CHAPTER 4

Working with Theories of Causal Mechanisms

Theoretical concepts are not self-explanatory. Nor are causal mechanisms self-evident from causal theorization of relationships between X and Y. Instead, causal theories need to be transformed so they offer a clear hypothesized mechanism describing how a type of outcome is produced. Therefore, theoretical concepts and causal mechanisms need to be carefully defined before they can be employed in process-tracing analysis. While there are numerous methodological texts relating to conceptualization techniques in political science, existing guidelines are not always applicable when we are using process-tracing methods.[1] This chapter adapts existing guidelines to each of the three variants of process-tracing.

We start by discussing some of the common challenges shared by the three variants of process-tracing methods. First, we discuss how concepts should be conceptualized, arguing that defining key concepts in set-theoretical terms is more compatible with process-tracing than viewing concepts as variables. We then elaborate on the common theoretical elements that causal mechanisms share in the three variants—in particular, the need to elaborate the content of each of the parts of the mechanism that are composed of entities that engage in activities. Activities are explicitly conceptualized in process-tracing to capture how causal forces are transmitted through a causal mechanism to produce an outcome. We then discuss how causal mechanisms differ across four types of theoretical explanation, the analytical level at which they operate, the degree of contextual specificity, and the time span in which they are theorized to operate.

The second section illustrates the specific challenge in working with theories in each of the three variants of process-tracing: theory-testing, theory-

building, and explaining-outcome. We show that the conceptualization phase of theory-testing process-tracing involves the mostly deductive task of using existing theorization to flesh out the causal mechanism between X and Y, whereas theory-building derives theories from a more inductive analysis of empirical material. In contrast, conceptualization is an ongoing and iterative process in explaining-outcome process-tracing, involving multiple stages of analysis of evidence and theoretical reformulation (see chapter 2).

4.1. Common Challenges Shared by All Three Variants

Defining Key Concepts

Key to any research design is the definition of the central concepts that form the basis for theoretical propositions. Adcock and Collier (2001) refer to this as the translation of abstract theoretical concepts into what they term *systematized concepts.* Most important theoretical concepts are contested, with multiple plausible meanings, and they are often quite ambiguous and vague at the abstract level. We therefore need clear systematized definitions that distinguish between what is included and not included in the concept that allow us to know it when we see it.

Defining systematized concepts is not an "anything goes" process. For most concepts, a considerable amount of theoretical work describes what the concept should include and not include, and our systematized concepts should be faithful to this work (Adcock and Collier 2001: 532). Conceptualization involves defining the constitutive dimensions of a concept and how they relate to each other (Goertz 2006). For example, a systematized definition of democracy could include the two characteristics of civil rights and competitive elections. These secondary characteristics obviously should not rule each other out. It is also important to note that concept formation in qualitative case study research such as process-tracing is attentive to the details of cases, resulting in more context-specific conceptual definitions that have a narrower scope than are commonly used in large-*n* analysis (Ragin 2008: 78–81).

There are important differences in how concepts are defined depending on whether they are considered to be variables or conditions. In the understanding put forth by King, Keohane, and Verba (KKV), theoretical concepts are variables. Causal relationships are then formulated in terms of an independent variable (X) (or set of variables) that are theorized to

cause variation in the dependent variable (Y) (King, Keohane, and Verba 1994). This type of theory usually describes a probabilistic causal relationship, where an increase in X raises the probability of Y occurring in a given case (Gerring 2005: 167). The term *variables* means that they must be able to vary. If we are conceptualizing democracy as a variable, then we also need to take into consideration whether democracy is a dichotomous variable (democracy—autocracy) or an interval scale variable (for example, from 0 to 7 as in the Polity measure of democracy [Marshall and Jaggers 2002]). This means that when viewing concepts as variables, we need to define the positive pole (full democracy), the negative pole (autocracy), and the scale of the intervening continuum (e.g., dichotomous, ordinal, interval, or continuous) (Goertz 2006: 35). In the Polity measure, the characteristics that compose the positive pole (democracy) are juxtaposed against another set of characteristics that define the negative pole (autocracy).

We argue that process-tracing methods are more closely aligned with the conceptualization of conditions as used in set theory than when they are understood as variables. Set-theoretical causal relationships describe a causal condition (or set of conditions) that is necessary and/or sufficient for the occurrence of an outcome.[2] Set-theoretical causal relationships are usually studied with comparative methods. For example, after comparing all of the cases that compose the set of social revolutions, Skocpol concludes that peasant village autonomy is a necessary causal condition for the occurrence of social revolutions (1979: 112–54).

When analyzing set-theoretical relationships, the focus is not on defining the full variation of the concept (differences in degree) but instead is on defining the concept itself and its negation (i.e., the concept is present or not present—differences in kind). For example, if we are conceptualizing democracy as a necessary condition, we need to have a full definition of the positive pole (democracy), but the negative pole (autocracy) would not have to be defined in the same manner, as only the presence or absence of democracy is under investigation. If we are studying a democratic peace mechanism, we are interested in studying democracy and its effects on peace (outcome). The negative pole (autocracy) is analytically irrelevant, since studying autocracy does not tell us anything about how democracy produces peace. Here, we would conceptualize democracy as having two poles: the characteristics associated with democracy, and a second pole where absence is defined as anything but democracy. Outcomes are conceptualized in the same manner, with the focus on the concept and its presence or absence and the negative pole defined as the absence of the outcome. In explaining-outcome process-

tracing, in contrast, the outcome is not defined as a theoretical concept (a case of . . .), but as the historical event to be explained (e.g., the Cuban Missile Crisis or the French Revolution).

When we engage in process-tracing, we need to define carefully the concept and its absence, but we do not need to include both the positive (democracy) and negative (autocracy) poles along with the nature of the continuum in-between. The definition of concepts in process-tracing is therefore usually closer to how concepts are defined in set theories, where we are interested in whether or not a concept is present. Defining concepts as variables, however, introduces superfluous elements.

