Theory-Based Evaluation Meets Ambiguity

Dahler-Larsen, Peter

Published in: American Journal of Evaluation

DOI: 10.1177/1098214017716325

Publication date: 2018

Document version: Peer reviewed version

Citation for published version (APA):
Theory-Based Evaluation Meets Ambiguity: The Role of Janus Variables

Peter Dahler-Larsen

Abstract

As theory-based evaluation (TBE) engages in situations where multiple stakeholders help develop complex program theory about dynamic phenomena in politically contested settings, it becomes difficult to develop and use program theory without ambiguity. The purpose of this article is to explore ambiguity as a fruitful perspective that helps TBE face current challenges. Literatures in organization theory and political theory are consulted in order to cultivate the concept of ambiguity. Janus variables (which work in two ways) and other ambiguous aspects of program theories are classified and exemplified. Stances towards ambiguity are considered, as are concrete steps that TBE evaluators can take to identify and deal with ambiguity in TBE.

Keywords
theory-based evaluation, program theory, participatory evaluation, evaluation use, ambiguity

Janus is a composite and obscure Roman god who is associated with beginnings, doorways, gates, choices, and time. He is usually depicted with two faces looking in opposite directions. In his hand, he holds a key.

Excerpts from Encyclopedia Mythica

Theory-based evaluation (TBE) is an important and long-standing school of thought in evaluation (Bickman, 1987; Chen, 1990). It is acknowledged in leading textbooks and histories of evaluation (Alkin, 2004; Fitzpatrick, Sanders, & Worthen, 2004; Rossi, Lipsey, & Freeman, 2004; Shadish, Cook, & Leviton, 1991). It has gained increasing interest over time (Chen, 2005; Donaldson, 2007; Rogers, 2008; Weiss, 1997a) in new policy areas and in several continents (Rogers, 2007; Stame, 2013). The vitality of TBE is demonstrated through innovative revisions of earlier models both inside and outside of the TBE tradition (Funnell & Rogers, 2011; Mark, Henry, & Julnes, 2000; Mayne, 2011; Pawson & Tilley, 1997).

1 Department of Political Science, University of Copenhagen, København, Denmark

Corresponding Author:
Peter Dahler-Larsen, Department of Political Science, University of Copenhagen, Øster Farimagsgade 5, 1353 København K, Denmark.
Email: pdl@ifs.ku.dk
In these endeavors, the theory–stakeholder interface plays a crucial role. After consulting with stakeholders, the evaluator usually formulates a unified program theory that will be consensually accepted as a framework for the evaluation. The effectiveness of TBE is assumed to hinge on consensus around “the program theory” (in the singular), but this term is not always helpful. Especially for neophytes in TBE, it leads to a futile search for the theory that represents program reality in a thing-like, objective way modeled after laws in natural science. The conventional design of the TBE process helps bury or conceal ambiguities that might otherwise lead to interesting and productive heuristic insights. By ambiguity, I mean the coexistence of multiple interpretations of a phenomenon among reasonable people while there is not necessarily an easy way to choose between the interpretations or eliminate some of them.

Alternatively, program theory can be viewed as a preliminary and pragmatic tool embedded in social interaction among stakeholders. Theory is a dynamic, interpretable, and thus inherently ambiguous phenomenon. Although boundary conditions of TBE have been found resting with “lack of stakeholder cooperation,” “disagreements about program theory,” and “highly dynamic programs and program theories” (Donaldson & Lipsey, 2006, p. 66), TBE is already pushing the envelope as more stakeholders become deeper involved in coproducing more complicated and complex program theories and as TBE expands into several forms of use. As a result, encounters with ambiguity become more likely.

Without recognition of ambiguity and of the politically charged nature of problem definitions, program theories, and outcomes, TBE runs the risk of becoming overly technical. Better cooperation may follow if stakeholders make sense of each other’s program theories and how they interconnect, and by embracing ambiguity, evaluators can help promote collective sense-making about complex interventions as part of a democratic evaluation process.

