

# Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation

American Journal of Evaluation  
31(3) 363-381  
© The Author(s) 2010  
Reprints and permission:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/1098214010371972  
<http://aje.sagepub.com>



Brad Astbury<sup>1</sup> and Frans L. Leeuw<sup>2,3</sup>

## Abstract

There is growing interest in the concept of “mechanism” across many areas of the social sciences. In the field of program and policy evaluation, a number of scholars have also emphasized the importance of causal mechanisms for explaining how and why programs work. However, there appears to be some ambiguity about the meaning and uses of mechanism-based thinking in both the social science and evaluation literature. In this article we attempt to clarify what is meant by mechanisms in the context of program evaluation by identifying three main characteristics of mechanisms and outlining a possible typology of mechanisms. A number of theoretical and practical implications for evaluators are also discussed, along with some precautions to consider when investigating mechanisms that might plausibly account for program outcomes.

## Keywords

theory-driven evaluation, social and behavioral mechanisms, realism, program theory

## Introduction

There is now a large and growing body of literature on “mechanisms” and the role of mechanism-based approaches to theory building in the social, behavioral, political, and economic sciences. A key theme in much of this literature is the notion that identifying mechanisms that link cause and effect relations is crucial for the development of deeper and more fine-grained explanations of social phenomena (e.g., Bunge, 1997, 2004; Elster, 1989, 2007; George & Bennet, 2004; Hedström & Swedberg, 1998; Lawson, 1997; Little, 1991; Machamer, Darden, & Craver, 2000; Mayntz, 2004; Steele, 2004; Stinchcombe, 1991; Tilly, 2001).

This widespread interest in mechanism-based explanation across the social science disciplines has slowly started to trickle over into the field of program and policy evaluation. This has occurred

---

<sup>1</sup>Centre for Program Evaluation, The University of Melbourne, Victoria, Australia

<sup>2</sup>University of Maastricht, Maastricht, The Netherlands

<sup>3</sup>Research, Statistics and Information Center, Ministry of Justice, the Netherlands Institute for Justice, Den Haag, the Netherlands

## Corresponding Author:

Brad Astbury, Centre for Program Evaluation, The University of Melbourne, Victoria 3010, Australia

Email: [brad.astbury@unimelb.edu.au](mailto:brad.astbury@unimelb.edu.au)

mainly through introduction of the term “mechanism” into various forms of theory-driven evaluation; most notably Chen’s (1990) “intervening mechanism evaluation” and the “realistic evaluation” of Pawson and Tilley (1997), with its emphasis on middle-range “context–mechanism–outcome” theories of programs. Although theory-driven evaluators largely agree that mechanisms are important, there seems to be some lingering confusion about the nature of mechanism-based theorizing. What exactly is a “mechanism” and how might the study of “mechanisms” be useful for evaluation research?

In this article, we explore these questions by drawing on the extant literature to offer a conceptualization of mechanisms and their role in evaluation research. We begin the article by orientating the reader with a quick history of the now familiar “black box” problem in evaluation and describe how this “problem” has led to the development of various types of theory-driven approaches to evaluation. Although the focus and form can vary, one key aim of theory-driven evaluation is to unpack programmatic “black boxes” and explain how and why programs work (or fail to work) in different contexts and for different program stakeholders. This is where the explicit use of mechanisms can play an important role in assisting theory-oriented evaluators to articulate more precisely the causal linkages between programs and their desired effects.

Next, we discuss at some length the concept of “mechanism” and attempt to elucidate what mechanisms are and what they are not. A typology of mechanisms is outlined and illustrated with an example to show how the identification of mechanisms can help evaluators unpack the assumptions underlying an intervention. Finally, we discuss theoretical and practical implications of mechanisms for the field of evaluation, as well as some precautions to consider when using mechanisms to explicate and test program theory.

## The Black Box “Problem” and Theory-Driven Evaluation

When evaluators talk about the black box “problem,” they are usually referring to the practice of viewing social programs primarily in terms of effects, with little attention paid to how those effects are produced. The antonym of “black box” evaluation is white box evaluation (sometimes also referred to as clear box evaluation). White box evaluation involves some attempt to “unpack” the black box so that the inner components or logic of a program can be inspected (Scriven, 1994). White box, or “theory-driven” evaluation as it is more commonly known, has a long history. Edward Suchman, in his 1967 book *Evaluative Research*, was perhaps the first evaluator to highlight the importance of opening up and empirically testing the “black box” of social programs. Weiss (1972, 1995, 1997a), Wholey (1979, 1983), and Chen (1989, 1990), among others, have also helped establish the importance of investigating the theory underlying social programs.

For the theory-driven evaluator, programs are embodiments of theories in at least two ways. First, they comprise an expectation that the introduction of a program or policy intervention will help ameliorate a recurring problem. Second, they involve an assumption or set of assumptions about how and why program activities and resources will bring about change for the better (Tilley, 2004). Another common thread running through the now vast literature on theory-driven evaluation is the concern that programmatic assumptions are often never made explicit and even on occasions when the underlying theory of a program is surfaced, it is not articulated and/or tested in a particularly robust way (Weiss, 1997b). This can be problematic, because if a program is based on a faulty theory, then it will not bring about desired changes, irrespective of how well it is implemented.

### *Developing Program Theory*

In contemporary evaluation practice, program theory is created in many different ways and used for a variety of purposes (Birkmayer & Weiss, 2000; Donaldson, 2007). Program theory can be

developed before a program is implemented (i.e., prospectively) or after the program has been running for some time (i.e., retrospectively). In addition, the way in which program theory is used will usually depend on a variety of factors, such as the particular circumstances of the evaluation; time and resource constraints; the stage of program development; methodological expertise; and the needs of program stakeholders. For example, focusing on program theory could be used as part of an evaluability assessment to help specify the program and determine the feasibility of implementing a full-scale impact study (Wholey, 1987). Alternatively, program theory development may be useful during a participatory evaluation to facilitate team building, staff buy-in, and stakeholder engagement. Policy makers might also benefit from applying theory-driven evaluation to clarify the design of a program prior to implementation and/or establish a performance monitoring framework. In addition, performance auditors can use it to check the robustness of a program's underlying assumptions. Another common application is the development of program theory during evaluation planning to help identify and prioritize key evaluation questions and guide the selection of data collection and analysis techniques.

The methodology for constructing or reconstructing program theory, as well as the level of detail and complexity, also varies significantly. Some examples of these include path analysis and causal modeling; observations of the program in action; interviews with staff to uncover implicit assumptions about how the program works; concept mapping exercises; formal (argumentational) analyses of program and policy documents; and detailed investigations of research on similar programs as well as social science theory (Chen, 1990; Leeuw, 1991; Lipsey, 1993; Trochim, 1989; Smith, 1990). Often evaluators recommend using a combination of these approaches, drawing on both primary and secondary data sources and multiple research methodologies (e.g., Donaldson, 2007, Connell, Kubish, Schorr, & Weiss, 1995; Leeuw, 2003; Pawson & Tilley, 1997).

Program theory is also often expressed in several different ways—a graphic display of boxes and arrows, a table, a narrative description, and so on. There are now many “logical templates” that are widely used by program staff and evaluators to assist in developing visual models of the hypothesized relationship between program resources, activities, and outcomes (e.g., McLaughlin, & Jordan, 1999; United Way of America, 1996; W. K. Kellogg Foundation, 2004). Regardless of the way program theory is developed and depicted, it should constitute a “plausible and sensible model of how a program is supposed to work” (Bickman, 1987, p. 5). It is also argued that theory-driven evaluation should ideally contain both a conceptual component (i.e., the development of the program theory or theories) *as well as* an empirical component (i.e., the systematic testing and refinement of the program theory; Rogers, Hasci, Petrosino, & Huebner, 2000).