For the purpose of case selection, in some research situations, information regarding varying degrees to which a cause or outcome is part of the set of a concept can be relevant (see chapter 8). Recently developed set-theoretical methods such as fuzzy-set QCA have opened the possibility of studying both differences in kind, defined as membership or nonmembership in a concept, and differences in degree of membership or nonmembership, ranging from full membership to more in than out (Ragin 2000, 2008). When we are, for example, selecting a most-likely case, it can be relevant to find a case where the theorized cause and outcome are present (differences in kind) as well as where the cause and outcome have high degrees of membership of the sets of the concepts. In contrast, a least-likely case will involve a cause and/or outcome that are merely more in than out, meaning that they are only marginally within the set of the concepts.

Concepts are not divorced from theory but are instead an intrinsic part of it (Goertz 2006). This view has implications for process-tracing in that the initial cause (e.g., democracy) should be conceptualized in a manner that includes the attributes that are causally relevant for the causal mechanism. If we want to study the democratic peace thesis, our definition of democracy (X) should include characteristics that lead to the mechanism that contributes to producing Y. For example, Owen defines democracy (liberal democracy) as "states that are dominated by liberal ideology, and that feature, in both law and practice, free discussion and regular competitive elections" (1994: 102). All of these characteristics are related to his theorized mechanism (chapter 5). In other words, conceptualizing democracy in this manner captures how causal forces are transmitted from X into the causal mechanism itself.

At the same time, it is important to be as conscious as possible about the causal hypotheses that are embedded within definitions of key concepts (Goertz 2006: 65–66). Goertz suggests that when conceptualizing in process-tracing, one should "think hard before hardwiring hypotheses into concepts.

Often hardwiring makes it difficult to test hypotheses down the road and often will raise problems when gathering data. . . . Avoid in particular hypotheses that come into play when using the concept on the dependent variable side. . . . Potential causes of the phenomenon should almost always be left out of the concept itself" (66). If one defines the outcome in this manner, a theoretical proposition can become a self-fulfilling prophesy.

The Implications of Taking Causal Mechanisms Seriously—Conceptualizing Causal Forces

In process-tracing, we theorize more than just X and Y; we also theorize the mechanism between them. Doing so involves defining a more encompassing number of theoretical concepts for all of the parts between X and Y. Conceptualizing in these terms enables us to capture the theorized process whereby causal forces are transmitted through a causal mechanism to produce an outcome; forces that are black-boxed in both frequentist and set-theoretical causal theorization. For example, Casillas, Enns, and Wohlfarth (2009) put forward a causal theory on the impact that public opinion (X) has on the decisions made by the U.S. Supreme Court (Y). They then collect data that enables them to test whether changes of public opinion in a more liberal direction are followed by changes in judicial decisions in a more liberal direction, however the causal mechanism linking X and Y is black-boxed in the analysis. In contrast, studying the causal mechanism(s) would require describing both the causal condition that starts the causal mechanism and the outcome as well as the theoretical mechanism between X and Y that produces the outcome. A process-tracing analysis of the theorized relationship between public opinion and Supreme Court decisions would analyze a causal mechanism that theorizes about the process whereby public opinion becomes salient for judges and how and when judges are theorized to respond to perceived shifts in public opinion.

Each of the parts of the causal mechanism should be conceptualized as composed of entities that undertake activities, as illustrated in figure 4.1. Entities engage in activities (the parts of the mechanism—i.e., toothed wheels), while activities are the producers of change, or what transmits causal forces through a mechanism (the movement of the wheels). Entities can be individual persons, groups, states, classes, or structural phenomena depending on the level of the theory. The theoretical conceptualization of the entities uses nouns, whereas activities should include verbs that define the transmitters of causal forces through the mechanism. In social science terms, social

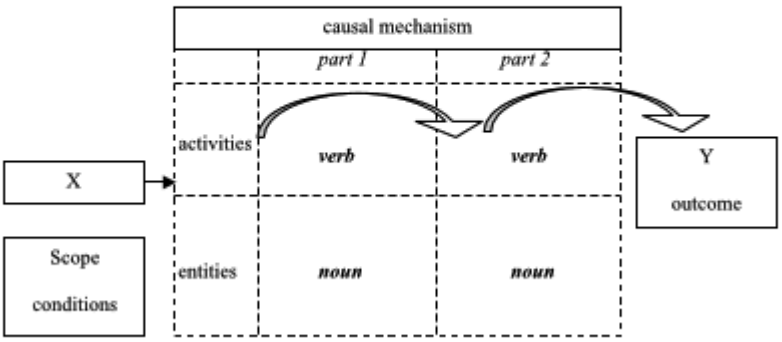

Fig. 4.1. A mechanismic conceptualization of a causal mechanism

entities have causal powers that can be understood as a "a capacity to produce a certain kind of outcome in the presence of appropriate antecedent conditions" (Little 1996: 37).

When conceptualizing the mechanism, each part should be seen as an individually insufficient but necessary part of the whole (see chapter 3). Parts have no independent existence in relation to producing Y; instead, they are integral parts of a "machine" that produces Y. Understood as one part of a causal mechanism, the engine of a car by itself has little utility in producing forward movement without a drivetrain or wheels. Understood in this manner, the necessity of parts of a mechanism has an important disciplining effect in our theoretical development, as redundant parts should be eliminated from the model.

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. If we then can confirm the existence of a hypothesized causal mechanism with process-tracing with a reasonable degree of certainty, we have produced strong evidence that shows how the theorized parts of the causal mechanism produce Y and shows that X and Y are causally connected by the mechanism (Bunge 1997, 2004).