This article explores fruitful implications of the role of ambiguity in TBE and argues for increased sensitivity of practitioners and students of TBE to ambiguity. I also offer a simple but innovative way of illustrating program theory so that ambiguity becomes visible. A key notion is what I have come to call a Janus variable. A Janus variable is a phenomenon that plays two (or more) roles in a program-theoretical model. It can be operationally identified when different stakeholders talk about the same phenomenon, so that their different causal understandings interfere with each other or even undermine each other’s consequences. I propose to display various stakeholder theories or partial theories overlaid in the same graphical model in a way that makes Janus variables appear visually.

The core of the article focuses on theory-making. Program theorizing is crucial, because it reflects understandings of the roots of social problems, it proposes how these problems can be ameliorated, and it plays a key role in the structuration of the entire TBE process. There are additional forms of ambiguity that may not be directly portrayed as part of program theory itself but are important ingredients in the sociopolitical context in which the evaluation process unfolds. Evaluators can be better pathfinders in this landscape if they recognize various forms of ambiguity.

First, I consult the literature on ambiguity. I show that the design of social processes where information is processed influences capacity to deal with ambiguity and facilitate sense-making. Second, I describe how recent trends in TBE (more complex theory, more participation, and more forms of use) send TBE into more ambiguous terrain, while the conventional TBE procedure lacks capacity to adequately embrace the full consequence hereof. Third, I propose a specific technique TBE evaluators can use to visually highlight ambiguity, in particular Janus variables. Fourth, I situate the use of this technique in a broader typology of forms of ambiguity and finally discuss conditions under which evaluators should pay particular attention to ambiguity in TBE.

**Ambiguity**

Ambiguity is discussed in philosophy, political theory, and organization theory. Existentially, ambiguity is indicative of our *thrownness*, a concept introduced by German philosopher Martin
Heidegger (1889–1976) to describe individual existences as “being thrown” into a world of collective and contested construction. As long as the constitution of meaning is an unfinished collective human undertaking (Best, 2008; C. Taylor, 2016), ambiguity is not going away.

Ambiguity is conceptually related to, but not the same as uncertainty (Best, 2008), contestedness (Gallie, 1956), and indeterminacy (Freeden, 2005). Whereas some ambiguity (more than one reading of a practice, image, or text) may be dealt with through disambiguation (an explanation that aims at removing all meanings but one; Freeden, 2005), only that part which has to do with uncertainty can be eliminated through clear communication. The remaining part of ambiguity has to do with contestedness. “Essentially contested” ideas such as democracy, equality, and quality continue to embody value conflicts and competing interests (Gallie, 1956).

Being a citizen requires learning to live with ambiguity and paradox (Stone, 2012). Some vagueness and ambiguity is fundamental for political coexistence (Freeden, 2005). For example, a sufficient level of ambiguity in the articulation of principles and goals is necessary in order to gain support for a proposal from a majority of otherwise diverse constituencies (Baier, March, & Sætren, 1986). Goals that signify virtue and humanity help justify political support for social interventions even if when problems cannot be solved (Hasenfeld, 1982). For these reasons, ambiguity is sometimes, but not always, a result of deliberately strategic communication.

Ambiguity is “channeled” for some time when a larger principle, value, or goal is specified, operationalized, and perhaps measured (Simon, 1996). However, not all of the components of rich concepts can be contained in any given formulation (Freeden, 2005). At some point in time, dominant definitions and interpretations become challenged, thus enhancing visible ambiguity. Political phenomena oscillate between the ambiguous and the disambiguated (Warren, 1999), between sweeping ambiguity under the carpet and exposing it to daylight.

