (1985) for a similar view of the parallels between causal mechanisms in the natural sciences and reason-based explanations in the social sciences. See also Abell (2004).
his statement remained rather vacuous. Although it may be difficult to explain why we observe the social phenomena we observe, I fail to see why, in principle, it would be impossible to explain social phenomena on the basis of the properties and activities of subunits and the way in which they are linked to one another. I return the question about emergence in chapter 4.

Where, then, does this leave us? Although the covering-law model has many attractive features, I do not think that the model as such is particularly useful for sociology. The main reasons are the following:

1. The deductive-nomological model is not applicable because the deterministic social laws that it presupposes do not exist.
2. The inductive-probabilistic model is not useful as an explanatory model because (a) it allows for and thereby legitimizes superficial theories and explanations, and (b) it does not give action and intentional explanations the privileged role they should have.

---

**Statistical explanations**

While the covering-law model is often referred to in discussions of explanatory strategies in sociology, it is rarely relied on in practice. The type of explanation to be discussed next, statistical explanation, is its opposite in this respect: it is at the heart of most empirical research in sociology but it is rarely discussed in contexts such as this.

The statistical type of explanation differs in important respects from a covering-law explanation. Most importantly, while covering-law explanations are theory-based in the sense that they use existing theories or laws to explain specific events through a deductive argument, statistical explanations are much more inductively oriented and typically do not presuppose any well-specified theories.

The defining characteristic of what here is referred to as a statistical explanation is that an appropriate explanation is at hand when we have identified factors that seem to make a difference to the probability of the event one seeks to explain. The identification of such factors is typically accomplished by decomposing the relevant population into different categories. The logic is best described with a hypothetical example. Much sociological research seeks to answer questions such as: ‘Why do we observe a gender gap in earnings?’; ‘Why has the support for political party X eroded over time?’; or ‘Why are revolutions more likely to occur in certain nations than in others?’ The questions refer to a difference between different entities – in these examples between men and women,
different time periods, and different nations, respectively – and the answers given are typically based on breaking down the relevant population into different subpopulations. In order to answer the question about the gender gap in earnings, for example, a decomposition as in figure 2.1 may be required.

In order to explain the observed difference in the average earnings between men and women, the population is decomposed. First we decompose it into four different groups based on gender and education, and we examine the average earnings within each of these groups. If at this point the gender differences disappear, that is, if men and women with the same educational levels are paid approximately the same, we may conclude that the explanation for the earnings gap might be that there are relatively more women than men with a low level of education (on the assumption that the highly educated earn more than the poorly educated). If the earnings gap does not disappear, we would further decompose the population. In this example, the next level of decomposition is in terms of work experience. If the gender differences disappear at this stage, we may conclude that the earnings gap is explained by gender differences in education and work experience. If not, we would continue the decomposition and introduce additional factors that could possibly explain the observed earnings differential.

This example is stylized. In reality, more fine-grained decompositions are used and partial accounts of observed differences are all that can be hoped for. In addition, such explicit decompositions are rarely seen in leading journals today; instead, some form of regression model is typically used to perform the same task. Nevertheless, the logic behind the analysis can best be understood in these terms: differences in some social states or events are considered to be explained if the decomposition eliminates them, and they are considered to be partially explained if it partially eliminates them.¹⁰

Unlike the search for ‘social laws’ discussed in the previous section, statistical analysis is a useful, and in many applied situations the most useful, strategy to pursue. In addition, when the objective is to develop explanatory sociological theory, such analyses may be important for ‘establishing the facts’ that need to be explained, as suggested by Goldthorpe (2000) and others. With reference to the previous example, for instance, rather different theories would seem to be needed if the

¹⁰ In the context of regression analysis, the extent to which observed differences are eliminated is typically measured as the proportion of the variance in the outcome variable being ‘removed’ when the explanatory factors are introduced.
Figure 2.1. Hypothetical decomposition used to answer the question ‘Why do we observe a gender gap in earnings?’
gender gap in earnings were due to educational differences between men and women rather than discriminatory practices at the workplace. Without a proper initial decomposition, we may end up focusing our attention on pseudoproblems or on problems that have little or no bearing on the empirical phenomenon that we seek to explain.11

Such statistical analysis is often described as a form of ‘causal analysis’. If a factor appears to be systematically related to the expected value or the conditional probability of the outcome, then the factor is often referred to as a (probabilistic) ‘cause’ of the outcome. Although it makes little sense to quibble over words, I would like to reserve the word cause for a less casual notion of causality.