Another issue that is commonly raised by proponents of theory-driven evaluation is the need to be more careful in the use of terminology (Davidson, 2006; Weiss, 1997b). Donaldson and Lipsey (2006) note, for example, that the contemporary evaluation landscape is littered with a confusing array of closely related terms such as “theory-based evaluation,” “theory-driven evaluation,” “program theory,” “theory of change,” “logic models,” “logical frameworks,” “intervention logic,” and so on.

In particular, while the terms “program theory” and “program logic” are often used interchangeably by evaluators, there appears to be growing recognition that they actually serve different functions (Chen, 2005; Leeuw, 2003; Rogers, 2007; Scheirer, 1987; Weiss, 1997a). Program logic is often used to identify and *describe* the way in which a program fits together, usually in a simple sequence of inputs, activities, outputs, and outcomes. Program theory goes a step further and attempts to build an *explanatory* account of how the program works, with whom, and under what circumstances. Thus, program theory might be seen as an elaborated program logic model, where the emphasis is on causal explanation using the idea of “mechanisms” that are at work.<sup>1</sup>

Although program logic and program theory can indeed be used in such a complementary fashion, this does not seem to be common practice in the field. Arguably, this is because the significance and importance of mechanisms is not well understood by evaluators. As Davidson (2000) has

observed “Despite the purported focus of theory-based evaluation on investigating the causal mechanisms by which a program achieves its effects, surprisingly few actually do this” (p. 18).

### *Use of the Term “Mechanism” in the Evaluation Literature*

Chen and Rossi were among the first evaluators to introduce the term “mechanism” and point out its importance for theory-driven evaluation. For example, in an early article, they argue that “the theory-driven approach avoids the pitfalls of black-box evaluation and provides better understanding of the *causal mechanisms* underlying the relationship between treatment and effects” (Chen & Rossi, 1987, p. 102, emphasis added). Later in Chen’s (1990) book, *Theory-Driven Evaluations*, again we see reference to the term “mechanism.” This time, a whole chapter is devoted to a particular type of theory-driven evaluation called “intervening mechanism evaluation.” This approach involves “identifying the causal processes that theoretically intervene between program treatment and outcome” (p. 191).

More recently, Chen (2005) identifies “two kinds of causal mechanisms [that] may underlie a program: mediating and moderating” (p. 240). He defines these as follows:

A mediating causal mechanism is a component of a program that intervenes in the relationship between two other components . . . [while] the second type of causal mechanism—moderating—represents a relationship between program components that is enabled, or conditioned, by a third factor (pp. 240–241).

Although Chen popularized the term “mechanism” in the early 1990s, in our view, a detailed treatment of the concept of “mechanism” did not appear in the literature until the publication of the book *Realistic Evaluation* by Pawson and Tilley (1997).<sup>2</sup> What makes the approach of Pawson and Tilley to program theory and mechanisms distinctive is that it is strongly based on the principles of causal explanation advanced by early pioneers associated with realist philosophy of science (e.g., Bhaskar, 1975; Harré, 1972). One of the main implications of the realist perspective for evaluation is that it is not enough to simply cite programs as a cause of outcomes—the mechanisms connecting causes and their effects must also be identified. As Pawson and Tilley put it: Programs work (have successful “outcomes”) only in so far as they introduce appropriate ideas and opportunities (“mechanisms”) to groups in the appropriate social and cultural conditions (“contexts;” p. 57).

A similar point is made by Carol Weiss (1997a) in a popular article reflecting on the past, present, and future of theory-driven evaluation. Weiss argues that it is very important for evaluators to distinguish between what she calls “implementation theory” (or as we would put it “logic models”), which provide operational details about how the program is carried out, and program theory which:

. . . deals with the *mechanisms* that intervene between the delivery of program service and the occurrence of outcomes of interest. It focuses on participants’ responses to program service. The mechanism of change is not the program service per se but the response that the activities generate (p. 46, emphasis in original).

So, according to a number of prominent evaluation theorists—Chen, Weiss, Pawson and Tilly and more recently Mark, Henry, and Julnes (2000) and Donaldson (2007)—if evaluators look closely inside the programmatic “black box,” they are likely to discover a mechanism or as is more often the case, several mechanism–context–outcome configurations.

Although the concept of mechanism is now part of the official evaluation lexicon, we are less confident that it is understood and applied well in practice. The argument developed in this article is that the potential role for mechanisms in evaluation research, as originally anticipated by several evaluation theorists, appears to have become overshadowed by a misunderstanding of what mechanisms are as well as a narrow focus on linear program logic modeling.<sup>3</sup> In the next section of this article, we look more

closely at why this might be the case and as a possible way forward offer a conceptualization of mechanisms, which is drawn from an extensive reading of the wider philosophical and social science literature.

## The Concept of Mechanism

Like so many words that are bandied about, “mechanism” can mean different things depending on the particular field of knowledge and context in which it is used.<sup>4</sup> As Mayntz (2004) observes, “. . . a survey of the relevant empirical and methodological literature soon bogs down in a mire of loose talk and semantic confusion about what ‘mechanisms’ are” (p. 238). If we are not careful, then, there is a risk that the word ‘mechanism’ like ‘theory’ may begin to “obscure rather than create understanding” (Merton, 1968).

In an effort to promote understanding, some scholars have compiled “running lists” of the range of definitions of “mechanism,” which can be found in the literature (Gerring, 2007; Hedström, 2005). For example, Mahoney (2001, 2003) identified 24 different definitions of the term mechanism, which he sourced from the writings of 21 different authors.<sup>5</sup> Rather than provide yet another definition, or another list of definitions, we will attempt to elucidate the concept of “mechanism” in two main ways. First, we discuss what mechanisms are not, particularly with respect to how the term is sometimes misunderstood in the context of program evaluation. Second, we identify three key characteristics of mechanisms that are located in various definitions put forward by sociologists, political scientists, evaluators, and philosophers of science in the last 40 years.

A common mistake is for evaluators to conflate the term mechanism with program activity. As noted earlier, we are not alone in pushing for a distinction between underlying “mechanisms of change” and a “to do” list of program activities (Rogers, 2007; Weiss, 1997a). Mechanisms appear too frequently as *unexplained “causal arrows”* that seem to flourish so well in the present climate of enthusiasm with visual logic models. This does not seem to be what theory-driven evaluators had in mind when they introduced the concept of “mechanism” to the evaluation community. It is worth quoting at length Weiss (1997a), who uses the example of a contraceptive counseling program to illustrate the distinction between mechanisms and program activities:

. . . if counselling is associated with reduction in pregnancy, the cause of change might seem to be the counselling. But the mechanism is not the counselling; that is the program activity, the program process. The mechanism might be the knowledge that participants gain from the counselling. Or it might be that the existence of the counselling program helps to overcome cultural taboos against family planning; it might give women confidence and bolster their assertiveness in sexual relationships; it might trigger a shift in the power relations between men and women. These or any of several other cognitive, affective, social responses could be the mechanisms leading to desired outcomes (p. 46).

Another, more complex concern, is when evaluators *directly* equate mechanisms with variables. Mechanisms are sometimes seen as independent causal variables (i.e., the  $X$  in the  $X \rightarrow Y$  formula), or more often, treated as an intervening variable or set of mediating or moderating variables that attempt to account for why a statistical correlation exists between an independent and dependent variable (i.e., the  $Z$  in the  $X \rightarrow Z \rightarrow Y$  formula). This does not distinguish adequately the idea of theoretical mechanisms from a statistical or “variable-centred type of theorising” (Hedström & Swedberg, 1998, p. 15; Mahoney, 2001, 2003; Pawson, 1989). Unlike variables, mechanisms are usually not observable attributes of some unit of analysis. Mechanisms attempt to explain why variables are related. In contrast, a mediator or mediating variable is an attempt to empirically measure the mechanism.<sup>6</sup>

To be sure, there are some parallels between “variable versus theoretical” perspectives on mechanisms. For example, statistical measurement and analysis can help to identify and describe

causal relationships between implementation variables and program outcomes. The results of this kind of quantitative causal modeling then provides the raw material for elaboration of theoretical models of the mechanisms that explain how statistical associations are generated. Although variables and mechanisms can perform complementary functions in evaluation research, it is important to avoid conflating the two. To do so risks losing the explanatory power of mechanisms. Thus, while the ontological status of mechanisms is still contested by philosophers, from a methodological point of view, we prefer to see mechanisms residing at a level of abstraction above variables.