Organization theory helps us understand how procedures, structures, and standard operating procedures (SOPs) help people deal with ambiguity. For example, scripts are routinized ways of organizing information in the form of charts, models, tables, indicators, and conceptual frameworks. Scripts help channel otherwise unstructured amounts of information into something interpretable within organizational frames of reference. By these means, organizations process particular versions of their environments upon which they act. Ambiguity does not disappear ontologically; it just becomes manageable as part of an organized process (March & Olsen, 1976). While existing SOPs are often formalistic, they are in fact a result of how the organization has dealt with its past experiences. SOPs are frozen learning.

An alternative school of thought focuses on an ongoing form of information processing called sense-making. Sense-making is about interpreting what is going on starting here and now (Weick, Sutcliffe, & Obstfeld, 2005). Sense-making is local and pragmatic. It hinges on diverse identities, experiences, and relevance structures. Relevance structures are expectations about how to act and pressures to do so that guides attention in particular directions (Berger & Kellner, 1981).

Sense-making is particularly intense when there is a pressure to act and available information is unusual, contradictory, or somehow sticks out from what could be expected, in other words, something ambiguous. Sense-making does not necessarily remove ambiguity but deals with it until further notice. Sense-making is “a way station” (Weick et al., 2005, p. 409) that takes place on the condition of throwness into an ongoing, unknowable, unpredictable stream of experience. For this reason, speech acts such as setting a goal, making a promise, and asking for information can only be understood as related to sense-making situated in a here-and-now perspective. When a user of information is asked to help design information that one may or may not need in the future, the individual in focus has no experiential access to the future situation, so the clues must come from the present situation. Situated usefulness is a distributed and dynamic phenomenon (Weick et al., 2005). Goals and promises are
reinterpreted retrospectively. An image of shared meanings may be misleading if it is not connected with ongoing experiential sense-making.

Faced with time pressure and ambiguity, organizations paradoxically revert to strategies that are known not to handle ambiguity well. Organizations centralize decisions, rely on hierarchy, and assume consensus without paying sufficient attention to ambiguous details. Instead, according to a sensemaking perspective, those who have crucial firsthand experiential insights should be actively and continuously involved, informally and immediately, especially because early sense-making of unusual information helps prevent later crises (Weick et al., 2005).

While a focus on organizational procedures and scripts highlights the role of formal organization in information processing, theories of sense-making emphasize the informal, more distributed, and less consensual ways of interpretation. These diverging approaches to ambiguity are not mutually exclusive. Any form of formal procedure channels some forms of ambiguity. The two perspectives are complementary and can both be seen as contributions to organizational learning (Levinthal & Rerup, 2006). Both unfold under restrictions concerning time, place, and attention.

I next look at TBE as a form of information processing that takes place in the tension between procedures and scripts on the one hand and sense-making on the other. I argue that TBE already traverses ambiguous territory, but the conventional mode of organizing the TBE process does not provide adequate space for highlighting ambiguity and making sense of it.

The Tradition of TBE and the Road Toward Ambiguity

I shall focus on how TBE is caught between the conventional SOP for theory-making and new challenges regarding ambiguity. Since systematic reviews of the development of TBE have already been carried out (Coryn, Noakes, Westine, & Schröter, 2011), I recognize that the following brief account does not do sufficient justice to the many advances in TBE (Chen, 1990; Donaldson & Lipsey, 2006; Funnell & Rogers, 2011; Leeuw, 2003; Mayne, 2011; Pawson & Tilley, 1997; Rogers, Petrosino, Huebner, & Hacsi, 2000; Stame, 2004; Vedung, 1997; Weiss, 1997a, 1997b), not to mention critics of TBE (Stufflebeam, 2001).

A word of clarification: No rigorous distinction is possible between theory-driven evaluation (science), TBE, theory-anchored evaluation, realistic evaluation, contribution analysis, evaluation based on a logic model, and a theory of change. I shall here refer pragmatically only to TBE as an umbrella term for all evaluation strategies and approaches where a core task is to construct and clarify a set of assumptions about why and how an intervention works explicitly in the form of a written or graphic description of a causal chain of events (the program theory).

