CHAPTER 3

What Are Causal Mechanisms?

This chapter focuses on debates about the nature of causality and the understanding of causal mechanisms that form the ontological and epistemological underpinnings of all three process-tracing variants. This chapter introduces the reader to the ontological debates within the philosophy of science that deal with the nature of causality itself to understand how the mechanismic understanding of causality used in process-tracing analysis differs from the other understandings of causality that are prevalent in social science, particularly large-*n* statistical analysis and comparative case study research. As this book is not a treatise on the lengthy philosophical debates on causality—a topic that has been the subject of heated exchanges since ancient Greece—the chapter only briefly reviews two key debates about the ontology (nature of) causality that are necessary to grasp how causal mechanisms are understood within theory-centric and case-centric process-tracing variants. The first debate relates to whether we should understand a causal relationship in a skeptical, neo-Humean fashion, where causality is seen purely in terms of patterns of regular association (regularity), or whether causality refers to a deeper connection between a cause and effect (e.g., a mechanism). The second debate deals with whether causal relations should be understood in a deterministic or probabilistic fashion.

The chapter then discusses the nature of causal mechanisms. After defining causal mechanisms in the mechanismic understanding, we identify a common core regarding how causal mechanisms are understood within process-tracing methods and how they differ from an understanding where mechanisms are seen as either empirical events or intervening variables. However, significant differences exist within process-tracing depending on the variant chosen. In case-centric analyses, a mechanism is often considered a loose conglomerate of systematic and nonsystematic parts that together ac-

count for a particular outcome. In contrast, theory-centric analyses operate with relatively simple causal mechanisms that include only systematic parts that can be generalized beyond the confines of the single case.

The chapter concludes with an in-depth discussion of several key points of contention about the nature of causal mechanisms. These points include the ontological debate about whether mechanisms should be understood as operating solely at the micro/actor level or whether macro/structural mechanisms also have a reality of their own as well as more epistemological debates about whether we can directly observe causal mechanisms or whether we can only observe the implications of their existence.

3.1. The Ontology of Causality in the Social Sciences

This section provides a brief overview of the main lines of debate in the philosophy of science regarding the nature of causality itself (mechanisms or regular association) and whether causality should be understood in a probabilistic or deterministic fashion.

Causality as Regular Association versus Causal Mechanisms

When we speak of a causal relationship between X and Y, what is the nature of causality in the relationship? Social science takes two main ontological positions on the nature of causal relations.[1] First, the skeptical, neo-Humean understanding of causality as patterns of regular empirical association has traditionally been the most prevalent in social science (Brady 2008; Kurki 2008). David Hume, in a reaction to the then-prevalent theory that saw causality as a necessary connection in the form of a "hook" or "force" between X and Y, contended that we cannot measure the "secret connection" that links causes and effects. We can observe that an object falls to the ground, but we cannot observe the gravitational forces that caused the object to fall. Given this inability to empirically verify that X caused Y, Hume argued that we should define causes merely in terms of constant conjunction (correlations) between factors; any theorization of "undetectable" mechanisms would quickly, in his opinion, degenerate into metaphysics (Brady 2008; Hume 1975). Causation is therefore taken to mean nothing but the regular association between X and Y, controlled for other relevant possible causes (Chalmers 1999: 214; Marini and Singer 1988).

Causation in the regularity approach is therefore understood in terms of regular patterns of X:Y association, and the actual causal process whereby X produces Y is black-boxed. Regularity can be analyzed by examining patterns of correlation between X and Y. For causality to be established, Hume argued that three criteria for the relationship between X and Y need to be fulfilled: (1) X and Y must be contiguous in space and time; (2) X occurs before Y (temporal succession); and (3) a regular conjunction exists between X and Y (Holland 1986). For example, a regular association between governments that impose austerity measures to cut deficits (X) and their inability to win the subsequent election (Y) would in this understanding suggest the existence of a causal relationship between X and Y, assuming that the three criteria are fulfilled.

The second ontological position in social science is a mechanismic understanding of causality, a position that underlies process-tracing methods (Bennett 2008b). Scientific realists such as Bhaskar (1978), Bunge (1997), and Glennan (1996) have contended that Descartes's mechanismic understanding of causal mechanisms, which was prevalent prior to Hume's treatise, should be reintroduced in a modified fashion. The defining feature of a mechanismic ontology of causation is that we are interested in the theoretical process whereby X produces Y and in particular in the transmission of what can be termed causal forces from X to Y. A mechanismic understanding of causality does not necessarily imply regular association. Indeed, a mechanism can be infrequent. What is necessary is that X actually produces Y through a causal mechanism linking the two (Bogen 2005).

The focus in mechanismic understandings of causality is the dynamic, interactive influence of causes on outcomes and in particular how causal forces are transmitted through the series of interlocking parts of a causal mechanism to contribute to producing an outcome. Philosopher Stuart Glennan, for example, defines a mechanism as "a complex system, which produces an outcome by the interaction of a number of parts" (1996: 52; Glennan 2002). Within social science research, Andrew Bennett has defined causal mechanisms as "processes through which agents with causal capacities operate in specific contexts to transfer energy, information or matter to other entities (2008b: 207). David Waldner has defined a mechanism as "an agent or entity that has the capacity to alter its environment because it possesses an invariant property that, in specific contexts, transmits either a physical force or information that influences the behavior of other agents or entities" (2012: 18).

In contrast to the regularity understanding, causality is therefore un-

derstood in more complex terms as the causal mechanism linking X to Y, depicted as X → mechanism → Y. By studying mechanisms, scholars gain what Salmon (1998) terms deeper explanatory knowledge.