There is considerable support for this view of mechanisms in much of the contemporary literature on mechanism-based analysis. Indeed, regardless of methodological or disciplinary traditions, most scholars who write about mechanisms offer a conceptualization that has been influenced to some extent, by realist accounts of causation (see e.g., George & Bennett, 2004; Henry, Julnes & Mark, 1998; Mahoney, 2003; Pawson & Tilley, 1997). According to this perspective, *mechanisms are underlying entities, processes, or structures which operate in particular contexts to generate outcomes of interest*. There are three essential clues located in a “realist” reading of mechanisms. These are that:

1. Mechanisms are usually hidden;
2. Mechanisms are sensitive to variations in context; and
3. Mechanisms generate outcomes.

### *Mechanisms are Usually Hidden*

It is common for realist philosophers and realist evaluation methodologists to emphasize that mechanisms are *underlying* and hence often unobservable or “hidden” (Pawson, 2008). This captures nicely the idea that to explain social regularities (or program outcomes), we cannot rely exclusively on repeated observations. Instead, we must go below the “domain of empirical,” surface-level descriptions of constant conjunctions and statistical correlations to identify the underlying mechanisms that account for “regularities in the joint-occurrence of events” (Bhaskar, 1975, p. 13; Pawson, 1989, p. 157).

A favorite metaphor used to demonstrate the spirit of realist, mechanism-based explanation is the clock. It is not possible to understand how a clock works by examining the surface—the numbers on the face and the movement of the hands. We need to prise the clock open and go beneath the “surface (observable) appearance” and delve into the “inner (hidden) workings” of the “balanced spring or the oscillation of caesium atoms” (Pawson & Tilley, 1997, p. 407).

Although useful in conveying the general tenor of mechanism-based explanation, a limitation of this “nuts and bolts” or “cogs and wheels” imagery is that it lends itself too easily to criticism on the grounds that human behavior cannot be understood in terms of mechanics or machines (Weber, 2006). Although this may be a reasonable *prima facie* objection, it does not lead automatically to the conclusion that studying mechanisms is not useful for social researchers and evaluators.

Suppose, for example, the results of successive evaluations reveal that students from disadvantaged backgrounds are statistically less likely than students from more privileged social positions to do well at school. The inquisitive evaluator and astute policy maker might ask: Why is this so? What is it about the nature of disadvantage that leads to underperformance at school? By identifying and testing plausible mechanisms underlying this empirical regularity, we “move beyond thinking about individual variables and the links between them to considering the bigger picture of action in its entirety” (Anderson et al., 2006, p. 103).

For instance, we might posit the existence of an underlying mechanism relating to expectations teachers have about disadvantaged students and their academic potential. This mechanism-based

conjecture might then be constructed as a propositional statement in terms of a warrant (Leeuw, 2003) as follows:

If a teacher expects disadvantaged students to underperform at school, then they will underperform (the proposition). This is because of the principle that expectations, even if initially false, are brought about because of the belief that they are true (the warrant).

To test this mechanism, we might randomly select 20% of students at the start of school term and tell their teachers that these students show unusual potential for intellectual growth. In the classic 1968 study “Pygmalion in the Classroom” Rosenthal and Jacobson did just this. They found that 8 months later these “gifted” children showed much greater gains in intelligence (as measured by standardized tests) than the remaining children. Rosenthal and Jacobson (1968) considered this to be strong evidence supporting the existence of an educational self-fulfilling prophecy or as others have called it a “belief-formation” mechanism (Hedström & Swedberg, 1998).

Basically, a belief-formation mechanism occurs when “an initially false belief of a situation evokes behaviour that eventually makes the false conception come true” (Hedström & Swedberg, 1998, p. 18). For instance, if we believe that we are about to meet a disagreeable person, we may approach that person so defensively that we turn them into a disagreeable person (Rosenthal & Jacobson, 1968). Belief-formation mechanisms are well documented and have been used to explain a variety of social phenomena, such as hypnosis; placebo and Hawthorne effects; the failure of banks and stock market crashes; and the processes of racial and religious prejudice (see Merton, 1968; Rosenthal & Jacobson, 1968, pp. 3–30; Schelling, 1998).

Belief-formation mechanisms are, of course, just *one* of potentially several, mechanism-based explanations that might account for why disadvantaged students underperform in school.<sup>7</sup> Belief-formation mechanisms may also interact and meld with other mechanisms. For example, teacher expectancies might shape the nature and quality of teacher–student relations, which in turn influences student motivation toward school (i.e., an academic self-efficacy mechanism). The important point is that mechanisms may not be observable, at least in a direct, empirical sense.

This, however, does not mean that mechanisms are not “real.” The vast majority of people do not deny the idea of electricity or gravity because they cannot see it; evidence of effects is usually sufficient (e.g., switching on a light or dropping a rock off a cliff). In general, then, it is possible to make a plausible case for the existence of underlying mechanisms by referring to observable effects *which can only be explained as products of underlying mechanisms*<sup>8</sup> (Sayer, 2000).

### *Mechanisms are Sensitive to Variations in Context*

A second key feature of a realist understanding of mechanisms is that mechanisms are sensitive to variations in context, as well as to the operation of other mechanisms in a particular context. Consider the following logic: In Context A, mechanism ( $M_1$ ) is not activated. That is,  $M_1$  is dormant; still possessing causal “tendencies” or “capacities” but not the conditions that “enable” it to be triggered. In Contexts B and C, the conditions are conducive to triggering  $M_1$ . However, in Context C, no effect or different effects are observed. This could be due to a countervailing mechanism ( $M_2$ ) that is present in Context C but not in Context B.

The implication of this logic is that mechanisms should not be seen as universal “covering-laws” that apply always and everywhere. Even in the physical world, casual “laws” vary (e.g., water only boils at 100 C when air pressure is at a certain level; Carter & New, 2004).

This is one of the reasons why mechanisms are often likened (after the sociologist Robert Merton) to middle-range theories that position themselves between universal social laws and description (Pawson, 2000, 2010). For Merton, mechanisms are “elementary building blocks of middle-range

theories” (Hedström & Swedberg, 1998, p. 6). Sayer (1981/1998) articulates the contingent nature of mechanisms (i.e., “causal powers”) as follows:

Gunpowder has the ‘causal power’ to explode in virtue of its unstable chemical structure. Copper can conduct electricity because of the presence of free ions in its chemical structure. Whether each of these causal powers are ever ‘realised’ or ‘activated’ depends upon contingently related conditions, such as the presence of oxygen, low humidity and a spark in the first case, and an electric current in the second (p. 124).

Recall the earlier example concerning teacher expectations of student performance. Self-fulfilling prophecies (or other belief-formation mechanisms) also do not operate in a deterministic, law-like fashion and may, or may not, be activated depending on the circumstances. Not all teachers, in all places, and at all times will hold the same belief about the educational potential of disadvantaged students. Some teachers, as well as students, may even actively resist the underlying mechanism theory.