The term program theory is habitual and should not be restricted to programs only because a similar logic can be applied in evaluations of policies, projects, and many kinds of interventions (Rogers, 2007). In a similar vein, the term “theory” has rough edges. A theory used in TBE may refer to something more or less explicit and articulate, more or less abstract and formal, and more or less stakeholder based versus anchored in general social science theory.

The program theory is logically connected to a “program field theory” that provides a causal understanding of the social problem that the intervention is intended to ameliorate (Vedung, 1997). The connection to a program field theory is necessary because the program theory is only effective if attacks (at least one of) the factors that help change the problem situation and circumvent causal obstacles. A model may comprise both program theory and program field theory. The term model has two meanings: a mental model and a graphic representation. (In the latter sense, one model can in principle comprise competing theories, in which case theory and “model” are not coterminous.)

Whether constructed as a part of an intervention or reconstructed for the purpose of the evaluation (Vedung, 1997), the program theory informs evaluation questions and helps “conceptualizing,
designing, conducting, interpreting, and applying an evaluation” (Coryn et al., 2011, p. 201). The centrality of program theory is reflected in how TBE evaluators allocate their attention. Much energy has been devoted to finding mechanisms and contexts (Pawson & Tilley, 1997) and making “better theories” (Weiss, 1997a). Perhaps more focus has been on the theory itself right (as an objectified tool) than on the social processes in which the articulation and use of program theory are embedded.

Some have classically viewed the delineation of a program theory as the product of an evaluation researcher situated on top of hierarchy of insights (Pawson & Tilley, 1997). According to this model, the evaluator invites various stakeholders to contribute bits and pieces to the program theory, but the ultimate responsibility for the program theory rests with the evaluator. Alternatively, a group of stakeholders who carry out their own evaluation can also ask an expert in TBE to help them develop their own program theories.

Most TBE takes place between these two extremes. Conflicting motivations to include stakeholders, but not too much, make participation in program theory-making gravitate toward what I call a conventional design of the TBE process. In this SOP, the evaluator has a main responsibility for delineating the program theory that best represents a converging consensus. Consensus around the unified program theory is important for acceptance of the subsequent evaluation. Notice how common it is in the field to talk about a program theory (or model) or the program theory (or model) in the singular as if one is the best number and as if theory and model are coterminous.

Although “competing theoretical perspectives make the empirical confirmation and/or disconfirmation (evaluation) process even more informative” (Donaldson, 2007, p. 224), a standard conception in the field is that the resulting program theory (in the singular) represents a “common understanding” among stakeholders (Balle Hansen & Vedung, 2010, p. 298), “an agreement that accords with their thinking” (Birckmayer & Weiss, 2000, p. 426), and a “common representation of the program that all parties can reference” (Donaldson & Lipsey, 2006, p. 65).

Most TBE evaluators recognize an intense time pressure on theory construction. Too much time spent on theory-making can “threaten the practicality” of TBE and “engender frustration or disengagement among stakeholders” (Funnell & Rogers, 2011, p. 144). Practitioners of TBE confirm that theory construction is time-consuming (Tronegård-Madsen, 2008). While it does not take a long time to draw a theory on paper, achievement of consensus is the real problem. As an SOP, the conventional design of the theory-making process is characterized by the need to condense the contributions from various stakeholders into a unified program theory in time for the remainder of the evaluation process to proceed. In doing so, it relies on an important script: a graphical model representing the unified program theory. As a result, ambiguity is compressed.

Prominent TBE evaluators have foreshadowed friction in the theory–stakeholder interface. Weiss (1997a) talked about “confusion” about components in program theories, potential “hyperrationalism,” a “long a frustrating process” (p. 507), and potential “indeterminacy” (p. 513). Friedman (2001) identified “gaps, inconsistencies, and conflicts,” “designed blindness,” and “defensive behaviors” in the making of program theory.