Probabilistic versus Deterministic Understandings of Causality

Another key ontological distinction is between probabilistic and deterministic understandings of causality. Probabilistic causality means that the researcher believes that we are dealing with a world in which there are random (stochastic) properties, often modeled using error terms (see, e.g., King, Keohane, and Verba 1994: 89 n. 11). This randomness can be the product either of an inherent randomness, where the social world is understood in terms analogous to quantum mechanics, or of complexity. In the latter case, while we might assume that the world is inherently deterministic, the social world contains nonlinear associations, feedback, and other complex features that make it appear as if there are stochastic elements. At the end of the day, whether the stochastic elements of the social world are inherent or the product of complexity is irrelevant, as their implications for probabilistic theorization are the same (Marini and Singer 1988).

Probabilistic theories therefore assume that there are both systematic and nonsystematic (i.e., random) features of reality. For example, in the study of genetics, many scholars contend that a large portion of any individual's cognitive abilities is inherited from his/her parents (e.g., Bouchard 2004; Haworth et al. 2010; Herrnstein and Murray 1994). However, geneticists do not expect that children will always have the same IQ as their parents; instead, heredity has an inherent degree of randomness that results in a relationship best expressed as a probability distribution in the form of a bell curve, where on average higher-IQ parents have higher-IQ children, and vice versa. This causal relationship is understood as probabilistic, where even precise knowledge of the parents' IQ would not enable us to make exact predictions of the IQ of any individual child, only a probable range of outcomes. Hypotheses in probabilistic causal models therefore take the form of "Y tends to increase when X increases."

A probabilistic ontology has methodological implications in that it only makes sense to investigate probabilistic causal relationships with cross-case methods, investigating mean causal effects of systematic parts across the population or a sample of the population of the phenomenon. A single case study comparing the IQ of a child and her parents tells us nothing about the strength of the relationship between IQ and heredity in the population,

Process-Tracing Methods : Foundations and Guidelines

as we have no way of knowing whether the correlation in the single case is the product of pure chance or is close to the predicted mean causal effects of heredity.

For statisticians, the term *deterministic causality* means a theoretical model where there is no error term (i.e., no random component), which basically means that, if properly specified, a deterministic model should explain 100 percent of the variance of a given dependent variable. Qualitative social scientists have a more pragmatic understanding of determinism. Mahoney succinctly summarizes this understanding: "The assumption of an ontologically deterministic world in no way implies that researchers will successfully analyze causal processes in this world. But it does mean that randomness and chance appear only because of limitations in theories, models, measurement and data. The only alternative to ontological determinism is to assume that, at least in part, 'things just happen'; that is, to assume truly stochastic factors . . . randomly produce outcomes" (2008: 420).

Deterministic causal relationships can be studied at the population level but are more often associated with small-*n* case study research (Mahoney 2008). For qualitative scholars, the term *deterministic* is used primarily to refer to discussions of necessary and sufficient causes in individual cases or combinations of these types of conditions (417). This means that what we are examining is not whether a given X tends to covary with Y in a population but whether X is either a necessary and/or sufficient cause of Y in an individual case (Collier, Brady, and Seawright 2010a: 145; Mahoney 2008: 417). A condition is necessary if the absence of it prevents an outcome, regardless of the values of other variables, whereas if a sufficient condition is present, the outcome will always take place.[2]

The Ontologies of Causality Adopted in Different Social Science Methods

Table 3.1 illustrates the four different logical combinations and the social science methods that have utilized them. Cell 1 illustrates the most widely used ontological position in social science methods, where regularity is coupled with a probabilistic understanding. X is theorized to increase the probability that outcome Y occurs in a population, and if we find that the three criteria for assuming causality are fulfilled (contiguity, temporal succession, and regular association), we can infer that X is a cause of Y. Methods that utilize this position include large-*n*, quantitative statistics and the adaption of them by King, Keohane, and Verba (KKV) to qualitative case study research (1994).

In contrast to KKV's pronouncements regarding probabilistic causal-

Process-Tracing Methods : Foundations and Guidelines

ity and case studies, most qualitative methodologists counter that this is unfaithful to the tenets of qualitative case-oriented methodology and that we instead should adopt deterministic understandings of causality (Blatter and Blume 2008; Mahoney 2008). Mahoney (2008) argues that it makes no sense to use a probabilistic understanding of causality when we are investigating single cases and their causes: "At the individual case level, the ex post (objective) probability of a specific outcome occurring is either 1 or 0; that is, either the outcome will occur or it will not. . . . [S]ingle-case probabilities are meaningless" (415–16). This makes cell 3 a logical impossibility in a single-case research design, whereas studying mechanisms is not feasible for larger-*n* studies (see chapter 5).

Instead, Mahoney (2008) contends that we should utilize a deterministic understanding in small-*n* case research. In cross-case studies, this can involve comparative case study methods (cell 2) where patterns of regular association between conditions are investigated. Congruence methods are within-case studies where the similarities between the relative strength and duration of the hypothesized causes and observed effects are assessed (George and Bennett 2005: 181–204; Blatter and Haverland 2012).

Process-tracing involves studying causal mechanisms in single-case studies. The basic point is that if we take mechanisms seriously, this implies that evidence from individual process-tracing studies cannot be compared with that gathered in other studies given that the evidence is case-specific. What is relevant evidence in one case cannot be meaningfully compared with evidence in another case, making cross-case comparisons more or less impossible. Therefore, if we are interested in studying causal mechanisms, we need to adopt the deterministic ontology of causality (Mahoney 2008). This is depicted as cell 4 in table 3.1.