Whether the causal “tendencies” of a particular mechanism is activated is largely dependent, then, on human reasoning and volition. This is because mechanisms work *through* human agents who have the (cognitive) capacity to think and act in terms of causalities and who also possess other capacities that make things happen. In practical terms, people do not react to programs like billiard balls that are hit; rather “programs only ‘work’ if people choose to make them work and are placed in the right conditions to enable them to do so” (Morén & Blom, 2003; Pawson & Tilley, 1994, p. 294). Thus, a key contextual aspect of the operation of mechanisms in the social world is human interpretation of social structures and events.<sup>9</sup>

This is not to say that mechanisms are necessarily context bound; just that context matters. Lawson’s (1997) notion of *demi-regularities* (“demi-regs”) is a useful term to depict the way context can affect mechanisms. For example, it can be argued that in total institutions like prisons “the strength of weak ties” mechanism (Granovetter, 1973), which points at the importance of having contacts with others outside personal friendship networks for finding information or jobs, works in a different way compared to what happens outside a prison context. One of the challenges for evaluation research is to detect “demi-regs,” which can prevent the belief that every program is unique and idiosyncratic.

### *Mechanisms Generate Outcomes*

A third characteristic of mechanisms is that they generate outcomes. The argument that unobservable causal entities (i.e., mechanisms) produce effects, differs considerably from standard depictions of causation promoted by David Hume and extended by John Stuart Mill. Hume believed that we can only know what we experience. Thus, causation cannot be directly observed—only inferred—by examining patterns of regular contingent relations between events.

This reasoning is familiar to many experimental evaluators who (often deliberately) treat programs as “black boxes” and their visible effects as light switches that can be turned on and off. House (1991) explains that one of the unfortunate legacies of Humean causality is that “things got turned around so that what was real was mistaken to be limited to only what we directly experience . . . and anything beyond was discredited as metaphysics” (p. 4).

In contrast, generative accounts suggest that analysis of causation should not stop with surface events (i.e., the things we can observe through our sense data). This is because reality is stratified, not flat, as Hume implied. This opens up the possibility of understanding the generation of events and regularities (such as program outcomes) *at different layers of reality*.<sup>10</sup> Deep, mechanism-based explanation focuses not only on outcomes themselves, and whether evaluators actually observe them happening, but also the underlying generative mechanisms that produce the



outcomes. Social programs, then, consist not just of what we observe (i.e., program inputs, activities, and outcomes) but also of interactions between mechanisms and contexts, which account for what we observe.

## **Types of Mechanisms and Levels of Analysis**

Although there are not yet repositories on mechanisms, several scholars have provided useful summaries of the research literature on mechanisms. Jon Elster (1989, 2007), for example, draws on insights from neuropsychology to economics and political science to identify and discuss some 20 mechanisms that underlie a range of social phenomena. Similarly, Farnsworth (2007) takes legal arrangements like laws and contracts as a starting point and dissects which types of mechanisms play a role when one wants to understand how these arrangements work. He discusses mechanisms such as the “slippery slope,” the “endowment effect,” and “framing effects”.<sup>11</sup> Theoreticians within the social sciences have also contributed to knowledge about mechanisms, as work by Festinger (1950, 1954, 1957), Merton (1968), and Olson (1971) has shown.<sup>12</sup>

In recent years, there have been some preliminary attempts to group mechanisms into common categories. For example, building on James Coleman’s (1986, 1990) classic macro–micro–macro model of social action, Hedström and Swedberg (1998) suggest that there are three interrelated types of mechanisms:<sup>13</sup> (a) situational mechanisms; (b) action-formation mechanisms; and (c) transformational mechanisms.

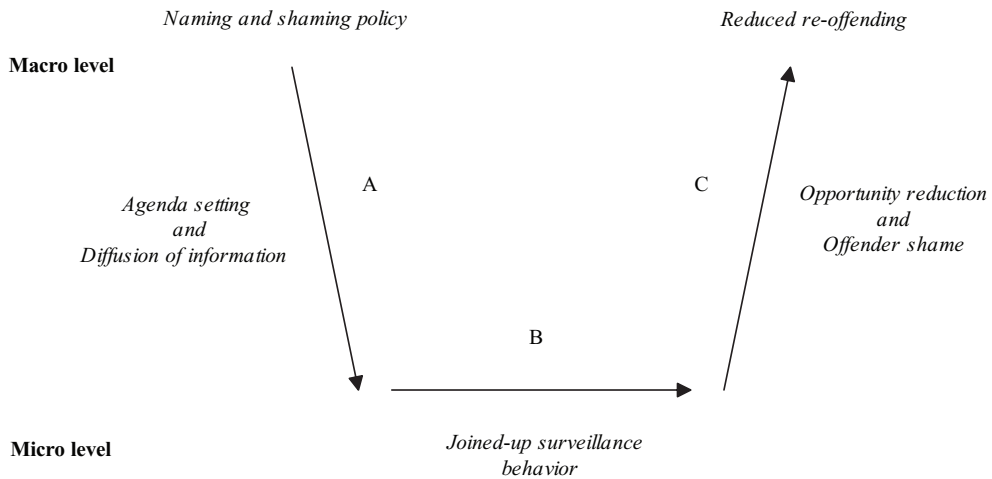
Situational mechanisms operate at the macro-to-micro level. This type of mechanism shows how specific social situations or events shape the beliefs, desires, and opportunities of individual actors. Belief-formation mechanisms such as the self-fulfilling prophecy are a good example of a situational mechanism. This mechanism has been shown to affect the way teachers interact with disadvantaged children in the classroom (as discussed earlier).

Action-formation mechanisms operate at the micro-to-micro level. This type of mechanism looks at how individual choices and actions are influenced by specific combination of desires, beliefs, and opportunities. Leon Festinger’s (1957) theory of cognitive dissonance illustrates different types of action-formation mechanisms that are used by individuals to reduce psychological distress that often arises when a person holds two contradictory ideas simultaneously. Smokers, for example, often use techniques of rationalization to avoid quitting despite strong evidence that smoking reduces life expectancy (e.g., “lung cancer only happens to heavy smokers”).

Transformational mechanisms operate at the micro-to-macro level and show how a number of individuals, through their actions and interactions, generate macro-level outcomes. An example is “cascading,” by which people influence one another so much that people ignore their private knowledge and rely instead on the publicly stated judgments of others. The “bandwagon phenomenon”—the tendency to do (or believe) things because many other people do (or believe) is related to this, as are “group think,” the “common knowledge effect,” and “herd behavior” (Elster, 2007).

### ***Applying the Typology of Mechanisms: A Selected Example***

How might this typology be useful for an evaluator who is seeking to open up a program or policy “black box”? Consider, for example, the case of “naming and shaming” interventions as analyzed by Pawson (2006). Over the last decade, policy makers, legislators, the police, and regulators have been increasingly active in implementing a range of naming and shaming programs. For example, dissemination of information via websites and newspapers about organizations that do not comply with rules and regulations (e.g., car manufacturing safety standards). Or establishing bulletin boards and other registration and notification activities about released sex



**Figure 1.** A basic model of mechanisms underlying “naming and shaming” of sex offenders. A = Situational mechanism; B = Action-formation mechanism; C = Transformational mechanism. Source: Adapted from Coleman (1986); Hedström and Swedberg (1998).

offenders, warning the community the ex-offender “is back.” It is believed that the naming and shaming of noncompliant and/or deviant behavior will lead to behavior changes in the desired direction.

We focus here on naming and shaming of released pedophilic sex offenders. Examples of interventions in this area are U.S. sex offender registration and community notification initiatives, such as the well-known Megan’s Law. This law was enacted in 1996 following the brutal rape and murder of Megan Kanka. Her killer was a released sex offender who, unbeknown to her parents, was living in their neighborhood. News of the case sparked a large public outcry, which resulted in the swift introduction of legislation allowing for a mandatory community notification system for convicted sex offenders (Pawson, 2006). In several other countries such as England, Scotland, the Netherlands, and Australia “look-alike” initiatives are discussed and sometimes implemented.