Richard Swedberg and I have discussed the limitations of this ‘variable approach’ in more detail elsewhere (see Hedström and Swedberg 1998b), and I will not repeat that discussion here (but I return to the variable approach and its limitations in chapter 5). In brief, however, I do not believe that a view of theories and explanations as lists of statistically relevant factors is conducive to the development of a rigorous body of sociological theory. Statistical regularities are rarely (if ever) as unequivocal and easily interpretable in causal terms as this view would seem to suggest. As Stinchcombe (1968: 13) once expressed it, ‘A student who has difficulty thinking of at least three sensible explanations for any correlation that he is really interested in should probably choose another profession.’ Phenomena such as gender differences in earnings, voting trends, political upheavals and most other concerns of sociologists are clearly the result of highly complex social processes. The belief that one should be able to ‘read off’ their causes by observing relationships between variables such as those discussed above has always seemed a bit naïve to me. Rather than trying to establish ‘phenomeno-phenomenological laws’ on the basis of statistical analyses, Boudon’s suggestion (1976: 117) that we should ‘go beyond the statistical relationships to explore the generative mechanism responsible for them’ seems to be a more promising path forward, and it is the one pursued in chapter 6. Statistical analyses are important for testing proposed explanations, but it must be remembered that a statistical analysis is a test of an explanation, not the explanation itself. This distinction is often obliterated in the statistically oriented tradition.

11 As is discussed in chapter 5, however, one should always be aware that statistical analysis, particularly when it is based on implausible assumptions or on ad hoc statistical models with numerous independent variables, can establish ‘artifacts’ rather than ‘facts’, thus hindering us rather than helping us to arrive at appropriate answers (see e.g., Freedman 1991).
Mechanism-based explanations

Although the search for law-like relationships between different social entities – in both their covering-law and causal-modelling guises – seems to be part of sociology’s past rather than its future, this does not preclude the possibility of developing rigorous explanatory theories. By focusing on the mechanisms that generate change in social entities, rather than on statistical regularities between variables, a foundation for powerful explanations can be established. Such mechanism-based explanations are at the core of the analytical approach, and the rest of this book is, in one way or another, concerned with explicating and refining this type of explanatory approach. In the remainder of this chapter I focus on the abstract logic of mechanism-based explanations and show how such explanations differ from covering-law and statistical explanations, and in later chapters I give the idea concrete expression in empirical examples.\(^\text{12}\)

The core idea behind the mechanism approach is that we explain not by evoking universal laws, or by identifying statistically relevant factors, but by specifying mechanisms that show how phenomena are brought about. Philosophers and social scientists have defined the mechanism concept in numerous ways (e.g. Bhaskar 1978; Bunge 1996; Elster 1999; Gambetta 1998; Glennan 1996; Hedström and Swedberg 1998b; Karlsson 1958; Little 1991; Mahoney 2001; Mayntz 2004; Mcadam, Tarrow and Tilly 2001; Pawson 2000; Salmon 1984; Schelling 1998). Figure 2.2 describes some of the currently most cited definitions of what characterizes a mechanism.

These definitions differ a great deal from one another. Some definitions refer to causal mechanisms in general, while others refer exclusively to social mechanisms; some definitions refer to concretely existing entities, while others refer to models or reasoning about such entities. Underlying them all, however, is an emphasis on making intelligible the regularities being observed – a mechanism explicates the details of how the regularities were brought about.