The clear elaboration of mechanisms into parts composed of entities engaging in activities should be thought of as an ideal typical conceptualization that sometimes is not attainable in practice. For example, we cannot always capture the activities of entities for each part of a mechanism in our conceptualizations, especially when theorizing at the macrolevel, where structural "activities" can be difficult to conceptualize. (For a good discussion of some of the challenges relating to studying structure, see Wight 2006.)

Theory-testing and theory-building process-tracing seek to generalize, aiming at developing and testing theories beyond the context in which they were developed, whereas explaining-outcome process-tracing focuses on crafting a minimally sufficient explanation of a particular case. This distinction is important to note when we are conceptualizing a causal mechanism, as it determines the types of parts that we include in our theories. In a theory-centric study, only systematic parts that have causal effects beyond a particular case for the whole population of the phenomenon are included in the theoretical model.[3] For example, in Owen's study of the democratic peace mechanism, the phenomenon under investigation was the impact of democracy on the war-proneness of states (see chapter 5). The theoretical proposition (democratic peace) is taken by many scholars to be close to universal, being applicable to almost all times and places. Theory-building process-tracing attempts to trace the mechanism whereby X (or a set of Xs) contributed to producing an outcome. For example, what are the mechanisms whereby economic development contributes to producing democratic change? Here both concepts would need to be carefully defined, involving careful thinking about the scope of the theoretical proposition. Is the ambition to uncover a broad theoretical mechanism or a more bounded proposition that applies for example to East Asia in the past fifty years?

In contrast, explaining-outcome process-tracing starts with defining the particular outcome of interest. The focus here is less upon testing or building theoretical propositions and more on the outcome for its own sake. The French Revolution is an important outcome in and of itself as a consequence of its substantive historical importance. However, even here it is important to think carefully about a case's outcome to enable the use of existing theories as starting points for crafting a sufficient explanation of the outcome. For example, the French Revolution can be considered a case of either the theoretical concept of "social revolutions" or "political revolution," enabling us to draw on a series of existing theoretical tools that are then supplemented by new hypotheses after the analytical first cut (see below for more on explaining-outcome process-tracing).

In explaining-outcome process-tracing, nonsystematic, case-specific mechanisms would be included. An example could be a case-specific mechanism that describes how inclement weather impeded voter turnout in a specific election—a mechanism that could be vital to understanding the particular outcome but that probably does not have systematic, cross-case effects. Usually in explaining-outcome process-tracing studies, we cannot build a minimally sufficient explanation without including certain nonsystematic,

case-specific parts that are particularly significant in the case. For example, we cannot fully account for Indian democratization without capturing the role played by the Congress Party (Mahoney 2008: 416). Yet the role played by the party in Indian politics is unique: a mass-incorporating party in all other contexts has other effects. Therefore the role played by the Congress Party is a case-specific, nonsystematic part of the explanation.

4.2. Different Types of Causal Mechanisms

While the conceptualization of theoretical propositions as causal mechanisms should contain certain common elements, they will also differ in terms of the type of theoretical explanation, analytical level, the extent of the scope conditions of the mechanism (from "universal" mechanisms that are applicable across space and time to mechanisms explaining particular outcomes), and the temporal dimension.

Different Types of Theoretical Explanations

There are differences in what elements a typical causal mechanism contains depending on the type of theoretical explanation. Parsons (2007) helpfully distinguishes between four types of explanation within the social sciences: structural, institutional, ideational, and psychological. In this chapter, we focus on the conceptual commonalities of the different types of explanation, thought of in terms of building blocks that they share. For example, institutional mechanisms share different common building blocks than do ideational mechanisms. Chapter 6 discusses the challenges these different types of explanations encounter when we test them empirically.

Structural causal mechanisms focus on the exogenous constraints and opportunities for political action created by the material surroundings of actors (Parsons 2007: 49–52). Common building blocks for structural mechanisms include how certain preferences and a given material structure dictate observed behavior (or in a looser sense, create a pattern of structural constraints and incentives) (65). Another building block of structural mechanisms is that action is theorized to be a rational process (52). For structure to have any impact, actors have to react in predictable (rational) ways to their structural positions (52).

An example of a structural mechanism is found in the theorization on

electoral realignment in the U.S. context, where realignments at the congressional and local level are theorized to be the product of changes in demographic factors and other slow-moving structural mechanisms (Miller and Schofield 2008).

Institutional mechanisms are distinct from structural ones in that institutions are man-made and thereby can be manipulated. Institutions can be defined as "formal or informal rules, conventions or practices, together with the organizational manifestations these patterns of group behavior sometimes take on" (Parsons 2007: 70). Typical institutional mechanisms deal with how certain intersubjectively present institutions channel actors unintentionally in a certain direction. The exact content of institutional mechanisms is determined by which of the different subtypes of institutional theory is being utilized—sociological institutionalist mechanisms that have norms and institutional cultures as common building blocks, rationalists who share a focus on institution-induced equilibria, historical institutionalists who conceptualize mechanisms in ways that capture the unforeseen consequences of earlier institutional choices, prioritizing the building blocks of path-dependency and temporal effects. An example of an institutional mechanism is Streek and Thelen's layering mechanism, where progressive amendments and revisions slowly change existing political institutions (2005: 22–23).

Ideational mechanisms share the argument that outcomes are (at least partially) the product of how actors interpret their world through certain ideational elements (Parsons 2007: 96). Here, the focus is not on how structures or institutions constrain behavior but instead on how ideas matter in ways that cannot be reduced to the objective position of an actor. Common theoretical building blocks include the view that actions reflect certain elements of ideas and that elements arose with a degree of autonomy from preexisting objective conditions (i.e., ideas are not just manifestations of structures). An example is Khong's (1992) mechanism that theorizes how historical analogies impact how actors interpret the world, making certain foreign policy choices more likely than would otherwise have been the case.