Applying the typology of mechanisms helps to reveal the (assumed) causal chain of this intervention. The result of our analysis can be found in Figure 1 which presents a basic model of mechanisms underlying naming and shaming policies for pedophilic sex offenders.

There are a range of possible mechanisms at work for each link in the concatenated chain. The links are also likely to be much more complex than the unidirectional arrows in the model suggest. However, for the purpose of illustrating how the typology of mechanisms might assist evaluation practice, we have decided to focus on just a few examples for each level of mechanisms. Where possible, we have also drawn from social science theory to make clear the significance of focusing on different types of putative mechanisms underlying social programs and policies.

The basic model identifies two examples of situational mechanisms, which can affect how sex offender naming and shaming policies work. *Agenda setting* describes the processes by which widespread and extensive media coverage can influence public opinion about how to respond to the issue of released sex offenders living in the community. For example, heightened salience and framing effects are likely to determine the way in which naming and shaming policy is put into practice (McCombs & Shaw, 1972; Scheufele, 1999). In a context of strong social pressure to “do something” about the problem, policy makers and law enforcement officials typically decide that investment in sex offender registration and information management systems are appropriate ways to enact policy.

It is assumed that these activities will then lead to a *diffusion* process, whereby appropriate persons will receive accurate and timely information about registered sex offenders who are residing in their local neighborhood. Diffusion describes how information about registered sex offenders spreads through particular communication channels over time among the members of a social system (Rogers, 2003). In practice, the dissemination of information can occur in a variety of ways. For example, mass media outlets such as the internet as well as interpersonal communication channels such as police-community consultative groups and informal meetings of concerned local citizens.

However, this is not enough to make naming and shaming work. The next link in the chain involves individual actors responding appropriately (or not) to information once it is received. According to the basic intervention theory of naming and shaming, police officers, members of the community, heads of school and teachers, sports mentors, and others will accept information about registered sex offenders, assume that is valid and start to act upon it. Acting upon implies increased *joined-up surveillance behavior*, which also implies that the framing of what suspicious behavior constitutes is more or less similar among different stakeholders in different contexts (Tilley, 1995). It also implies that the results of surveillance activities are fed back into information management systems.

Transformational mechanisms aggregate the individual joint (surveillance) actions to such a level that they are believed to make a difference for the behavioral choices of registered sex offenders. Aggregated surveillance increases the (perceived) likelihood of being caught leading to the possible activation of an *opportunity reduction* mechanism (Clarke, 1992). It also has the potential to trigger a *shame* inducing mechanism among potential re-offenders (Wortley, 1996). Finally, the joining up of the different mechanisms, so the theory goes, will lead to a reduction in re-offending.

This brisk reconstruction of the “supposed to do” theory of legislators and policy makers is, of course, just a starting point for planning a potential evaluation of naming and shaming interventions for sex offenders. No doubt there are other possible mechanisms at work that will be discovered as the evaluation process unfolds. Nevertheless, as it currently stands, our preliminary chain of inter-linked mechanism-based propositions can be immediately useful in several ways.

First, it places the evaluator in a position to start focusing the evaluation design by identifying questions and data collection methods to “test” the way in which the theory works (or fails to work) in practice. For example, situational mechanisms, such as diffusion processes are identified in the policy-maker’s theory as an important channel for delivering information to local residents. One way of systematically investigating diffusion would be through a survey approach that considered the extent to which information is actually getting to individuals and community groups in a timely fashion. A local case study analysis of communication networks might also be useful for addressing the question of whether information is actually getting to the right people in the right way. If this is not the case, then unintended side effects such as vigilantism may occur. In both scenarios, data collection efforts are guided by theory, are more focused and address whether the diffusion of information mechanism is working as intended. If not, then the causal chain breaks down and outcomes such as reductions in re-offending will not be achieved.

Second, tapping into existing social science theory relating to mechanisms is often useful for policy makers and evaluator as it may provide important insights into the likely functioning of mechanisms. Consider again the example of situational mechanisms, such as diffusion, where there is a readily available and vigorous body of scholarly research (Rogers, 2003). One important finding is that the type of interpersonal network structure can influence information exchange greatly. Radial personal networks comprise a set of individuals linked to a focal individual but not interacting with each other, while interlocking personal networks consist of individuals who interact mainly with each other. The former would seem to be much more effective in ensuring that information about local sex offenders is exchanged with a wider environment, thereby increasing surveillance

behavior. However, in some communities, this may lead to vigilantism and tighter controls over information networks and communication channels may be required.

A third way in which a more explicit focus on mechanisms in evaluation can be useful is that it helps contribute to knowledge development about social programming more generally. Mechanisms are often “portable” in the sense that they are building blocks for middle-range program theories, which may be transferable to different contexts and policy domains. For example, Pawson (2006) shows that sex offender naming and shaming interventions are not really that unique. The basic idea is also used by policy makers in relation to school league tables and safety indices for car manufacturers, among many others. This suggests that although the precise details and context are likely to vary across different policy situations, this does not necessarily mean that we need to start anew every time a “naming and shaming-type” intervention is launched. Instead of treating all interventions as completely novel, it is our hope that building a knowledge and theoretical base about “families of interventions” (Pawson, 2006), including the different types of mechanisms that underlie them may over time reduce “policy amnesia” and constant reinvention of the wheel.

### **Using Mechanisms in Evaluation: Some Further Considerations**

Although we would argue that evaluators almost always need knowledge about social and behavioral mechanisms to substantiate causal claims, this does not necessarily mean that all evaluations should deal with mechanisms. Certainly, given the potential demand on time and resources, careful consideration must be given to the nature of the evaluand, as well as whether stakeholders actually desire to know how and why the program works. This will provide some clues as to whether an explanatory focus is appropriate. There is nothing inherently wrong with an evaluator deciding to treat a program as a black box, if the purpose of the evaluation is primarily about judging merit, worth, or significance. Sometimes “black box” evaluation is necessary “when no theory [or knowledge of mechanisms] is available or needed, or when getting the relevant expertise would take resources of time or money we do not have” (Scriven, 1998, p. 59). As Scriven might say, even more frankly, “You don’t want to waste all your time and resources fishing for mechanisms, when it is not appropriate to be doing so”.

If the circumstances are conducive to an explanatory approach, and an evaluator is fortunate enough to have the luxury of “fishing” for mechanisms, then it is important to be aware of some additional precautions. When investigating mechanisms that might plausibly account for program outcomes, it is important not to be “mechanistic.” There are no set procedures to be followed rigidly, no columns, rows, or logic model boxes to be filled with generic examples of program “inputs,” “outputs,” or “activities.” The evaluator is not a “box-filling” administrator (Gasper, 2000).

Instead, the evaluator is an applied theorist,<sup>14</sup> who draws on a range of social and behavioral theories to combine diverse ideas in imaginative ways. Weber (2006) summarizes this point well in the context of organizational studies, although it might equally apply to the field of evaluation. He warns that:

In the quest to identify mechanisms and to assemble them into causal models, researchers end up with theory that is mechanistic and neither interesting nor generative. The temptation is to focus too much on input-output relationships, on linear chains of causality, and on building tightly knit models of arrows and boxes . . . . This would be a dangerous path to take as the complexity and situatedness of much organisational activity begs for a style of theory that preserves some ambiguity (p. 120).

Program theory building with mechanisms involves constant shuttling between theory and empirical data, using both inductive and deductive reasoning. Realist philosophers and evaluators sometimes

refer to this particular logic of inquiry as “retroduction” or “principled discovery” (Bhaskar, 1975; Mark et al. 2000).