\(^{12}\) A focus on mechanisms similar to the one advocated here can be found in the work of Bhaskar and other ‘critical realists’ (see Archer et al. 1998 for an overview). Much of what the critical realists have to say about explanatory strategies in the social sciences I find useful and interesting, but others have said most of this with greater precision. The set of ideas that can be said to be unique to the critical realists tends to be too vague and too normatively oriented to be of much scientific use. Hence, despite somewhat overlapping concerns, there are not many references to the work of the critical realists in the following pages, except in chapter 4 where I critically discuss their notion of a stratified reality.
The most satisfactory conceptual analysis of the mechanism concept is found in Machamer, Darden and Craver (2000). The spirit of their approach is very similar to the Elster I and Hedström–Swedberg approaches. If one builds upon these ideas, mechanisms can be said to consist of entities (with their properties) and the activities that these entities engage in, either by themselves or in concert with other entities. These activities bring about change, and the type of change brought about depends upon the properties of the entities and the way in which they are linked to one another. A social mechanism, as here defined, describes a constellation of entities and activities that are organized such that they regularly bring about a particular type of outcome. We explain an observed phenomenon by referring to the social mechanism by which such phenomena are regularly brought about.

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

Figure 2.2. Alternative mechanism definitions.
From the mechanism perspective, correlations and constant conjunc-
tions do not explain but require explanation by reference to the entities
and activities that brought them into existence. The explanatory mech-
anisms that we seek to develop should be ‘final’ in Boudon’s (1998a)
sense of the term. That is to say, the mechanism should not include any
glaring black boxes which simply give rise to additional why-questions.
As discussed in chapter 3, this means that action-based explanations are
at the core of all social mechanisms.13

According to Craver (2001), theories in the biological sciences typic-
ally refer to mechanisms that are hierarchically nested: that is, they refer
to mechanisms nested within other mechanisms. This is also the case in
the social sciences, and this is why Stinchcombe (1991) once defined
mechanisms as theories-within-theories or as pieces of theory that give
knowledge about components of another theory. As emphasized above,
sociological theories typically seek to explain social outcomes such as
inequalities, typical behaviours of individuals in different social settings,
and social norms. In such theories individuals are the core entities and
their actions are the core activities that bring about the social-level
phenomena that one seeks to explain. The way in which these actors
are linked one to another defines the structure of interaction, and this is
likely to influence in its own right the social outcomes brought about.
That is to say, the same entities (individual actors) strung together in
different ways can be expected to regularly bring about different types
of outcome.14 In this sense, different types of structural configurations of
actors can be said to constitute different social mechanisms.

Nested within these ‘molecular’ mechanisms are more elementary
mechanisms that explain the actions of individual actors. Also in this
case the mechanisms can be described in terms of their entities (and
their properties) and the way in which the entities are linked to one
another. The core entities are different, however, and now include the
beliefs, desires and opportunities of the actors, but the explanatory logic
is the same: we explain an observed phenomenon, in this case an indi-
vidual action, by referring to the mechanism (that is, the constellation of

13 It should be noted that the criterion of what constitutes a ‘final’ explanation varies from
discipline to discipline. For example, while in sociology an intentional explanation can
be considered final because it allows us to understand why actors do what they do, a
neuroscientist would consider it to be a black-box explanation. I return to the issue of
discipline-specific stopping rules later in this chapter.
14 The ways in which social networks or relational structures influence the social outcomes
actors are likely to bring about are examined in some detail in chapter 4.
beliefs, desires and opportunities) by which such phenomena are regularly brought about.\(^{15}\)

One possible objection to explanations that seek to explicate generative mechanisms ‘beneath’ the surface of observed regularities is that they may lead to an infinite regress (e.g., Kincaid 1996; King, Keohane and Verba 1994). For example, is not the insistence on mechanisms as theories-within-theories simply a way of moving the ‘black box’ down a level – from that of the theory to that of the theory-within-the-theory? To be consistent, should not the mechanisms of the theory-within-the-theory also be specified in terms of yet deeper mechanisms (that is, with a theory-within-the-theory-within-the-theory), and these, in turn, in terms of even deeper mechanisms? This regress could in principle continue for ever, or at least until we have reached the level of inexplicable laws of nature. In the end, then, we may be forced to accept a traditional regularity view of causation, and the critical reader may wonder if it is not better, or at least more consistent, to adopt a traditional Humean approach from the very start.