TABLE 3.1. The Ontological Assumptions regarding Causality of Different Social Science Methodologies

	Probabilistic	Deterministic
Regularity	(1) Large-*n* quantitative statistical methods, KKV's qualitative case study methods	(2) Congruence case studies (within case), comparative cross-case study methods (small-*n*), and qualitative comparative analysis (QCA) (medium-*n*)
Mechanisms	(3) Not logically possible in single case studies, not feasible to examine mechanisms in larger-*n* study	(4) Process-tracing methods (single case)

3.2. Causal Mechanism—A Mechanismic Definition

In the mechanismic understanding introduced earlier, a causal mechanism is defined as a theory of a system of interlocking parts that transmits causal forces from X to Y (Bhaskar 1979; Bunge 1997, 2004; Glennan 1996, 2002). Bunge defines a theoretical causal mechanism as "a process in a concrete system, such that it is capable of bringing about or preventing some change in the system as a whole or in some of its subsystems" (1997: 414). Another good definition that summarizes this position is that "a mechanism is a set of interacting parts—an assembly of elements producing an effect not inherent in any one of them. A mechanism is not so much about 'nuts and bolts' as about 'cogs and wheels'—the wheelwork or agency by which an effect is produced" (Hernes 1998: 78).

The mechanism linking a cause and outcome can be understood using a machine analogy. Each part of the theoretical mechanism can be thought of as a toothed wheel that transmits the dynamic causal energy of the causal mechanism to the next toothed wheel, ultimately contributing to producing outcome Y. We use the machine analogy merely as a heuristic aid in the conceptualization and operationalization of a given causal mechanism. We by no means imply that all social causal mechanisms exhibit machine-like qualities; indeed, many social causal mechanisms are more dynamic (Bunge 1997; Pierson 2004). They are not necessarily neutral transmission belts. A small trigger can have a disproportional effect, as the forces are amplified through a causal mechanism. A strong cause also can have its effect muted through a causal mechanism. Moreover, the transmission of causal forces can be nonlinear through a mechanism, or the workings of the mechanisms can result in the alteration of the causal forces in another direction. This implies that causal mechanisms can have effects that cannot merely be reduced to the effect of X, making it vital to study causal mechanisms together with causes instead of causes by themselves.

Each of the parts of the causal mechanism can be conceptualized as composed of entities that undertake activities (Machamer 2004; Machamer, Darden, and Craver 2000). Entities are the factors engaging in activities (the parts of the mechanism—i.e., toothed wheels), where the activities are the producers of change, or what transmits causal forces through a mechanism (the movement of the wheels).

What is the logical relationship between the parts of a mechanism and the whole in the mechanismic understanding? We adapt the terminology of necessary and sufficient conditions to this relationship. In comparative methods, explanatory conditions are viewed as necessary, sufficient, or some

combination of the two, such as being INUS conditions (Braumoeller and Goertz 2000; Mahoney 2000; Ragin 1988). Mackie defines INUS conditions as an insufficient but necessary part of an unnecessary but sufficient condition (1965). Necessary conditions are conditions that have to be present for an outcome to occur and where the absence of X results in an absence of the outcome. In contrast, sufficiency describes a situation where a condition (or set of conditions) is able to produce an outcome. If X, then always outcome Y.

While the cross-case comparative literature usually describes causes (Xs) as conditions, there is no logical reason why we cannot adapt the logic of necessary and sufficient conditions to the present purposes of analyzing the mechanisms that produce an outcome. The difference is that while a comparativist thinks of a condition as X, we argue that we can also adapt the language to the analysis of the parts of a mechanism and the whole.

Each part of a mechanism can be illustrated as $(n_n \rightarrow)$, where the n_n refers to the entity (n) and the arrow to the activity transmitting causal energy through the mechanism to produce an outcome. * is used to refer to logical *and*. As a whole, a causal mechanism can therefore be portrayed as

$$X \rightarrow [(n_1 \rightarrow) * (n_2 \rightarrow)] Y$$

This should be read as X transmits causal forces through the mechanism composed of part 1 (entity 1 and an activity) and part 2 (entity 2 and an activity) that together contribute to producing outcome Y. This is a "context-free" mechanism, and a proper study would also detail the contextual conditions that enable the mechanism to become activated (Falleti and Lynch 2009).

An analogy can be made to a car, where X could be the motor and Y is the movement of the car. However, without a driveshaft and wheels, the motor by itself cannot produce forward movement. Here the driveshaft and wheels can be thought of as the causal mechanism that transmits forces from X (motor) to produce Y (movement).

We contend that all three variants of process-tracing research strategies share an understanding of the parts of a causal mechanism, where they should be conceptualized as insufficient but necessary parts of an overall mechanism. Each part of the mechanism is by itself insufficient to produce an outcome Y, as it only functions together with the rest of the "machine." Second, explicit in a mechanismic ontology is a view that the parts that we include in our conceptualization of a given causal mechanism are absolutely

Process-Tracing Methods : Foundations and Guidelines

vital (necessary) for the "machine" to work, and in the absence of one part, the mechanism itself cannot be said to exist. This is denoted by the logical *and*. If we have conceptualized a three-part causal mechanism as $(n_1 \rightarrow) * (n_2 \rightarrow) * (n_3 \rightarrow) = Y$, and $(n_2 \rightarrow)$ is either empirically or theoretically superfluous, then the mechanism should be reconceptualized as $(n_1 \rightarrow) * (n_3 \rightarrow) = Y$. This introduces a disciplining effect when we attempt to model a given theory as a causal mechanism. Basically, if a logical, theory-based argument cannot be formulated for why the particular part is a vital (necessary) part of a causal mechanism and in particular describes how the specific entity or entities engage in activities that transmit causal forces from X to Y, then the part should be eliminated from the theoretical model as being redundant. In addition, if our empirical analysis has found that a part is not necessary, then the mechanism should also be reconceptualized to exclude it.

This disciplining effect means that when we engage in the theoretical modeling of mechanisms, we do not suffer from the problem of infinite digression into a plethora of parts at an ever more microlevel of explanation (for a particularly good discussion of this problem, see Roberts 1996); instead, we model only the parts of a mechanism that are theorized as absolutely essential (necessary) to produce a given outcome. This approach also helps us remedy some of the problems of adding various ad hoc explanations for why a theory or hypothesis might eventually hold in a specific case when we are engaging in theory-centric process-tracing, as we need to be able to argue that all of the parts of the mechanism can hypothetically exist in other cases.