Finally, *psychological mechanisms* deal with mental rules that are hardwired into the human brain, resulting in behavioral regularities. Common building blocks include theorization about how and how much internal psychological dispositions interacted with other factors to produce action. An example is Janis's groupthink mechanism, where the innate social needs of individuals are theorized to produce a mechanism that results in poor decision-making processes that are dominated by premature consensus.

Different Analytical Levels, Contextual Specificity, and Temporal Conditions

Another key difference between mechanisms is whether they are theorized to be at the microlevel, macrolevel, or linking the two. In chapter 3, we introduced the debate about analytical levels in theorizing causal mechanisms, distinguishing between purely macrolevel, purely microlevel (action-formation mechanisms), and micro-macro mechanisms (transformational) and macro-micro (situational) mechanisms. The pragmatic middle ground that we espouse in this book states that mechanisms can hypothetically exist at the macro- or microlevels or can span the two levels (situational and transformative mechanisms). The choice of level at which to analyze a causal mechanism depends pragmatically on at which level the empirical manifestations of a theorized mechanism are best studied.

Another dimension of difference regards the degree of contextual specificity of mechanisms, defined as the scope conditions that are necessary for a given mechanism to function (Falleti and Lynch 2009; Walker and Cohen 1985). Context can be defined as the relevant aspects of a setting where the initial conditions contribute to produce an outcome of a defined scope and meaning through the operation of a causal mechanism (Falletti and Lynch 2009: 1152).[4]

When defining context, it is first important to make clear what phenomenon is under investigation. In other words, what is the phenomenon a case of, and what is the population of cases to which a theoretical proposition refers (Gerring 2007a)? Theoretical mechanisms can range from very broad, law-like propositions to bounded propositions applicable only to a small population of cases delimited in time and space or even case-specific mechanisms explaining the causes of a particular outcome. The bounds of applicability of the causal mechanism need to be explicitly theorized by defining the context within which the mechanism is expected to operate (Abbott 1997; Falletti and Lynch 2009). Is it theorized to be a causal mechanism that is broadly applicable across space and time or applicable within a narrow context, or is it case-specific? Defining the context in which a mechanism is expected to function is vital, as the same causal mechanism placed in two different contexts can hypothetically contribute to producing two very different outcomes (Falleti and Lynch 2009: 1160).

Finally, causal mechanisms also vary on the temporal dimension according to both the time horizon of the causal forces that produce an outcome and the time horizon of the outcome. We adapt Pierson's (2003, 2004) theorization on the temporal dimension in causal theory to theorization of causal mechanisms in process-tracing.

Many scholars traditionally have theorized causal relationships in terms of short time horizons with regard to the cause, the mechanism, and the outcome. Khong's (1992) process-tracing analysis of the impact of analogous reasoning in U.S. government decision making in the escalation of the Vietnam War is an example of a short-term mechanism in terms of both mechanism and outcome.

Yet theories can vary depending on the length of the time within which the mechanism is theorized to be acting and the time horizon of the outcome (Pierson 2004). Incremental causal mechanisms have causal impacts that first become significant only after they have been in action over a long time period. In threshold-type causal mechanisms, after an incremental mechanism has been in play for a long period, a cusp is reached, after which the outcome becomes immediately apparent. For example, many analyses of treaty reform in the EU take a relatively short-term view, focusing on the grand bargains between governments on more integration (e.g., Dyson and Featherstone 1999; Moravcsik 1998). Other scholars working with historical institutionalist types of explanations have contended that this snapshot-type analysis misses the important long-term effects of an institutional mechanism of "informal constitutionalization," which is theorized as an incremental process of small decisions by actors that over time accumulate, resulting in the creation of a structure that forms a pro-integrative context for governmental decisions (Christiansen and Jørgensen 1999; Christiansen and Reh 2009).

In addition, one can theorize that the outcome of a causal mechanism can first become apparent over a longer time period (Pierson 2004: 90–92). An example is seen in the work of Campbell (2005), who contends in an analysis of globalization (X) and institutional change (Y) that institutional change is not always self-evident. Different forms of institutional change are theorized to vary depending on their time span, including evolutionary change composed of small incremental steps along a single path; punctuated equilibrium, where nothing happens for long periods followed by a period of relatively rapid and profound institutional change; and punctuated evolu-

TABLE 4.1. The Temporal Dimension of Causal Mechanisms

		Time Horizon of Outcome	
		Short	Long
Time horizon of mechanism producing an outcome	Short	Normal "Tornado-like"	Cumulative effects "Meteorite/extinction"
	Long	Thresholds "Earthquake-like"	Cumulative causes "Global warming"

Source: Adapted from Pierson 2003: 179, 192.

tion, where the two forms of change are combined (long periods of evolutionary change followed by a rapid alteration) (33–35).

It is therefore important to theorize explicitly about the time dimension involved in the workings of a mechanism along with how an outcome manifests itself. A longer-term mechanism will look very different from a short-term mechanism; in particular, these differences manifest themselves in the types of observable implications that an incremental, long-term mechanism will be expected to have in comparison to a short-term mechanism (see chapter 6). In an incremental mechanism, we should expect small, almost unnoticeable empirical traces that will be apparent only if one knows what one is looking for.

Threshold mechanisms are also challenging to study empirically. In this type of mechanism, there is very little observable evidence available until the mechanism has reached a cusp, after which a very sudden development (outcome) occurs. This type of mechanism could be mistaken for a short-term mechanism, but incorrectly theorizing that the outcome resulted from a short-term mechanism would cause us to miss the longer-term incremental process that was most important for producing the outcome.

We now turn to the challenges relating to the conceptualization phase for each of the three variants of process-tracing research strategies.