Three trends in TBE have sent this school of thought into ambiguous terrain. Program theory becomes less of an objective, reified, scientific tool, and more of a dynamic forum for contested collective processing of information as a consequence of (1) more complex program theories; (2) more forms of participation; and (3) more forms of use of TBE. In the following, I will show how these three movements have put the conventional design of the TBE process under pressure and revealed its lack of capacity to embrace ambiguity.

**Exploring More Complex Theory**

Early advocates of TBE found that even simple theories with a small number of logical steps would constitute a major advance (Weiss, 1997a). Weiss (1997b) recommended that TBE be primarily used
in case of “one or a few change efforts of moderate complexity” (p. 520). In Coryn, Noakes, Westine, and Schröter’s (2011) review, most program theories until 1998 were simple linear models. TBE evaluators need more complex models today because the interventions they evaluate are more complex and the environments in which these interventions unfold are large, global, diverse, turbulent, technologically complicated, and risky.

Program theories today include complicated (Funnell & Rogers, 2011) and complex (Patton, 2012) phenomena. As a consequence, the potential number of different roles and functions of phenomena represented in program theories and the number of types of connecting links have increased dramatically. Symptomatically, the illustrative simple model of program theory in Coryn et al. (2011) displays eight phenomena in two graphic forms with three similar arrows between them. Mathematically speaking, such elements can be combined in 48 permutations. The contrasting nonlinear model (Coryn et al., 2011, p. 202) shows 14 phenomena in three graphic forms connected by 17 arrows of four kinds, which corresponds to 2,856 permutations. Not all of these will be relevant in practice, but the rise in number of presentational possibilities is astonishing.

Complex models leave more discretionary space for interpretation. Complex models often play with graphic effects that may advance a visual grasp of the program theory but also blur the line between what looks attractive and what is logically convincing in afterthought (Gargani, 2001). Complex models increase the toll on interpretive skills of those who are to understand and use the representation and allow for more numerous interpretations. In other words, these models make ambiguity more likely.

Increasing openness to context and to context–mechanism–outcome configurations (Pawson & Tilley, 1997) requires more differentiated presentations and accounts. People who have experiences with variations of the program in different contexts bring different experiential accounts of the program logic into the evaluation process. Should contextual variables be understood as givens (as physical environments) or as social constructions that can be changed (over time)? If the latter is the case, the validity and effectiveness of a particular program theory depends on who holds control over relevant contextual variables. In some situations, programs themselves influence their contexts in positive or negative ways (Dahler-Larsen, 2001).

Especially, in crowded policy spaces, the context in which an intervention operates very much consists of other interventions. As a consequence of complexity and interaction between interventions, the conventional distinction between “program theory” and “problem field theory” becomes blurred. The distinction only holds under the classical assumption that one intervention and one program theory be under consideration at one time. If several mental models interact in their consequences, one person’s program theory is embedded in another person’s problem field theory. This fundamental source of ambiguity is suppressed when consensus about the program theory is assumed in the conventional design of the TBE process.

**Involving Stakeholders in More Advanced Ways**

Over the years, TBE has invited broader and deeper stakeholder participation. Involvement of stakeholders increases relevance, ownership, and use (Cousins, 2003). Through dialogue, potential surprises can be detoxified early on, and stakeholders can gain trust in the quality of the evaluation. These factors impinge on use (Ledermann, 2012). Involvement helps stakeholders overcome blind spots in their own theory-making (Friedman, 2001).

However, there are many passage points on the road toward a unitary program theory (Gargani, 2001). Stakeholders have tacit knowledge and different motivations to share or not share their mental models with others. Tacit mental models may change as a function of their explication. Formal and informal rules regulate how stakeholders interact. Stakeholders bring various resources,
intellectually and communicatively, to the table, and they have organizational positions and power
bases with them. Editing rules regulate how to compose a program theory out of the elements that
stakeholders contribute. The evaluator may keep elements of his or her own preferred program
theory in or out in deliberate or inadvertent ways. There are rules and procedures for officially
acknowledging a theory as “the theory.”