The understanding of the parts of a causal mechanism as individually necessary requires that a deterministic ontology of causality is adopted, enabling causal inferences about the existence of the individual parts of a causal mechanism. Using a probabilistic ontology, if we empirically investigate whether a specific part of a causal mechanism exists in a case and find no confirming evidence, we are left in the dark as to whether we should disconfirm its existence or merely ascribe it to the randomness of any individual case in circumstances where there is otherwise a strong mean causal effect across the population. The methodological prescription here would be to increase the number of cases (increase the *n*), as KKV suggest, but doing so is inappropriate for process-tracing given that the aim is to test the presence/absence of mechanisms within single-case studies.

In contrast, in process-tracing methods, each part of a mechanism is conceptualized as an individually necessary element of a whole. Yet although these examples are single-case studies, we contend that we can make infer-

ences about the presence of the parts of a mechanism using the Bayesian logic of inference (see chapter 5).

Further, if we find strong disconfirming evidence for one part of a multipart mechanism, we disconfirm the existence of the whole hypothesized causal mechanism. In this situation, there are two ways forward: One can either completely discard the hypothesized mechanism or engage in theoretical revision using more inductive tools in an attempt to detect an underlying mechanism (theory-building). If we find strong confirming evidence of the existence of each part of the mechanism, we can infer that the mechanism actually exists (with a certain degree of confidence).

3.3. Debates about the Nature of Causal Mechanisms

While it is relatively easy to differentiate the mechanismic and deterministic understanding of causality used in process-tracing from the regularity and probability understandings that underlie methods such as large-*n* statistical analysis, a considerable degree of ambiguity exists about what causal mechanisms actually are (Gerring 2010; Mahoney 2001).

Many social science scholars contend that they are studying causal mechanisms with a variety of research methods: Some scholars believe that causal mechanisms should be understood as systems that transmit causal forces from X to Y, while others see them as series of empirical events between the occurrence of X and Y. Still others have considered causal mechanisms in terms of intervening variables between X and Y. We illustrate how taking mechanisms seriously implies instead a certain understanding of mechanisms as systems whereby X contributes to producing an outcome.

Within process-tracing methods, disagreements have arisen about the exact nature of these mechanismic systems. As introduced in chapter 2, there is disagreement across the theory/case-centric divide regarding whether they should be seen as relatively parsimonious and singular or as case-specific conglomerates. Additional disagreements that do not span this divide include whether causal mechanisms operate solely at the micro/actor level or whether there are also macro/structural-level mechanisms that cannot be reduced to the microlevel? Finally, can we observe causal mechanisms in action or only indirectly observe the implications of their existence? On these two points, we suggest that there are no logically imperative reasons for choosing one or the other position, suggesting that we should remain agnostic about levels or whether mechanisms are observable.

Causal Mechanisms Are More Than Just Empirical Events

Many scholars contend that they are studying causal mechanisms but are in reality only tracing an empirical process, understood as a series of empirical events that are temporally and spatially located between the occurrence of X and the outcome Y (an empirical narrative).

Figure 3.1 depicts how some scholars misuse the term *process-tracing* to refer to scholarship that traces an empirical process (sequence of events) occurring between X and Y but where the causal mechanism linking them is in effect black-boxed (Bunge 1997). This type of research takes the form of an empirical narrative: "Actor A did X to actor B, who then changed his position on issue Y, and so forth." This type of scholarship is a valuable form of descriptive inference that describes a series of empirical events and provides valuable historical knowledge regarding what happened but tells us little about the underlying how and why an outcome occurred. The focus is on events instead of a theory-guided analysis of whether evidence suggests that a hypothesized causal mechanism was present.

In process-tracing, the research focuses on the causal mechanism through which X contributes to producing an outcome Y. This is most evident in what we termed the theory-centric variants of process-tracing (figure 3.2), where a relatively simple causal mechanism is front and center in the analysis. In the theory-testing variant, the mechanism is explicitly theorized along with the empirical manifestations of each part of the mechanism. The case study then assesses whether we find the predicted empirical evidence. For each part of the hypothesized causal mechanism, we investigate whether the predicted empirical manifestations of the mechanism were present or absent. Different types of evidence are gathered, depending on what is best suited to enable us to update our confidence in the presence/absence of the mechanism.

For example, a theorized rational decision-making mechanism could be theorized as having the four parts: decision makers would (1) gather all relevant information; (2) identify all possible courses of action, (3) assess the alternatives based on the decision makers' utility function, and finally (4) choose the alternative that maximizes the expected utility (Oneal 1988). The empirical manifestations of these four parts would be very different, and different types of evidence would be used to assess whether part 1 or part 4 was present. Determining whether part 1 took place before part 2 could be a manifestation of the presence of part 1 measured by investigating the temporal sequence of events. However, the sequence of events would not be enough to establish whether parts 1 or 2 were present. Instead, for example,

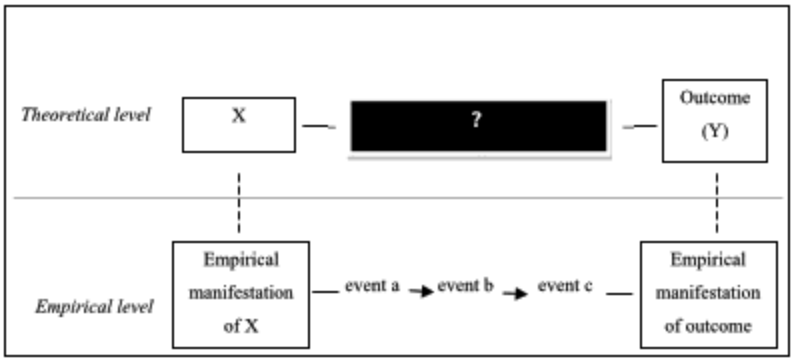

Fig. 3.1. Empirical narratives are the black-boxing of mechanisms.

to assess part 2, we would test whether other predicted manifestations were also present, including evidence that showed whether or not all possible courses of action were assessed. Although we are testing the empirical manifestations of each part of the mechanism, we are in effect tracing the underlying theoretical causal mechanism, illustrated in figure 3.2. Our analysis is structured as focused empirical tests of each part of a mechanism instead of a narrative empirical presentation of the story of events of the case. Theory is guiding our process-tracing analysis irrespective of whether or not we have a theory-centric or case-centric ambition. In contrast, by tracing events, we gain no knowledge of the underlying causal mechanism.