4.3. Theory-Testing Process-Tracing

In theory-testing process-tracing, we know both X and Y, and we either have existing conjectures about a plausible mechanism or are able to deduce one relatively easily from existing theorization. If a causal mechanism is theorized to require both X and Z to function, both should be included when conceptualizing. This can take the form of (X + Z) → mechanism → Y. In the rest of this section, we discuss the simpler situation where a causal mechanism has a monocausal start (only X).

Conceptualization in theory-testing process-tracing starts as a deductive exercise. Using logical reasoning, we formulate a plausible causal mechanism whereby X contributes to producing Y, along with the context within which we expect it to operate. In practice, theory-testing has many inductive elements—for example, when we review existing empirical work for ideas about how we can flesh out the logical steps in a mechanism to transform a causal theory (X → Y) into a causal mechanism.

The amount of logical work necessary to flesh out a causal mechanism depends on whether existing theories are formulated in terms of mere cor-

relations (X:Y), as plausible causal links between X and Y (e.g., intervening variables), or as full-fledged causal mechanisms. Most common is the situation where we know X and Y but where the process (i.e., causal mechanism) whereby X causes Y has not been explicitly conceptualized. For example, in the three political science journals with the highest impact factor, in the period from 2005 to 2010, fewer than 10 percent of articles even mentioned mechanisms in their theorization of causal propositions.[5]

Even if causal mechanisms have been formulated in prior research, we need to ensure that they are conceptualized in a manner that explicitly captures the transmission of causal forces through a mechanism. Despite using the term *mechanism*, most studies are conceptualized as a series of intervening variables where the transmission of causal forces is not explicitly theorized (see also section 3.3.). For example, Finkel, Pérez-Líñan, and Seligson (2007) conceptualize what they term a causal mechanism linking democratic promotion by external powers through foreign assistance (X) with democratic change in a receipt country (Y). They then describe both an indirect mechanism and a direct effect mechanism of foreign assistance. For illustrative purposes, we focus only on the direct effect mechanism, noting that the indirect effect mechanism suffers from the same conceptual problem. The direct effect mechanism is formulated in the following manner:

> Targeted democracy assistance, by contrast, works to educate and empower voters, support political parties, labor unions, and women's advocacy networks, strengthen human rights groups, and otherwise build "constituencies for reform"; it thus attempts to influence democratic outcomes in both the short term and the medium term. (410)

While this conceptualization describes the first part of the mechanism (from assistance to constituencies for reform), it does not explicitly theorize how the constituency for reform produces democratic change. Instead, it merely posits that a condition (existence of a constituency for reform) potentially acts as an intervening variable that can produce democratic change. The analytical result of this lack of conceptualization of the causal forces whereby X contributes to producing Y is that their empirical analysis only investigates inputs (foreign assistance) and outputs (democratic change), with no investigation of the theoretical process linking X and Y.

We argue that if scholars want to take seriously the study of causal mechanisms, they need to conceptualize mechanisms in a manner that enables us to study what happens between X and Y. For Finkel, Pérez-Líñan, and Seligson, the mechanism could be reconceptualized by logically gaming

through the theoretical process whereby foreign democracy assistance can plausibly influence democratic outcomes, with more explicit theorization of how democracy assistance builds constituencies for reform (part 1) and then how these constituencies pressure the government to engage in democratic reform (part 2). Conceptualizing the mechanism in this manner would force us to investigate empirically the process within the black box of the mechanism that is theorized to link X with part 1, part 2, and the outcome.

A good starting point for conceptualizing a plausible causal mechanism for a given theorized causal relationship is to start with a thorough reading of the existing theorization on the phenomenon. Review articles are often particularly useful, as are descriptions of the state of the art in peer-reviewed journal articles and books. It is important to note that this reading should be as encompassing as possible. In theorizing about the mechanisms whereby nonstate supranational actors such as the EU Commission can wield political power in EU negotiations, Moravcsik (1999) not only incorporates theoretical scholarship on the phenomenon itself but also casts his net wider by finding inspiration from general bargaining theory, international relations theory, and theories of American and comparative politics.

The next step is to game through the different steps of a hypothesized mechanism, filling in the dots between X and Y to detail the nuts, bolts, wheels, and cogs between them. One way to start is to draw a mind map of plausible links between X (or a set of Xs) and Y, using boxes to illustrate each part of the mechanism. Each of the parts is insufficient to produce the outcome on its own, but all are necessary. In particular, we must focus on conceptualizing the entities and their activities. In practice, we cannot always do this, especially with regard to macrolevel mechanisms, where the activities of structural entities are not always self-evident.

Examples—Moravcsik's Study of EU Negotiations and Ghiecu's Study of Norms

An example of how the conceptualization of mechanisms can look in theory-testing process-tracing can be seen in Moravcsik (1999). While Moravcsik uses the term *variables* for the parts of his institutional mechanism (independent, intervening, and dependent), given that the intervening variables have no independent existence but only make sense as part of the theorized mechanism, it makes more theoretical sense to think of them as parts of a causal mechanism, each of them necessary to produce supranational influence over negotiation outcomes (Y) (275). Here, the parts of the mechanism

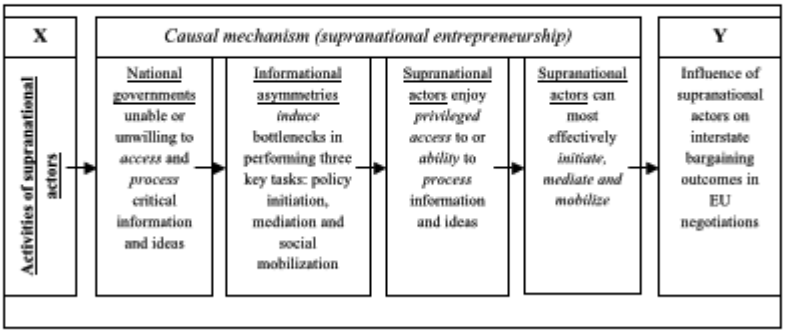

Fig. 4.2. A causal mechanism of how supranational actors can influence EU negotiations. (Based on Moravcsik 1999: 275.)

are depicted in terms of the entities engaging in activities, although activities can also take the form of nonactivity (see part 1 in figure 4.2). X is the activities of supranational actors (EU Commission), which produces supranational influence through the supranational entrepreneurship causal mechanism that is conceptualized as composed of four parts, each of which can be thought of as individually insufficient but necessary parts of the mechanism.