Gargani’s penetrating analysis suggests that the sheer number of permutations at different stages
in the participatory theory-making process makes it impossible to make all aspects of the theory-
making process transparent and legitimate. His analysis exemplifies the compressions of ambiguity
needed to produce a unified program theory. Funnell and Rogers (2011) agree:

We have learned that negotiating a common path when very fundamental differences remain can lead to
a program theory to which no one relates in any practical way and is not a useful touchstone for either
program implementation or program evaluation. (p. 130)

Involving stakeholders in creating a unitary and consensual program theory may happen only at
the price of having considerable ambiguity onboard.

Gargani (2001) then asks under what circumstances and why a stakeholder may endorse a final
model even though it contradicts his or her own version (p. 13). From a sensemaking perspective
(which is always anchored here and now), we can answer: Maybe stakeholders have been asked to
accept the theory, and they cannot predict the consequences of doing so, because use of the evalua-
tion and pressure to act comes later. If that is true, consensus around the unified program theory will
be temporary and hollow.

**Exploring Multiple Forms of Use**

Complex models make testing of theory less clear-cut. Theories may be accepted or rejected for
social, political, and institutional reasons some of which are resistant to hypothesis testing (D. Taylor
hypothesis includes theory-related factors and context-specific factors at the same time, so it is
unclear whether a hypothesis failed because the theory failed or because some of many helping
conditions were not in place in the testing situation, including variables that were not properly
recognized and beyond the control of participants. It may be necessary to explain to commissioners
that their expectations about conclusive evidence cannot be met when programs are complicated and
dynamic (Patton, 2012).

Without finally departing from the idea of ultimately testing program theories, an important
strategic move in TBE is to emphasize the many alternative forms of use that are more attractive,
more easy to achieve, or more pressing. A good program theory communicates clearly to external
partners and funders what a program is trying to achieve and how it seeks to do so. A program theory
can help design a good intervention and focus energy invested in the intervention in an optimal way.
It may also be used to develop a program formatively and/or to coordinate the efforts of different
partners whose contributions are essential to the program. A good complex program theory can serve
as a dynamic learning frame in complex interventions and identify systemic tipping points that
require immediate collective action (Patton, 2011). Program theories can also be used to motivate
participants and celebrate successes (Behn, 2003). Overall, the participation in articulation of
program theory may—like process use in general (Forss, Rebien, & Carlsson, 2002)—enhance
reflexivity, learning, and evaluative thinking as well as enlightenment (Weiss, 1977).

As a broad spectrum of forms of use is developed and emerges along the way, some stakeholders
may be left with ambiguous expectations about the consequences of TBE. Acknowledging ambi-
guity about the use of TBE may make a common project possible among a broad set of stakeholders
with different views, interests, and goals (Baier et al., 1986). High hopes about the use of evaluation
(such as positive learning or development) may be shared as long as they remain unspecific. Stakeholders can also positively engage in TBE if they trust that process use is an important form of use and a positive experience, but process use is by definition difficult to specify in advance, so few may get exactly what they expected, and ambiguity may follow. Now that the program theory helped us get funding, does the same theory still demand our loyalty? How much revision can the “same” program theory survive? Do we test the theory or just reflect on the construction process, and can we call all this learning?

No wonder that the basis for stakeholder’s commitments to a “consensual” program theory may be uncertain or shaky. Perhaps these observations help explain why many TBE evaluations do not complete the evaluation process in full compliance with TBE’s own principles and why some seem to forget the program theory as evaluation results come in (Coryn et al., 2011). In the following, I propose an approach to TBE that does not sweep ambiguity under the carpet but instead cultivates it as a source of insight.

**How Evaluators Can Highlight Ambiguity**

Let me focus on a form of ambiguity that can be portrayed directly in a graphic representation of program theory (and later come back to forms of ambiguity which belong to other aspects of the evaluative process). I suggest to juxtapose various partial program theories as early as possible and to use the highlighted ambiguity as an input into a collective sensemaking process. To do so, I propose a simple but innovative technique, which is to simultaneously represent in one model multiple sensible program theories that compete in their interpretation of the same phenomenon.