Complicating the picture slightly is the difference in how causal mechanisms are understood in theory-centric and case-centric process-tracing variants. In the theory-centric variant, mechanisms are understood as mid-range theories of mechanisms that transmit causal forces from X to Y and are expected to be present in a population of cases, assuming that the context that allows them to operate is present. Analysis seeks to determine whether a single mechanism such as learning or policy drift is present in a particular case, but given that most social outcomes are the product of multiple mechanisms, no claims of sufficiency are made. This means that we are studying singular mechanisms instead of complex conglomerates.

In contrast, when the purpose of analysis is to craft a sufficient explanation of a particular outcome, we almost always need to combine mechanisms into an eclectic conglomerate mechanism to account for a particular outcome. Evans, for example, writes, "Cases are always too complicated to vindicate a single theory, so scholars who work in this tradition are likely

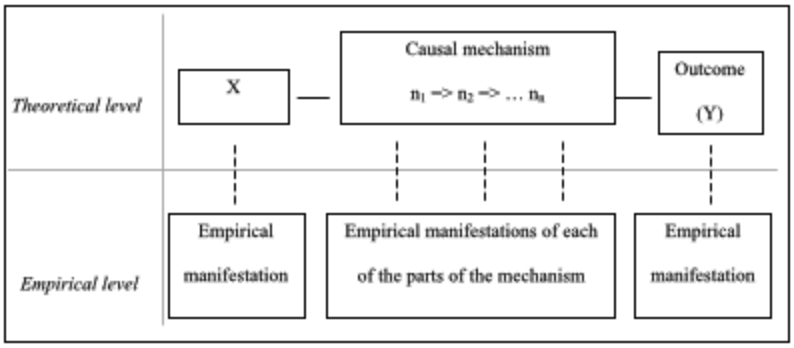

Fig. 3.2. Theory-centric process-tracing methods

to draw on a mélange of theoretical traditions in hopes of gaining greater purchase on the cases they care about" (1995: 4). Schimmelfenning notes that "eclecticism is the unintended result of research that seeks to explain specific events as well as possible" (cited in Sil and Katzenstein 2010: 191). The result is more complicated, case-specific combinations of mechanisms (see chapter 4).

Further, given that the ambition is to craft a minimally sufficient explanation of a particular outcome, it is usually necessary to include nonsystematic parts in the causal mechanism, defined as a mechanism that is case-specific. While Elster contends that mechanisms have to be at a level of generality that transcends the particular spatiotemporal context (i.e., they are systematic mechanisms) (1998: 45), thereby excluding the use of nonsystematic mechanisms in explaining outcomes, other scholars have more pragmatically argued that mechanisms that are unique to a particular time and place also can be defined as mechanisms. Wight, for example, has defined mechanisms as the "sequence of events and processes (the causal complex) that lead to the event" (2004: 290). Nonsystematic mechanisms can be distinguished from systematic ones by asking whether we should expect the mechanism to play any role in other cases.

The importance of nonsystematic mechanisms in explaining a particular outcome makes explaining-outcome process-tracing sometimes more analogous to the historical interpretation of events (Roberts 1996). However, these nonsystematic parts will almost never stand alone, given that social reality is not just a random hodgepodge of events but includes mechanisms that operate more generally across a range of cases within a bounded population.

Further, the inclusion of nonsystematic mechanisms that are sometimes depicted as events has an important advantage in that it enables us to capture actor choice and the contingency that pervades historical events, immunizing our research from the criticisms of social science from historical scholars (Gaddis 1992–93; Roberts 1996; Rueschemeyer 2003; Schroeder 1994). In the words of Lebow, "Underlying causes, no matter how numerous or deep-seated, do not make an event inevitable. Their consequences may depend on fortuitous coincidences in timing and on the presence of catalysts that are independent of any of the underlying causes" (2000–2001: 591–92).

The admission of case-specific mechanisms does not mean that they are preferable (Gerring 2006). "To clarify, single-outcome research designs are open to case-centric explanation in a way that case study research is not. But single-outcome researchers should not assume, *ex ante,* that the truth about their case is contained in factors that are specific to that case" (717). What differentiates explaining-outcome process-tracing from historical research is both the causal-explanatory focus—where the analysis is theory-guided—and the ambition to go beyond the single case (Gerring 2006; Hall 2003). With regard to the ambition to go beyond the single case, this involves attempts to identify what mechanisms are systematic and nonsystematic in the specific case study. This is best seen in book-length works, where lessons for other cases are developed in the conclusions. For example, what parts of the conglomerate mechanism do we believe can be systematic based on the findings of our study and in light of what we know from other research? What findings can be exported to other cases, and to what extent are they unique to a particular case? Individual causal mechanisms can be exported, but the case-specific conglomerate (usually) cannot be exported.

Figure 3.3 illustrates the more complex nature of the causal mechanisms that are being traced in explaining-outcome process-tracing. The analysis still focuses on the theoretical level of causal mechanisms, although these are understood in a broader and more pragmatic fashion.