It is important to consider how many parts of the mechanism are logically necessary. A given mechanism should be kept as parsimonious as possible when we are engaging in theory-testing process-tracing. But at the same time, the process whereby X produces Y needs to avoid large logical gaps. In the case of Finkel, Pérez-Líñan, and Seligson, a substantial gap exists between the constituencies for reform and the outcome (democratic reforms), and additional parts of the mechanism should be theorized to fill in the gap.

Another example of a causal mechanism is found in Gheciu (2005), where she develops two ideational causal mechanisms that theorize the impact that the activities of international institutions such as NATO as an institution (X) have on the socialization of state actors to the norms of the institution (Y). Drawing on existing theorization in several different fields, Gheciu develops two different theoretical mechanisms that can result in socialization: persuasion and teaching. Gheciu fleshes out the causal relationship by developing the scope conditions under which the persuasion mechanism is theorized to work, thereafter detailing a simple two-part ideational mechanism where (1) actors engaged in persuasion use arguments to get other actors to rethink their conclusions, and (2) actors try to present a given course of action as the "right thing to do" (981–82).

In theorizing a mechanism, it is important to keep in mind that the theory should be internally consistent, meaning that two parts cannot contradict one another. Further, the outcome must be an outcome, not a potential cause. In addition, one should choose to include observable concepts whenever possible.

4.4. Theory-Building Process-Tracing

> It is of the highest importance in the art of detection to be able to recognize out of a number of facts which are incidental and which vital. Otherwise your energy and attention must be dissipated instead of being concentrated.
>
> —Sherlock Holmes (A. C. Doyle 1975: 138)

In its purest form, theory-building process-tracing starts with empirical material and uses a structured analysis of this material to detect a plausible hypothetical causal mechanism whereby X is linked with Y. While this is at heart an inductive exercise, existing theorization is usually used to inspire us in collecting evidence on which we can build theories.[6]

Theory-building process-tracing is utilized in two different types of research situations.[7] The first situation is when we know that a correlation exists between X and Y but are in the dark regarding potential mechanisms linking the two. In this form of X-Y-centric theory-building, the analyst examines a typical case to uncover a plausible causal mechanism that can be tested empirically in subsequent research. The second variant of theory-building process-tracing is when we know an outcome (Y) but are unsure what caused it to happen (i.e., X is unknown).

While theory-building process-tracing has some significant overlap with the explaining-outcome variant, a number of key differences exist (see also chapter 2). Of relevance here is the difference in the type of causal mechanisms being traced. Theory-building seeks to detect a systematic and relatively simple mechanism that contributes to producing an outcome across a bounded context of cases, whereas explaining-outcome uses eclectic conglomerates of systematic and nonsystematic mechanisms instrumentally to craft a minimally sufficient explanation for a particular outcome.

Conceptualization prior to empirical analysis is necessary in theory-building process-tracing only with regard to X and Y in X-Y-centric or Y in Y-centric theory-building. The actual mechanism between X and Y is naturally not conceptualized at the start.

Theory-building then proceeds to investigate the empirical material in

the case, using evidence as clues about the possible empirical manifestations of an underlying causal mechanism between X and Y that fulfills the guidelines for a properly conceptualized causal mechanism. This process involves an intensive and wide-ranging search of the empirical record: In the words of Sherlock Holmes, "'Data! Data! Data!' he cried impatiently. 'I can't make bricks without clay'" (A. C. Doyle 1892: 343).

Theory-building process-tracing seeks to uncover middle-range theories formulated as a causal mechanism that works within a bounded context—for example, spatially (by region, such as Southeast Asia) or temporally (such as post–World War II). According to Evans, to "be useful, these configurations had to be conceptualized in ways that were potentially separable from the settings in which they were originally derived" (1995: 6).

As chapter 8 develops further, theory-building process-tracing is usually part of a larger mixed-method research design, where the theory that is developed is then tested using either process-tracing or another form of theory test (e.g., a fsQCA analysis).

Observable evidence does not speak for itself. Theory-building often has a deductive element in that scholars seek inspiration from existing theoretical work and previous observations. For example, an analyst investigating socialization of international administrative officials within international organizations could seek inspiration in theories of domestic public administration or in psychological theories of small-group dynamics while reading more descriptive accounts of the workings of international organizations as sources for plausible causal mechanisms. In other situations, the search for mechanisms is based on hunches drawn from puzzles for which existing work cannot account.

An Example—Janis on Groupthink

Janis (1983) attempts to build a causal mechanism that details how conformity pressures in small groups can have an adverse impact on foreign policy decision making, using a selection of case studies of policy fiascoes (Y) that were the result of poor decision-making practices by small cohesive groups of policymakers (X). He uses the term *groupthink* to describe the causal mechanism that details how conformity pressures in small groups produce premature consensus.

The first exploratory case he uses is the Bay of Pigs fiasco. He notes that the groupthink mechanism is not expected to be the sole cause of fiasco (Janis 1983: 32), but he also notes a puzzle for which existing explanations

cannot account: Why did the "best and the brightest" policymaking group in the Kennedy administration not pick to pieces the faulty assumptions underlying the decision to support the intervention. "Because of a sense of incompleteness about the explanation," Janis "looked for other causal factors in the sphere of group dynamics" (32–33).