What is perceived to be a ‘black box’ and a ‘mechanism’ certainly depends upon the resolution of the theoretical lens through which we view a problem. It also has an important historical dimension.\(^{16}\) But from these observations it does not follow that the insistence on mechanism-based explanations is unfounded or that a traditional regularity view of causation and explanation would be preferable. Even if it were possible to carry out a reduction in the manner described in the previous paragraph – impossible in practice – the resulting explanation is not likely to be of much sociological relevance. There exist discipline-specific relevance criteria and ‘stopping rules’ (Miller 1987) that at least roughly stipulate what types of explanatory factors are considered relevant within different academic disciplines. Although, as noted above, sociologists differ in what they consider to be the most appropriate stopping rules – Blau and Black, for instance, advocate different stopping rules from those that I do – non-explainable laws of nature are far outside the domain of sociological relevance. For this reason I would

---

\(^{15}\) The way in which different constellations of beliefs, desires and opportunities are likely to influence individuals’ actions is examined in chapter 3.

\(^{16}\) As noted by Patrick Suppes (1970: 91), ‘From the standpoint of either scientific investigation or philosophical analysis it can fairly be said that one man’s mechanism is another man’s black box. I mean by this that the mechanisms postulated and used by one generation are mechanisms that are to be explained and understood themselves in terms of more primitive mechanisms by the next generation.’
not consider the infinite-regress objection to be of much importance, and I would therefore maintain that mechanism-based explanations are what sociological theory should be all about.

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 more precise and intelligible explanations. Another important reason is that a focus on mechanisms tends to reduce theoretical fragmentation. For example, we may have numerous different theories (of crime, organizations, social movements or whatnot) that are all based on the same theory-within-the-theory, that is, they all refer to the same set of mechanisms of action and interaction. Focusing on the mechanisms as such avoids unnecessary proliferation of theoretical concepts and may help to bring out structural similarities between seemingly disparate processes. Finally, it is the knowledge about the mechanism as such, that is, knowledge about why the constellation of entities and activities referred to in the explanation can be expected to regularly bring about the type of outcome we seek to explain, 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 correlation.

Although the explanatory focus of sociological theory is on social entities, an important thrust of the analytical approach is that actors and actions are the core entities and activities of the mechanisms explaining such phenomena. There are at least three important reasons why this is the case. First, it is a well-established scientific practice that theories should be formulated in terms of the processes that are believed to have generated the phenomena being studied. In sociology, this realist principle assigns a unique role to actions because actions are the activities that bring about social change. The causal efficacy of actions would be readily seen if we were able to press a pause button that suddenly froze all individuals and prevented them from performing any further actions. All social processes would then come to an immediate halt. Second, action-based explanations are, in one particular respect, more intellectually satisfactory than the available alternatives. Focusing on actions and explaining actions in intentional terms provides a deeper and more emphatic understanding of the causal process than do other

17 As is discussed in more detail in chapter 4, this does not mean that ‘social structure’ is unimportant, only that ‘structural’ effects need actors in order to be instantiated.
non-action-based explanations.\(^{18}\) Third, action-based explanations tend to reduce the risk of erroneous causal inferences. As noted by Skog (1988) and others, there is considerable risk of mistaking spurious correlations for genuine causal relationships when one focuses on macro-level trends and correlations. One telling example used by Skog is the high correlation often found between sunspot activity and various social phenomena. The correlation between sunspot activity and the prevalence of intravenous drug use in Stockholm during the period 1965–70, for example, was as high as 0.91. Action-based explanations can help to eliminate such spurious causal accounts in the following way: if it proves impossible to specify how the phenomenon to be explained could have been generated by the actions of individuals, or if the account must be based on highly implausible assumptions, one’s faith in the proposed causal account is sharply reduced.