Causal Mechanisms Are More Than Just Intervening Variables

Another common misunderstanding about mechanisms is made by those who conceptualize mechanisms as a series of intervening variables. The most widely used definition of causal mechanism sees them as a series of intervening variables through which an explanatory variable exerts a causal effect on an outcome variable (e.g., Falleti and Lynch 2009: 1146; George and Bennett 2005: 6; Gerring 2007b, 2010; Hedström and Ylikoski 2010; King, Keohane,

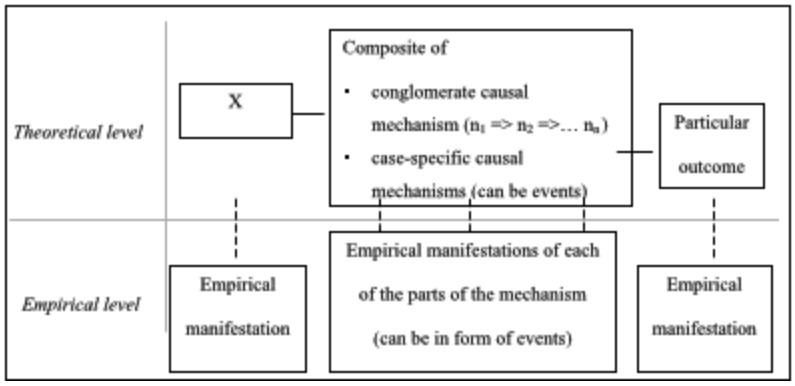

Fig. 3.3. Mechanisms in explaining-outcome process-tracing

and Verba 1994; Waldner 2011). This understanding is best exemplified in the definition of causal mechanisms given in KKV's book, *Designing Social Inquiry*: "This definition would also require us to identify a series of causal linkages, to define causality for each pair of consecutive variables in the sequence, and to identify the linkages between any two of these variables and the connections between each pair of variables" (1994: 86). For KKV, mechanisms are simply chains of intervening variables that connect the original posited cause and the effect on Y (87). These intervening variables are usually expressed as nouns.

The causal linkages are viewed as variables in this definition. This means that the values they can take vary and that they have an existence independent of each other, as each variable is in effect a self-contained analytical unit. Variance implies that a probabilistic understanding of causality is utilized—something that makes little sense when we are engaging in single-case studies.

Second, the use of intervening variables usually has the practical consequence that the linkages between the variables are neglected. The neglect of the causal linkages between variables results from the simple fact that when a causal mechanism is conceptualized as being composed of a series of intervening variables, it is far easier to measure the presence/absence of an intervening variable than the linkages between them. The analytical focus on variables instead of linkages is strengthened by the regularity understanding of causality that is used by King, Keohane, and Verba (1994), among others.[3]

The result is that the intervening variable understanding ends up gray-

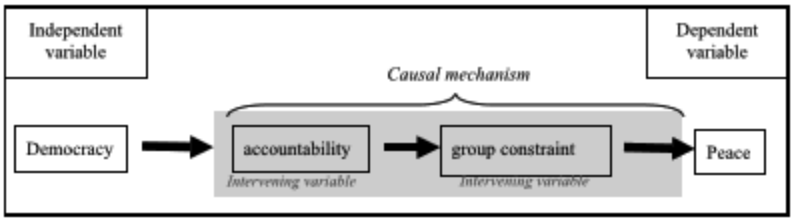

Fig. 3.4. Causal mechanism defined as a series of intervening variables. (Drawn from Rosato 2003.)

boxing the causal mechanism itself (Bunge 1997; Mahoney 2001; Waldner, 2012). While the mechanism is slightly unpacked, the actual transmission of causal forces from X that produce Y is not explicitly studied (Bunge 1997). Waldner goes as far to say that "mechanisms explain the relationship between variables because they are not variables" (2012: 18). Furthermore, causal mechanisms themselves can affect how causal forces are transmitted from X to Y—for example, by magnifying the effects of X through a mechanism. Mechanisms are therefore more than just sets of intervening variables.

An example of the prevalent intervening variable understanding of causal mechanisms can be seen in an article by Rosato (2003) that critically examines democratic peace theory. He contends that he is conceptualizing the mechanisms linking democracy and peace but then describes one mechanism as being composed of two intervening variables, accountability and group constraints. Accountability means that political elites will be voted out of office if they adopt unpopular policies. In the model, leaders are theorized to be especially responsive to the wishes of domestic antiwar groups, creating a constraint emerging from groups that affects the ability of leaders to engage in war (585). This theory is depicted in figure 3.4.

However, the actual causal forces that are transmitted through a mechanism to produce the outcome (peace) are left out of the analysis; they are, in effect, gray-boxed. Both accountability and group constraints are theorized to be linked with peace, but the conceptualization prevents us from analyzing *how* democracy produces peace, as the causal linkages between the intervening variables are not explicitly theorized. Instead, the analyst would measure only the presence/absence of the intervening variables as well as potentially the covariance between each intervening variable and the outcome (Y). If strong evidence exists, the analyst can first infer that the intervening variable is present. If each of the intervening variables was found to be present, a further leap can be made by inferring that the theorized mechanism

was present. But it is important to underline that we are not studying how democracy produced peace but only measuring a series of covariants.

We agree with social scientists such as Bennett (2008a) and Waldner (2012) who contend that we should take seriously the distinct ontological nature of the mechanismic understanding of causality in process-tracing. This understanding has important methodological benefits, as conceptualizing mechanisms as systems results in an empirical analysis that has a more explicit focus on the causal linkages between X and Y that produce the outcome, enabling stronger within-case inferences to be made.

The mechanismic understanding involves opening up the black box of causality as much as possible (Bennett 2008a; Bhaskar 1978; Bunge 1997; Glennan 1996). Mechanisms are more than just a series of intervening variables. Instead, mechanisms are invariant with regard to both the whole mechanism and each individual part. Either all of the parts of a mechanism are present, or the mechanism itself is not present (Glennan 2005: 446).