It is also important that theory-driven evaluators do not replace substantive social and behavioral science theory with a focus only on putative mechanisms. Again, Weber (2006) is instructive here when he cautions against the temptation to avoid serious theorizing by “plug[ging] and play[ing] with a few handy mechanisms to explain phenomena of interest” (p. 120). Mechanisms “are usually specified in relation to and often only make sense as part of a larger body of theory. They elaborate, sharpen, transpose, and connect theories, but they do not substitute for them” (Weber, 2006, p. 120).

## Concluding Remarks

Social programs and policies are theories incarnate and a focus on mechanisms in program theory has much to offer the field of evaluation. Theorizing with mechanisms strengthens our understanding of how and why programs work, with whom, and under what circumstances. This is an adage of realist, theory-driven evaluation with which we strongly agree. Although the argument that mechanisms are important for evaluation is not necessarily new, it is one that needs greater emphasis in contemporary evaluation practice.

This is because much of what passes as theory-driven evaluation today seems to pay scant attention to mechanisms (especially in terms of the way we have conceptualized mechanisms in this article). Often, evaluators unpack a programmatic “black box” by laying out the components of the evaluand and then order them into some logical sequence. That is nothing more than dissecting the “operational logic” but not the “conceptual logic,” as the U.S. General Accounting Office (US GAO, 1986) once called it. The “operational logic,” as we contend, does not constitute an explanatory theory of a program.

This argument has theoretical and practical implications for evaluators. First, a more explicit focus on underlying generative mechanisms might help to counter what appears to be a growing trend toward oversimplified versions of program theory in the form of linear logic models (Gasper, 2000; Rogers, 2007; Weiss, 1997a). The realist notion of explanatory mechanisms is useful for evaluation practitioners who are seeking to complement and extend the way in which they currently develop and use program logic models in evaluation. Evaluators who use primarily quantitative techniques such as path analysis and causal modeling to articulate and test moderators and mediators in program theory might also derive greater explanatory power from their studies if they allow for a distinction between statistical and theoretical conceptions of mechanisms.

Second, while we are optimistic that greater attention to mechanisms will be beneficial for evaluation, it also carries with it some challenges, such as the potential to inadvertently confuse the term “mechanism” with program activity or variable. There is also the danger that mechanism-based explanation becomes associated with a “machine” imagery of social programs or evaluators treat mechanisms as if they are stand-alone little theories that replace the need to engage with substantive social science theory. In this article, we have attempted to address these challenges by reinforcing the argument that mechanism are not a to-do list of program activities nor are they variables that can easily be entered into a statistical regression model. When applied to social policies and programs, mechanisms are the underlying processes or “hidden causal levers” that account for how and why a program works to bring about desired changes in the reasoning and behavior of participants.

Third, mechanism-based theorizing may also help to stimulate and guide much needed research on evaluation. This is because evaluation activities themselves also seem to “work” by triggering particular mechanisms in particular contexts. For example, Mark and Henry (2004) identify how evaluation use, or more precisely evaluation influence, is brought about through the interaction of

a range of social and behavioral mechanisms. Mark and Henry argue that their mechanism-based theory of evaluation influence has the potential to open up new areas of inquiry by elucidating how and why evaluation works to bring about desired change in a more precise way. Furthermore, identification and clarification of the mechanisms underling evaluation influence may also provide better guidance for evaluation practitioners who are seeking to improve the influence of evaluation.

In some circumstances, evaluations can also trigger mechanisms that lead to unintended side effects. One example is the mechanism of “designed blindness.” It illustrates that stakeholders and evaluators can sometimes become such a strong believer in the program theory they articulate and test that empirical findings are largely, or only, framed as corroborations of this theory (a specific kind of tunnel vision; see Friedman, 2001). Campbell’s Law, which relates to the mechanism of “corruptibility of social indicators”<sup>15</sup> (Campbell, 1979), also shows that in some situations, a too strong belief in the relevance of indicators and performance measures and their contribution to “enlightenment” can produce undesired effects such as gaming and goal displacement. Of course, these are just some of the potentially harmful but largely unnoticed mechanisms that lie beneath evaluation activities. Therefore, we anticipate that further research on mechanisms underlying evaluation itself would also be extremely worthwhile.

Finally, getting more involved in (explanatory) theories about mechanisms adds value to the evaluation enterprise because it helps avoid the problem of one-off, discrete evaluations that do little to develop generalizable knowledge about social programming. As policy makers and program officers are not always “au courant” about the mechanisms they assume to be at work, evaluators can demonstrate that although the specific program or intervention they have designed and implemented looks different from other interventions, in fact the same underlying mechanisms are called upon to make the policy work. By sharing and using the accumulated evidence on the level of the *mechanisms at work* (instead of the specific intervention as such), policy makers and evaluator may come to realize that many supposedly “novel” interventions share common underlying mechanisms of change. Knowledge of these mechanisms could then be used to better inform the design and evaluation of social policies and programs.

## Notes

1. This useful way of distinguishing between program logic and program theory was suggested to the first author a number of years ago by Dr. Gerald Elsworth.
2. Although the importance of “mechanisms” was discussed at length in an earlier book by Pawson (1989), see also an article by Pawson and Tilley (1994) in the *British Journal of Criminology* where the scientific realist approach to evaluation was foreshadowed.
3. Gasper (2000) has pointed out a number of concerns regarding the overly simplistic application of the “logical framework approach” (or LogFrames) in international development evaluation. He refers, for example, to problems such as “logic-less frames” where prescriptive templates are used leading to an illusion of logic, “lack-frames,” which omit critical aspects of a program, and “lock-frames,” which restrict program learning and adaptation.
4. For example, Gerring (2007) suggests that in contemporary social science literature there are at least nine distinct, but sometimes contradictory and overlapping, meanings of the term mechanism.
5. Missing from Mahoney’s (2003) list of definitions is an important description of mechanisms by the realist philosopher Roy Bhaskar. According to Bhaskar (1975):  
“The world consists of mechanisms not events. Such mechanisms combine to generate the flux of phenomena that constitute the actual states and happenings of the world. They may be said to be real, though it is rarely that they are actually manifest and rarer still that they are empirically identified by men. They are the intransitive objects of scientific theory. They are quite independent of men—as thinkers, causal agents, and perceivers. They are not unknowable, although knowledge of them depends upon a rare blending of

intellectual, practico-technical and perceptual skills. They are not artificial constructs. But neither are they Platonic forms. For they can become manifest to men in experience. Thus, we are not imprisoned in caves, either of our own or of nature's making. We are not doomed to ignorance. But neither are we spontaneously free. This is the arduous task of science: the production of the knowledge of those enduring and continually active mechanisms of nature that produce the phenomena of our world" (p. 47).