For each case study, Janis starts by drawing on psychological theories of group dynamics; relevant political science theories, such as Allison's (1971) organizational model; and Janis's own previous research for clues about potential systematic mechanisms. His search for parts of the mechanism is also informed by accounts of the Bay of Pigs decision. For example, when Janis "reread Schlesinger's account, I was struck by some observations that earlier had escaped my notice. These observations began to fit a specific pattern of concurrence-seeking behavior that had impressed me time and again in my research on other kinds of face-to-face groups. . . . Additional accounts of the Bay of Pigs yielded more such observations, leading me to conclude that group processes had been subtly at work" (1983: vii). Here we see the importance that imagination and intuition play in devising a theory from empirical evidence, although the search is also informed by existing empirical research.

Step 1 involves collecting empirical material to detect potential observable manifestations of underlying causal mechanisms. Empirical evidence is then used to infer that observable manifestations existed (step 2), resulting in the secondary inference that an underlying mechanism was present in step 3. Janis writes, "For purposes of *hypothesis construction*—which is the stage of inquiry with which this book is concerned—we must be willing to make some inferential leaps from whatever historical clues we can pick up. But I have tried to start off on solid ground by selecting the best available historical writings and to use as my springboard those specific observations that appear to be solid facts in the light of what is now known about the deliberations of the policy-making groups" (1983: ix). Further, "What I try to do is to show how the evidence at hand can be viewed as forming a consistent psychological pattern, in the light of what is known about group dynamics" (viii).

Janis's presentation of the empirical evidence is not in the form of an analytical narrative describing events or causal steps between X and Y. Instead, he writes, "Since my purpose is to describe and explain the psychological processes at work, rather than to establish historical continuities, I do not present the case studies in chronological order. The sequence I use was chosen to convey step-by-step the implications of group dynamics hypotheses" (1983: viii–ix). He describes four different "symptoms" of groupthink that

can be understood as observable manifestations of a groupthink mechanism, including the illusion of invulnerability held in the group, the illusion of unanimity within the group, the suppression of personal doubts, and the presence of self-appointed mind guards in the group. For example, the shared illusions of invulnerability and unanimity helped members of the group maintain a sense of group solidarity, resulting in a lack of critical appraisal and debate that produced a dangerous level of complacent overconfidence.

Janis concludes, "The failure of Kennedy's inner circle to detect any of the false assumptions behind the Bay of Pigs invasion plan can be at least partially accounted for by the group's tendency to seek concurrence at the expense of seeking information, critical appraisal, and debate" (47).

4.5. Explaining-Outcome Process-Tracing

Explaining-outcome process-tracing refers to case studies whose primary ambition is to explain particular historical outcomes, although the findings of the case can also speak to other potential cases of the phenomenon. Examples of explaining-outcome process-tracing include Allison's (1971) classic study of the Cuban Missile Crisis, Wood's (2003) study that attempted to explain the puzzle of insurgent collective action in the high-risk circumstances of the Salvadoran civil war, and Schimmelfenning's (2001) study of why countries that were skeptics about EU enlargement decided to support it.

Explaining-outcome process-tracing is an iterative research process where theories are tested to see whether they can provide a minimally sufficient explanation of the outcome. Minimal sufficiency is defined as an explanation that accounts for an outcome, with no redundant parts. In chapter 2, we described the inductive and deductive paths in explaining-outcome process-tracing.

The first stage of conceptualization in explaining-outcome process-tracing involves examining existing scholarship for potential mechanisms that can explain the particular outcome. Here, one suggestion is to discuss what Y is a potential case of, although historical outcomes typically contain multiple theoretical phenomena. Wood's (2003) study, for example, views the Salvadoran civil war as a case of insurgent mobilization.

In most explaining-outcome studies, existing theorization cannot provide a sufficient explanation, resulting in a second stage in which existing theories are reconceptualized in light of the evidence gathered in the preceding empirical analysis. The conceptualization phase in explaining-outcome process-tracing is therefore an iterative research process, with initial mecha-

nisms reconceptualized and tested until the result is a theorized mechanism that provides a minimally sufficient explanation of the particular outcome. The revised theoretical mechanism is then tested on its own terms on new evidence gathered from the same case.

In developing sufficient explanations, the following strategies can be utilized: combining existing mechanisms (eclectic theorization), developing new theories (or parts thereof), or incorporating nonsystematic parts into an explanation to account for the outcome. Here we discuss eclectic theorization and the incorporation of nonsystematic parts; the question of developing new theories using inductive research was discussed in section 4.4.

Eclectic theorization can be conceived as the combination of different mechanisms in a complex composite to craft a minimally sufficient explanation of a particular outcome (Sil and Katzenstein 2010). It "offers complex causal stories that incorporate different types of mechanisms as defined and used in diverse research traditions [and] seeks to trace the problem-specific interactions among a wide range of mechanisms operating within or across different domains and levels of social reality" (419). According to Hirschman, "Ordinarily, social scientists are happy enough when they have gotten hold of one paradigm or line of causation. As a result, their guesses are often farther off the mark than those of the experienced politician whose intuition is more likely to take a variety of forces into account" (quoted in Sil and Katzenstein 2010: 413–14).

Eclectic theorization does not seek to create synthetic grand theories; rather, it is a more pragmatic strategy aimed at capturing the multiplicity of mechanisms that produce particular historical outcomes. For this reason, eclectic theorization is also termed *problem-oriented research*. According to Evans, "Cases are always too complicated to vindicate a single theory, so scholars who work in this tradition are likely to draw on a mélange of theoretical traditions in hopes of gaining greater purchase on the cases they care about" (1995: 4).