The difference between a theoretical conceptualization of a mechanism as a system and one composed of intervening variables is illustrated in figure 3.5, where the same mechanism described in the example of the democratic peace thesis is conceptualized in terms of a system that produces an outcome through the interaction of a series of parts of the mechanism. Each part is composed of entities that engage in activities. In the example, instead of conceptualizing the mechanism as two intervening variables (accountability and group constraints), the mechanism is unpacked further by focusing on the two entities (liberal groups and governments) and their activities ("agitate" and "respond") that make up the two parts of this simple mechanism. Each of the entities can be thought of as a wheel, where the activities are the movement that transmits causal forces to the next part of the mechanism.

Part 1 of the mechanism is liberal groups agitating against war before the government, and part 2 is the government responding to this agitation by adopting conciliatory foreign policies that result in peaceful relations. Together, the two parts comprise the simple theorized causal mechanism. In figure 3.5, we see that the entities that undertake the activities transmit causal forces through the mechanism are first "liberal groups" and then "governments." A subsequent process-tracing study of this mechanism would involve testing whether the hypothesized parts of the mechanism were present in a given appropriately chosen case.

By explicitly conceptualizing the activities that produce change, the mechanismic approach to causal mechanisms draws our attention to the actions and activities that transmit causal forces from X to Y—that is, how the mechanism produces an outcome and the context within which the

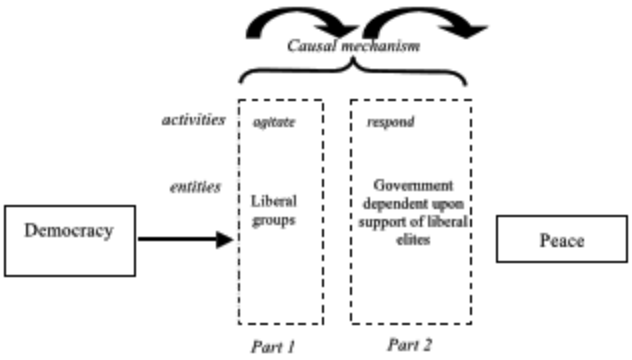

Fig. 3.5. A democratic peace example of a causal mechanism

mechanism functions. If we then can confirm the existence of a hypothesized causal mechanism in a theory test, we have produced strong evidence that shows how the theorized parts of the causal mechanism produce Y and how X and Y are causally connected by the mechanism (Bunge 1997, 2004). Understanding mechanisms in these terms enables us to capture the process whereby causal forces are transmitted through a causal mechanism to produce an outcome, forces that are black-boxed or gray-boxed when they are understood as events or intervening variables.

Are Mechanisms Only at the Micro/Actor Level?

Philosophers of social science have engaged in considerable debate about whether mechanisms always have to be reducible to the microlevel (Hedström and Swedberg 1998) or whether there are also macrolevel mechanisms that cannot be reduced to the microlevel (Bunge 2004; Mayntz 2004; McAdam, Tarrow, and Tilly 2001). Should we reduce every causal mechanism to the microlevel, investigating the actions of individuals, or are there causal mechanisms that have macrolevel properties?

To introduce the level of causal mechanism debate, Hedström and Swedberg have a helpful figure that illustrates how the level debate relates to the study of causal mechanisms (1998: 22) (figure 3.6). However, they take an extreme position that there are no purely macrolevel mechanisms. George and Bennett take a similar position when they state that mechanisms are "processes through which agents with causal capacities operate" (2005: 137).

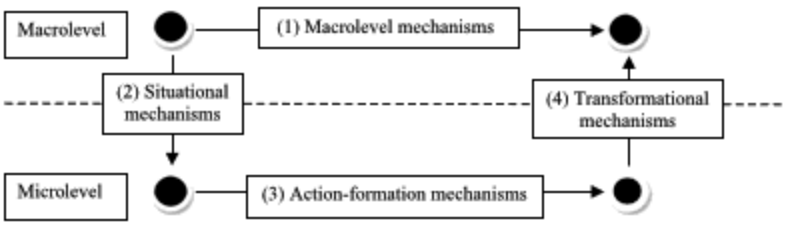

Fig. 3.6. Levels of causal mechanisms. (Based on Hedström and Swedberg 1998: 22.)

They go on to define mechanisms as the microfoundations of a causal relationship that involve the "irreducibly smallest link between one entity and another" (142).

In our view, this viewpoint unnecessarily restricts the uses of process-tracing methods to purely microlevel examination, usually at the level of individual actors and their behavior in specific decision-making processes. Yet many of the most interesting social phenomena we want to study, such as democratization, cannot meaningfully be reduced solely to the actor level but in certain situations can be better analyzed empirically at the macrolevel (McAdam, Tarrow, and Tilly 2001). Given that this conundrum is in essence the classic debate between agent and structure, we argue for an agnostic and pragmatic middle-ground position, where the choice of level that is theorized is related to the level at which the implications of the existence of a theorized causal mechanism are best studied. Mechanism may occur or operate at different levels of analysis, and we should not see one level as more fundamental than another (Falleti and Lynch 2009: 1149; George and Bennett 2005: 142–44; Mahoney and Rueschemeyer 2003: 5; McAdam, Tarrow, and Tilly 2001: 25–26).

Macrolevel mechanisms are structural theories that cannot be reduced to the actions of individuals (Type 1). Many sociologists claim that the search for the microlevel foundations of behavior is futile and that much of the capacity of human agents derives from their position in society (structure) (Mahoney 2001; McAdam, Tarrow, and Tilly 2001; Wight 2004). Many of the important causal mechanisms in the social world are arguably macrolevel phenomena that are "collaboratively created by individuals yet are not reducible to individual" action (Sawyer 2004: 266). Sawyer terms this the concept of "emergence," which means that macrolevel mechanisms have their own existence and have properties that cannot be reduced to the microlevel.