A concern often raised is that a focus on actions and micro-level mechanisms may lead to a loss of valuable information and therefore to a biased understanding of the phenomena being studied. Sometimes it is even suggested that macro-levels and micro-levels are partly independent of each other, or that it is at least useful to assume this to be the case (e.g., Brante 2001). A similar view seems to have motivated Jackson and Pettit (1992b) to advocate a particular type of explanation called a ‘programme’ explanation. They describe their core idea in the following way:

The idea is that a structural factor may explain a given social fact, not through producing it in the same basic way as individual factors, but through more or less ensuring that there will be some individual-level confluence of factors – perhaps this, perhaps that – sufficient to produce it. (Jackson and Pettit 1992b: 120)\(^ {19}\)

The situation Jackson and Pettit describe is illustrated in figure 2.3. The solid lines represent causal processes through which \(C_1\) and \(C_2\) can bring about \(E\), and the dotted line represents the ‘ensuring condition’, which states that in situation \(S\), should \(C_1\) fail to occur, \(C_2\) would occur, and \(C_2\) would bring about \(E\) (and vice versa should \(C_2\) fail to occur).

While I find the argument that social entities sometimes have these ‘effects’ to be persuasive, I do not find the arguments for ‘programme’ explanations equally so. The reason for this can be stated as follows:

\(^{18}\) The notion of ‘intentional explanations’ is discussed in chapter 3.

\(^{19}\) See also Jackson and Pettit (1992a). A well-known sociological work that rests on this type of argument is Skocpol’s (1979) so-called structural theory of revolutions.
1. Most of us would agree that the empirical observation that $E$ is more likely to be brought about in situation $S$ than in not–$S$ provides important and useful information (although many of us would not think of it as explanatory).

2. Most of us would agree that detailing the mechanisms through which various $C$s can bring about $E$ is of obvious explanatory importance. This is also true in situations of causal overdetermination (like the one described above) because we thereby reduce the causal possibility space.

3. Given the empirical information in (1) and the information about the causal processes in (2), the introduction of the programme explanation does not seem to add any new information unless the causal processes linking $S$ and the $C$s are explicated, but then we are not talking about a programme explanation any more.

Thus, it seems that reasonable programme explanations can be expressed as mechanism-based explanations, albeit of a more complex nature. If this is not possible, the programme explanation does not provide any information not already available from (1). This conclusion can, I believe, be generalized to any problem describable in these terms.

**Differences and similarities**

Let me highlight some of the most important differences and similarities between the three explanatory traditions discussed. With their focus on theory and systematic deductive arguments, mechanism-based explanations are in many ways closer to covering-law explanations than to statistical explanations. While a statistical explanation consists of an assembly of factors that appear to make a difference to the probability of the event being studied, a mechanism-based explanation, like a covering-law explanation, can be characterized as a theoretical deductive argument. The theoretical arguments differ in numerous ways, however. For instance, covering-law explanations typically refer to causal factors

![Figure 2.3. Components of a programme explanation.](image)
and not to causal processes (Mayntz 2004). Furthermore, the laws of covering-law explanations are typically seen as perfectly general and without exceptions while the mechanisms of mechanism-based explanations are not (Elster 1989b), and mechanism-based explanations are action-based while covering-law explanations typically are not.

An obvious difference between statistical explanations and mechanism-based explanations is that randomness and the stochastic nature of social processes are more central concerns of the statistical tradition. Although I certainly do not wish to deny the importance of randomness, I believe it is useful to formulate theories and explanations as if the world were deterministic. One important reason for this is that only as a last resort should one attribute the unexplained to the unexplainable. As we all know, all cats look grey at night, and we should therefore evoke randomness as an explanation only when all plausible alternatives have been proven unsatisfactory. Second, a deterministic language is preferable because of its economy of expression. As long as we are aware that theoretical statements refer to tendencies and not to actual processes, a deterministic language seems better suited to communicate the core ideas about the mechanisms assumed to govern the processes.20