However, while mechanisms from different research traditions can be combined, it is also important to make sure that key concepts and theoretical assumptions are compatible with one another to explain a concrete problem (Sil and Katzenstein 2010: 414–15). For example, one cannot combine a theorized ideational mechanism that contends that subjective beliefs drive actor behavior with an institutionalist mechanism where behavior is driven purely by the rational maximization of material interests. Here, one would have to reconceptualize the two mechanisms by, for example, developing a form of bridging theory that explains interaction between the two mecha-

nisms and develops scope conditions for when one or the other mechanism is expected to dominate.

Given the ambition to craft a minimally sufficient explanation of a particular outcome, it is usually necessary to include nonsystematic mechanisms or parts of mechanisms in the explanation. An illustrative example of how a nonsystematic mechanism can be added to an explanation is found in Layne (2006), which employs explaining-outcome process-tracing to explain why U.S. grand foreign policy strategy after the early 1940s was dominated by an extraregional hegemony strategy, an outcome that cannot be explained using other theories. The outcome being explained is not great power behavior in general but rather a particular historical case (U.S. grand strategy after the early 1940s vis-à-vis Western Europe).

Layne (2006) undertakes an analytical first cut that uses Mearsheimer's (2001) offensive realist theory to test whether a structural causal mechanism based solely on relative power and geography can explain the aggressive U.S. hegemonic strategy. Mearsheimer's structural mechanism contends that based on the strong relative power of the United States, we should expect that the system pushes the United States towards a global hegemonic strategy, but that the stopping power of water prevents the United States from attaining that goal. The United States cannot project enough power outside of the North American region to dominate powers in other regions such as China or Russia. However, Layne finds that Mearsheimer's theory can explain only a more limited version of hegemony—what he terms offshore balancing—and cannot account for the outcome (U.S. extraregional hegemony in Western Europe after World War II).

Layne (2006) then draws on existing historical scholarship to build a case-specific ideational mechanism that can be termed the "Open Door" mechanism, illustrated in figure 4.3.[8] This ideational mechanism links strong U.S. relative power with the particular outcome. As the figure shows, the parts of the causal mechanism are case-specific, relating to factors that are based on U.S. domestic beliefs and ideas that are unique to the United States, temporally specific to the post-World War II period, and geographically restricted to Western Europe. They therefore cannot be used to account for the grand strategy of other great powers in other cases, making the mechanism nonsystematic, or what we term case-specific.

Layne (2006) contends that without this case-specific Open Door mechanism, we cannot explain U.S. extraregional hegemony strategy vis-à-vis Western Europe. Further, he notes that the mechanism explains puzzles relating to the case for which other theories cannot account, such as why

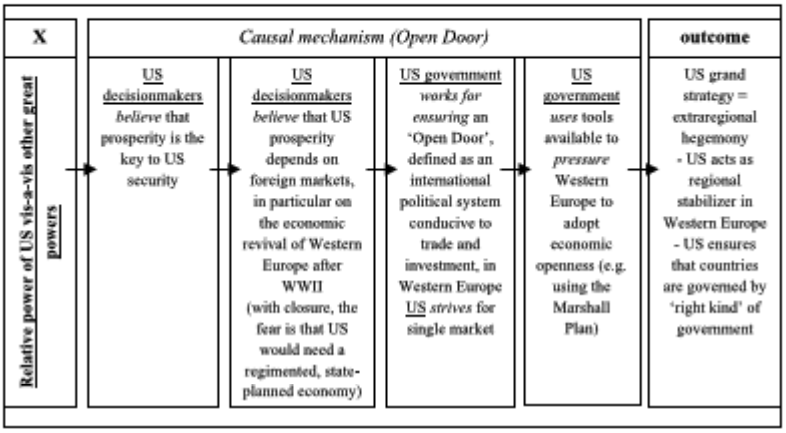

Fig. 4.3. Layne's case-specific Open Door mechanism. (Based on Layne 2006.)

the United States has remained in Western Europe after the Soviet Union's collapse (38).

Scholars choose explaining-outcome process-tracing precisely because they prioritize accounting for cases about which they care more than they prioritize theoretical parsimony. Here, claims about what makes a good theoretical explanation are based primarily on an explanation's ability to account for particular outcomes. Therefore, both eclectic theorization and the inclusion of nonsystematic mechanisms are prevalent.

An Example—Schimmelfennig's Study of Eastern Enlargement

A good example of how theories of causal mechanisms are developed in explaining-outcome process-tracing can be seen in Schimmelfennig (2001). The article attempts to explain a particular empirical puzzle—why France and other countries that initially opposed eastern enlargement of the EU eventually supported it (49).

Schimmelfennig's case study uses three iterations of the deductive path (see figure 2.4). He takes as his point of departure two competing theorized causal mechanisms from rationalist and sociological theories of international cooperation to explain the existing EU member states' positions regarding eastern enlargement. He finds that a rationalist mechanism can account for initial national preferences but not for the final decision to enlarge. In-

formed by the findings of his first empirical analysis, he tests whether a sociological mechanism can account for the full outcome, finding that it can account for France's final decision to accept enlargement but not for its change of position. Thus, neither mechanism can fully explain the outcome (neither is sufficient alone, and they are not sufficient in combination with each other), finding them both "wanting in the 'pure' form" (2001: 76). He then uses the empirical results of the first two iterations to formulate an eclectic combination of the two mechanisms that attempts to "provide the missing link between egoistic preferences and a norm-conforming outcome" by developing the idea of rhetorical action (the strategic use of norm-based arguments) (76).

The eclectic conglomerate mechanism composed of the three individual mechanisms provides a sufficient explanation of the historical outcome. Sufficiency is confirmed when it can be substantiated that there are no important aspects of the outcome for which the explanation does not account. In all three iterations, Schimmelfennig is tracing causal mechanisms. However, taken individually, the rationalist, sociological, and rhetorical action mechanisms are more generally applicable, whereas the eclectic combination of all three is much more case-specific and therefore cannot be exported per se to other historical cases.