Institutional roles, norms, and relational structures can play a significant role for actor behavior, and mechanisms can have structural properties that cannot be defined solely by reference to the atomistic attributes of individual agents (266). For example, system-level theories in international relations include neorealism, where Waltz (1979) theorizes the balancing mechanism as solely the product of macrolevel factors.

Three different types of mechanisms are related to the microlevel actions of agents, and two of them combine macrolevel properties with microlevel actions. At the microlevel are action-formation mechanisms (Type 3), or what Hedström and Ylikoski term "structural individualism," where all social facts, their structure and change, are in principle explicable in terms of individuals, their properties, actions, and relations to one another (2010: 59). Purely microlevel theories relate to how individuals' interests and beliefs affect their actions and how individuals interact with each other (Type 3). However, this does not mean that actors are necessarily individual humans. Social science operates with many forms of collective actors that are treated as if they were individuals, most bravely captured by Wendt's contention that "states are people too" (1999: 194). One example of a purely microlevel theory is Coleman's (1990) rational-choice-based theory of social action, where even actions such as altruism are reduced solely to individual self-interested motivations (desire for reciprocity in a long-term iterated game).

Situational mechanisms link the macro- to the microlevel (Type 2). Situational mechanisms describe how social structures constrain individuals' action and how cultural environments shape individuals' desires and beliefs (Hedström and Swedberg 1998). Examples of a macro-micro-level mechanisms include constructivist theories of actor compliance with norms that are embedded at the macrolevel (structural).

Transformational mechanisms describe processes whereby individuals, through their actions and interactions, generate various intended and unintended social outcomes at the macrolevel (Type 4) (Hedström and Swedberg 1998). An example of this type of micro-macro-level mechanism could be socialization processes, whereby actors through their interaction create new norms at the macrolevel. Another example is from game theory, where individual actions in situations like the prisoner's dilemma create macrolevel phenomena such as the tragedy of the commons.

There is no single correct answer to the question of at which level a causal mechanism should be theorized. Here we argue for a pragmatic approach. There are social mechanisms whose observable implications can best be theorized and measured at the macrolevel. Therefore, we agree with Stinchcombe's conclusions that

> Where there is rich information on variations at the collective or structural level, while individual-level reasoning (a) has no substantial independent empirical support and (b) adds no new predictions at the structural level that can be independently verified, theorizing at the level of [individual-level] mechanisms is a waste of time. (1991: 380)

Further, by disaggregating macrolevel mechanisms from their microlevel components, we risk the problem of infinite regress. Kincaid puts it well when he states that if we want to study the mechanism linking two macrolevel factors, "Do we need it at the small-group level or the individual level? If the latter, why stop there? We can, for example, always ask what mechanism brings about individual behavior. So we are off to find neurological mechanisms, then biochemical, and so on" (1996: 179). If we go down this path, the result is that no causal claims can be established without absurd amounts of information (Steel 2004).

The pragmatic middle ground advocated in this book states that mechanisms can hypothetically exist at both the macro- and microlevels, along with mechanisms spanning the two levels (situational and transformative mechanisms). The choice of level for our theorization of causal mechanisms in process-tracing depends on the pragmatic concern of the level at which the empirical manifestations of a given causal mechanism are best studied. If the strongest tests of a given mechanism are possible at the macrolevel, then it should be theorized and studied empirically at this level, whereas if the empirical manifestations are better observed at the microlevel, then we should conceptualize and operationalize our study at this level.

Can We Observe Causal Mechanisms?

Can we directly measure mechanisms or only infer their existence through their observable implications? Many scholars hold the view that causal mechanisms are unobservable. For example, George and Bennett posit that mechanisms are "ultimately unobservable physical, social, or psychological processes through which agents with causal capacities operate" (2005: 137). Hedström and Swedberg (1998) argue that causal mechanisms are merely analytical constructs that do not have a real-world existence.

In contrast, other scholars contend that the parts of a mechanism should be understood as having a "kind of robustness and reality apart from their place within that mechanism" (Glennan 1996: 53). In the words of Bunge,

Process-Tracing Methods : Foundations and Guidelines

"Mechanisms are not pieces of reasoning but pieces of the furniture of the real world" (1997: 414). Reskin (2003) suggests that we can answer the question of how an outcome was produced only by investigating observable causal mechanisms, thereby excluding many cognitive and macrolevel mechanisms.

As with the question of the level of analysis of mechanisms, a full answer to this question would involve a lengthy philosophical discussion that is outside the scope of this book. Our position here is pragmatic: We agree with scientific realist scholars such as Bunge and Glennan that our ambition should be to attempt to get as close as possible to measuring the underlying causal mechanism but that this ideal may not be achievable for theoretical and empirical reasons.

Some types of causal mechanisms can be conceptualized and operationalized in a manner that permits quite close observation of actual mechanisms and where plentiful evidence exists that enables us to measure the mechanism quite closely. For example, Owen (1994, 1997) conceptualizes and operationalizes a democratic peace mechanism that results in empirical tests that come quite close to measuring the mechanism directly (see chapter 5). Other types of causal mechanisms, such as groupthink mechanisms that deal with conformity pressures in small-group decision making, are so complex and involve such difficult measurement issues relating to access to confidential documents and problems relating to measurement of sociopsychological factors that we can measure the mechanism only in an indirect fashion through proxies (indicators) of the observable implications (see Janis 1983).

What are the methodological implications of the choice between adopting an understanding of causal mechanisms as observable or unobservable? If we believe that mechanisms can be observed quite directly, when we operationalize a mechanism, we are aiming to examine the fingerprints that the mechanism should have left in the empirical record. In contrast, if we believe that mechanisms are ultimately unobservable, we should instead think in terms of the observable implications that a mechanism should leave. In practice, the two positions result in similar forms of operationalization (see chapter 6).

The next chapter turns to the more practical question of working with theories of causal mechanisms.