A mechanism should thus be seen as an empirical commitment on the part of the theorist as to how a process would unfold if the assumptions upon which it rests were well founded. Even if the mechanism were well founded, however, we may not observe the outcome suggested by the mechanism. One important reason for this is that social phenomena are generally the result of several different causal processes operating simultaneously, and these processes can influence and even counteract one another. For this very reason, John Stuart Mill emphasized the importance of treating all theoretical statements as statements about empirical tendencies and not about actualities:

Doubtless, a man often asserts of an entire class what is only true of a part of it; but his error generally consists not in making too wide an assertion, but in making the wrong kind of assertion: he predicated an actual result, when he should only have predicated a tendency to that result – a power acting with a certain intensity in that direction. With regard to exceptions; in any tolerably advanced science there is properly no such thing as an exception. What is thought to be an exception to a principle is always some other...

20 It could be argued that using a deterministic language to describe a stochastic world represents a form of fictionalism. As is discussed on page 62, one should make a distinction between descriptively false and descriptively incomplete sentences, and I would consider the use of a deterministic language to be an example of the latter.
and distinct principle cutting into the former: some other force which impinges against the first force, and deflects it from its direction. There are not a law and an exception to that law — the law acting in ninety-nine cases, and the exception in one. There are two laws, each possibly acting in the whole hundred cases, and bringing about a common effect by their conjunct operation. (Mill 1844: 161–2)

Following Mill, mechanism-based explanations should be viewed as propositions about particular aspects of a causal ‘totality’, with no claim that the tendency in question is the dominant one. For this reason, to quote Mill once again, all mechanisms ‘in consequence of the liability to be counteracted, require to be stated in words affirmative of tendencies only, and not of actual results’ (1874: 319). In more modern language, mechanism-based explanations can be described as propositions about probabilities of different outcomes conditional upon general ceteris paribus clauses (see Gibson 1983).

A concrete example may help to identify the distinguishing features of the three types of explanations. As mentioned above, one of Peter Blau’s concerns was to explain why the formal organizational structures of different organizations vary as they do (see Blau 1970). His strategy was to specify general covering laws that could subsume, and in this respect explain, his concrete observations. From the point of view of the statistical approach, one would not seek to specify general laws, but instead use data about a large number of organizations to find statistically relevant factors that appear to make a difference to the probability of the organizations having a specific formal structure. The mechanism perspective, finally, would explain the change in organizational structures by referring to a constellation of actors and their actions that typically bring about such changes in organizational structures, and would then use statistical and other types of empirical analyses to test the assumptions and predictions of the theory. The details of how such mechanism-based explanations and empirical tests are carried through is the main focus of the rest of this book.

Summary

Let me close this chapter by briefly recapitulating the main thread of the argument. I started by noting that theories, in order to be explanatory, must provide at least partial accounts of why events happen, why

---

21 In practice, of course, the reason that we focus on one tendency and not on another is likely to be based on a belief that the tendency in question is more important than the other. But ‘important’ is not always the same as ‘dominant’.
something changes over time, or why states or events co-vary in time or space.

I then distinguished between three types of explanations that differ in terms of what are considered to be appropriate answers to such why-questions: (1) covering-law explanations, (2) statistical explanations, and (3) mechanism-based explanations. The discussion led to the conclusion that mechanism-based explanations are the most appropriate ones for sociological theory. Statistically oriented variable approaches once were a progressive force that moved sociology forward. But, as Abbott (1999: 216) has noted, this approach now ‘is old and tired’. To move out of the current impasse it seems essential to bring mechanism-based explanations to the fore. A social mechanism is a constellation of entities and activities that are organized in such a way that they regularly bring about a particular type of outcome. We explain an observed social phenomenon by referring to the social mechanism by which such phenomena are regularly brought about, and this entails a focus on the social outcomes that interacting actors are likely to bring about. In developing such theories we must not forsake the high standards set by the statistical tradition. We need to use the most appropriate statistical techniques when testing our theories, and we need to be as precise in formulating our theories as are the best sociologists in the statistical tradition when they specify and diagnose their statistical models.