6. This distinction is somewhat similar to the difference between a "hypothetical construct" and "intervening variable" made by MacCorquodale and Meehl in their seminal 1948 paper.
7. Others mechanisms might include the constraining effects of "cultural capital" (Bourdieu & Passeron, 1977) and "rational choice" theories of educational inequality (Boudon, 1974).
8. This is, perhaps, what Mahoney (2003) is referring to when he describes mechanisms as "ultimate causes" that do not require explanation themselves.
9. Benton (1981/1998) offers a useful distinction between mechanisms in the "natural world" and mechanisms in the "human world." In the former, mechanisms are "person independent," whereas in the latter, mechanisms are "person dependent."
10. This point relates to questions about the validity of the observation theory that is used when testing a policy theory or an intervention theory. Popper (1972) and Lakatos (1980) have shown that data collection is based on (sometimes) implicit observation theories that can differ in breadth, depth, and width, as is also the case with substantive (i.e., policy) theories. Detecting psychopathic behavior using only a questionnaire is, for example, probably based on an observation theory that is less advanced compared to trying to detect that kind of behavior using magnetic resonance imaging/positron emission tomography (MRI/PET) scans *and* questionnaires. Of course, much more could be said about the relationship between theory and observation, but this would require a lengthy treatment, which is beyond the scope of this article.
11. The "slippery slope" mechanism refers to the proposition or argument that a seemingly small first step inevitably leads to a chain of related events culminating in some significant impact. This is analogous to giving a ball a small push on the edge of a downward slope. In economic theory, the "endowment" mechanism proposes that people often demand much more to give up something than they would be willing to pay to acquire it. The "framing effect" is familiar to opinion researchers who find that presenting the same option in different formats can alter people's decisions. For example, individual inconsistencies have been found depending on whether a question is framed in a negative or positive way.
12. Festinger's cognitive dissonance theory and his social comparison theory (Festinger, 1954, 1957) point at (sociocognitive) mechanisms, whereas Merton's (1973) work on the sociology of science, for example, points at the Matthew effect (explaining how eminent scientists will often get more credit than a comparatively unknown researcher even if their work is similar). Olson's (1971) theory about why large groups do not contribute voluntarily to the production of collective goods is another famous example, as is work by Granovetter (1973) on the strength of weak ties.
13. Essentially, this typology is a categorization of mechanisms according to the way in which they interact at different strata of social reality or levels of analysis. For example, the connection between two macro-level conditions (like the example of low parental income and poor educational outcomes for children) can be explained best by taking into account how macro-phenomena influence the beliefs and actions of individual actors, who over time interact and generate new macro states. As Hedström and Swedberg (1998) explain: This way of conceptualising social action lends itself in a very natural way to a typology of mechanisms . . . instead of analysing relationships between phenomena exclusively on the macro level, one should always try to establish how macro-level events or conditions affect the individual, how the individual assimilates the impact of these macro-level events, and how a number of individuals through their actions and interactions, generate macro-level outcomes (p. 21–22).
14. The notion of an evaluator being an "applied theorist" is appealing to us and was first coined by Charles McClintock (1990) in an article describing the use of program theory as part of an evaluation of a local hospice agency.

15. Campbell's law, which was originally proposed in an occasional paper series in 1976, states that "the more any quantitative indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell, 1979, p. 35). Some historical examples of indicators susceptible to these pressures, as cited by Campbell, include "voting statistics," police crime "clearance rates," enemy "body counts" in the Vietnam War, and productivity indicators for factories. A contemporary example would be standardised educational testing in classrooms.

### Declaration of Conflicting Interests

The author(s) declared no conflicts of interest with respect to the authorship and/or publication of this article.

### Funding

The author(s) received no financial support for the research and/or authorship of this article.

### References

- Anderson, P. J. J., Blatt, R., Christianson, M. S., Grant, A. M., Marquis, C., Neuman, E. J., & . . . Sutcliffe, K. M. (2006). Understanding mechanisms in organisational research: Reflections from a collective journey. *Journal of Management Inquiry, 15*, 102-113.
- Benton, T. (1981/1998). Realism and social science: Some comments on Roy Bhaskar's 'The possibility of naturalism.' In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical realism: Essential readings* (pp. 297-312). London, England: Routledge.
- Bickman, L. (1987). The importance of program theory. In L. Bickman (Ed.) *Using program theory in evaluation. New Directions for Program Evaluation* (No. 33, pp. 5-18). San Francisco, CA: Jossey-Bass.
- Birckmayer, J. D., & Weiss, C. H. (2000). Theory-based evaluation in practice: What do we learn? *Evaluation Review, 24*, 407-431.
- Bhaskar, R. (1975). *A realist theory of science*. Sussex, Brighton: Harvester Press.
- Boudon, R. (1974). *Education, opportunity and social inequality*. New York, NY: John Wiley.
- Bourdieu, P., & Passeron, J. C. (1977). *Reproduction in education, society and culture*. London, England: SAGE.
- Bunge, M. (1997). Mechanism and explanation. *Philosophy of the Social Sciences, 27*, 410-465.
- Bunge, M. (2004). How does it work? The search for explanatory mechanisms. *Philosophy of the Social Sciences, 34*, 182-210.
- Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning, 2*, 67-90.
- Carter, B., & New, C. (Eds.). (2004). *Making realism work: Realist social theory and empirical work*. London, England: Routledge.
- Chen, H.-T. (1989). The conceptual framework of the theory-driven perspective. *Evaluation and Program Planning, 12*, 391-396.
- Chen, H.-T. (1990). *Theory-driven evaluations*. Newbury Park, CA: SAGE.
- Chen, H.-T. (2005). *Practical program evaluation*. Thousand Oaks, CA: SAGE.
- Chen, H.-T. & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10*, 95-103.
- Clarke, R. V. (1992). Introduction. In R. V. Clarke (Ed.), *Situational crime prevention: Successful case studies* (pp. 2-41). Albany, NY: Harrow and Heston.
- Coleman, J. S. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology, 91*, 1309-1335.
- Coleman, J. S. (1990). *The foundations of social theory*. Cambridge, MA: The Belknap Press of Harvard University.



- Connell, J., Kubish, A., Schorr, L., & Weiss, C. (1995). *New approaches to evaluating community initiatives: Concepts, methods, and contexts*. Washington, DC: Aspen Institute.
- Davidson, E. J. (2000). Ascertaining causality in theory-based evaluation. In P. J. Rogers, T. A. Hasci, A. Petrosino, & T. A. Huebner (Eds.), *Program theory in evaluation: Challenges and opportunities. New Directions for Evaluation* (No. 87). San Francisco, CA: Jossey-Bass.
- Davidson, E. J. (2006). Editorial: The “bagging” of theory-based evaluation. *Journal of Multidisciplinary Evaluation*, 3, iii-xiii.
- Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. Mahwah, NJ: Erlbaum.
- Donaldson, S. I., & Lipsey, M. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. Shaw, J. Greene, & M. Mark (Eds.), *The Sage handbook of evaluation* (pp. 56-75). Thousand Oaks, CA: SAGE.
- Elster, J. (1989). *Nuts and bolts for the social sciences*. Cambridge, UK: Cambridge University Press.
- Elster, J. (2007). *Explaining social behaviour: More nuts and bolts for the social sciences*. Cambridge, UK: Cambridge University Press.
- Farnsworth, W. (2007). *The legal analyst: A toolkit for thinking about the law*. Chicago, UK: University of Chicago Press.
- Festinger, L. (1950). Informal social communication. *Psychological Review*, 57, 271-282.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson.
- Friedman, V. J. (2001). Designed blindness: An action science perspective on program theory evaluation. *American Journal of Evaluation*, 22, 161-181.
- Gasper, D. (2000). Evaluating the ‘logical framework approach’: Towards learning-oriented development evaluation. *Public Administration and Development*, 20, 17-28.
- George, A. L., & Bennett, A. (2004). *Case studies and theory development in the social sciences*. Cambridge, MA: MIT Press.
- Gerring, J. (2007). The mechanistic worldview: Thinking inside the box. *British Journal of Political Science*, 38, 161-179.
- Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78, 1360-1380.
- Harré, R. (1972). *The philosophies of science: An introductory survey*. London, England: Oxford University Press.
- Hedström, P. (2005). *Dissecting the social: On the principles of analytical sociology*. Cambridge, UK: Cambridge University Press.
- Hedström, P., & Swedberg, R. (Eds.). (1998). *Social mechanisms: An analytical approach to social theory*. Cambridge, UK: Cambridge University Press.
- Henry, G. T., Julnes, G., & Mark, M. M. (Eds.). (1998). *Realist evaluation: An emerging theory in support of practice. New Directions for Program Evaluation* (No. 78). San Francisco, CA: Jossey-Bass.
- House, E. R. (1991). Realism in research. *Educational Researcher*, 20, 2-9, 25.
- Lakatos, I. (1980). *The methodology of scientific research programmes*. Cambridge, UK: Cambridge University Press.
- Lawson, T. (1997). *Economics and reality*. London, England: Routledge.
- Leeuw, F. L. (1991). Policy theories, knowledge utilization, and evaluation. *Knowledge and Policy*, 4, 73-92.
- Leeuw, F. L. (2003). Reconstructing program theories: Methods available and problems to be solved. *American Journal of Evaluation*, 24, 5-20.
- Lipsey, M. (1993). Theory as method: Small theories of treatments. In L. B. Sechrest, & A. G. Scott (Eds.), *Understanding causes and generalizing about them. New Directions for Program Evaluation* (No. 57, pp. 5-38): San Francisco, CA: Jossey-Bass.
- Little, D. (1991). *Varieties of social explanation: An introduction to the philosophy of social science*. Boulder, CO: Westview Press.

- MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review*, 55, 95-107.
- Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67, 1-25.
- Mahoney, J. (2001). Beyond correlational analysis: Recent innovations in theory and method. *Sociological Forum*, 16, 575-593.
- Mahoney, J. (2003, August 28). *Tentative answers to questions about causal mechanisms*. Paper presented at the annual meeting of the American Political Science Association, Philadelphia, PA. Retrieved October 7, 2008, from [http://www.allacademic.com/meta/p62766\\_index.html](http://www.allacademic.com/meta/p62766_index.html)
- Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation*, 10, 35-57.
- Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey-Bass.
- Mayntz, R. (2004). Mechanisms in the analysis of social macro-phenomena. *Philosophy of the Social Sciences*, 34, 237-259.
- McClintock, C. (1990). Evaluators as applied theorists. *Evaluation Practice*, 11, 1-12.
- McCombs, M. E., & Shaw, D. L. (1972). The agenda setting function of mass media. *Public Opinion Quarterly*, 36, 176-187.
- McLaughlin, J. A., & Jordan, G. G. (1999). Logic models: A tool for telling your program's performance story. *Evaluation and Program Planning*, 22, 65-72.
- Merton, R. K. (1968). *Social theory and social structure* (3rd ed.). New York, NY: Free Press.
- Morén, S., & Blom, B. (2003). Explaining human change: On generative mechanisms in social work practice. *Journal of Critical Realism*, 2, 37-60.
- Olson, M. (1971). *The logic of collective action: Public goods and the theory of groups*. Harvard, UK: Harvard University Press.
- Pawson, R. (1989). *A measure for measures: A manifesto for empirical sociology*. London, England: Routledge.
- Pawson, R. (2000). Middle-range realism. *Archives Européennes de Sociologie*, 41, 283-325.
- Pawson, R. (2006). *Evidence-based policy: A realist perspective*. London, England: SAGE.
- Pawson, R. (2008). Invisible mechanisms. *Evaluation Journal of Australasia*, 8(2), 3-13.
- Pawson, R. (2010). Middle range theory and program theory evaluation: From provenance to practice. In J. Vaessen & F. L. Leeuw (Eds.), *Mind the gap: Perspectives on policy evaluation and the social sciences. Comparative Policy Evaluation* (Vol. 16, pp. 171-203). London, England: Transaction Publishers.
- Pawson, R., & Tilley, N. (1994). What works in evaluation research? *British Journal of Criminology*, 34, 291-306.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, England: SAGE.
- Popper, K. R. (1972). *Conjectures and refutations: The growth of scientific knowledge*. London, England: Routledge and Kegan Paul.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.
- Rogers, P. J. (2007). Theory-based evaluation: Reflections ten years on. In S. Mathison (Ed.), *New Directions for Program Evaluation* (No. 114, pp. 63-67). San Francisco, CA: Jossey-Bass.
- Rogers, P. J., Hacsí, T. A., Petrosino, A., & Huebner, T. A. (Eds.) (2000). *Program theory in evaluation: Challenges and opportunities. New Directions for Evaluation* (No. 87). San Francisco, CA: Jossey-Bass.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom*. New York, NY: Holt, Rinehart and Winston, Inc.
- Sayer, A. (1981/1998). Abstraction: A realist interpretation. In M. Archer, R. Bhaskar, A. Collier, T. Lawson, & A. Norrie (Eds.), *Critical realism: Essential readings* (pp. 120-143). London, England: Routledge.
- Sayer, A. (2000). *Realism and social science*. London, England: SAGE.
- Scheirer, M. A. (1987). Program theory and implementation theory: Implications for evaluators. In L. Bickman (Ed.), *Using program theory in evaluation. New Directions for Program Evaluation* (No. 33, pp. 59-76). San Francisco, CA: Jossey-Bass.

- Schelling, T. C. (1998). Social mechanisms and social dynamics. In P. Hedström & R. Swedberg (Eds.), *Social mechanisms: An analytical approach to social theory* (pp. 32-44). New York, NY: Cambridge University Press.
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49, 103-122.
- Scriven, M. (1994). The fine line between evaluation and explanation. *Evaluation Practice*, 15, 75-77.
- Scriven, M. (1998). Minimalist theory: The least theory that practice requires. *American Journal of Evaluation*, 19, 57-70.
- Smith, N. L. (1990). Using path analysis to develop and evaluate program theory and impact. In L. Bickman (Ed.), *Using program theory in evaluation. New Directions for Program Evaluation* (No. 33, pp. 53-57): San Francisco, CA: Jossey-Bass.
- Steele, D. (2004). Social mechanisms and causal inference. *Philosophy of the Social Sciences*, 34, 55-78.
- Stinchcombe, A. L. (1991). The conditions of fruitfulness of theorizing about mechanisms in social science. *Philosophy of the Social Sciences*, 21, 367-388.
- Suchman, E. (1967). *Evaluative research: Principles and practice in public service and social programs*. New York, NY: Russell Sage Foundation.
- Tilly, C. (2001). Mechanisms in political processes. *Annual Review of Political Science*, 4, 21-41.
- Tilley, N. (1995). Seeing off the danger: Threat, surveillance and modes of protection. *European Journal on Criminal Policy and Research*, 3, 27-40.
- Tilley, N. (2004). Applying theory-driven evaluation to the British Crime Reduction Programme. *Criminal Justice*, 4, 255-276.
- Trochim, W. M. K. (1989). Outcome pattern matching and program theory. *Evaluation and Program Planning*, 12, 355-366.
- United Way of America. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: Author.
- U.S. General Accounting Office. (1986). *Teenage pregnancy: 500,000 births a year but few tested programs*. Washington, DC: Author.
- Weber, K. (2006). From nuts and bolts to toolkits: Theorizing with mechanisms. *Journal of Management Inquiry*, 15, 119-123.
- Weiss, C. H. (1972). *Evaluation*. Upper Saddle River, NJ: Prentice Hall.
- Weiss, C. H. (1995). Nothing as practical as good theory. In J. Connell, A. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New approaches to evaluating community initiatives*. New York, NY: Aspen Institute.
- Weiss, C. H. (1997a). Theory-based evaluation: Past, present, and future. *New directions for evaluation*. (No. 76). San Francisco, CA: Jossey-Bass.
- Weiss, C. H. (1997b). How can theory-based evaluation make greater headway? *Evaluation Review*, 21, 501-524.
- Wholey, J. S. (1979). *Evaluation: Promise and performance*. Washington, DC: Urban Institute.
- Wholey, J. S. (1983). *Evaluation and effective public management*. Boston, MA: Little, Brown.
- Wholey, J. S. (1987). Evaluability assessment: Developing program theory. In L. Bickman (Ed.), *Using program theory in evaluation. New Directions for Program Evaluation Theory* (No. 33, pp. 77-92). San Francisco, CA: Jossey-Bass.
- W. K. Kellogg Foundation. (2004). *Logic model development guide*. Battle Creek, MI: Author.
- Wortley, R. (1996). Guilt, shame and situational crime prevention. In R. Homel (Ed.), *The politics and practice of situational crime prevention* (pp. 115-132). New York, NY: Criminal Justice Press.