The theories displayed are diverse, but they interact through at least one common variable. Although the reader may find one of the presented theories more plausible or useful than its counterpart, ambiguity exists in practice when both theories have *some degree plausibility* as clues for sense-making for reasonable people in the practical situation at hand. Whether an influence is positive or negative is defined from the perspectives of the diverse theories. Remember that an ambiguous phenomenon is one about the meaning of which reasonable people may disagree in a way that cannot be easily and immediately settled. Sense-making is situated and dispersed. It is not enough to agree with oneself, given the interactive nature of contemporary TBE.

To call forward differences in interpretation, I represent various theories overlaid in the same graphic model. This heuristic initiative follows Balle Hansen and Vedung (2010) in their recommendation that a unitary model should not be sought routinely but departs from their recommendation to keep multiple program theories separate. Their recommendation is based on model/theory alignment, or in other words, that a joint model is by definition harmonious. Instead, I juxtapose disparate theories into one model that is *not* harmonious. This move highlights ambiguity. The point is not whether a model portrays reality in the only correct way. The point is to tweak the model, so that it portrays ambiguity in a way that is heuristically useful and socially productive.

Consider a very simple program-theoretical example (see Figure 1).

In order to prevent people from throwing cigarette butts on the ground outside the entrance to a public building, the technical staff puts up a large ashtray outside the entrance. Their program theory says: If ashtray, then fewer butts (Arrow 1).

Then, there is another theory: A sign posted at the front door, facing outward, states that smoking is prohibited, which is consistent with legislation. The undergirding program theory says the sign will reinforce a social norm against smoking here and ultimately lead to a smoke-free environment (Arrows 2 and 3).

In separation, both theories, the theory of the ashtray and the theory of the sign, are meaningful. Both consist exclusively of positive phenomena linked together by positive causal influences. But there is ambiguity in their interaction. In real life, smokers make sense of the ashtray as an approval
of their smoking. They ignore the sign or erroneously conclude that smoking is prohibited only after you have entered the building. But they do have a point: Why is there an ashtray if smoking is not allowed?

Arrow 4 symbolizes the unfortunate negative influence that the ashtray has on the norm against smoking. The arrow is drawn not because it is a part of the original program theory for the ashtray but because it is a part of the program field theory of the sign which seeks to enhance a nonsmoking environment. If smoking is perceived as legitimate, smoking will not be eliminated, and Arrow 5 shows the impact hereof on butts on the ground. (Arrow 5 is in the model because it is a part a program field theory of Theory 1.)

According to the views of technical staff, the ashtray helps reduce cigarette butts on the ground. According to antismokers, it undermines the program theory in support of a norm against smoking. The ashtray is a Janus variable. The ambiguity is easy to see visually, because there are two arrows from the same phenomenon, one positively valenced and the other negatively.

Depending on the relative strengths of the forces represented in the model, the ashtray may paradoxically lead to more, not fewer butts on the ground. Even in this simple model, conflicting but interconnected logics are set in motion. I see the root of these in the two arrows (one positive, one negative) from the ashtray. Janus the God looks in two directions, but he holds the key to insight.

Should I have presented the two program theories side by side, one about ashtrays (Arrow 1) and another about nonsmoking signs (Arrows 2 and 3)? No. A separation of the theories would not have made their unfortunate links visible (especially Arrow 4, but also 5). It is not the sum of diverse theories that make a difference. It is their interaction. The overlay of theories helps call this effect forward.

Others may argue that this example does not illustrate evaluation because data have not been collected yet. I do not agree. Critical thinking is part of evaluation. One of the most constructive uses of TBE is to get evaluative thinking started early. A critical analysis of the program theory is helpful, the sooner the better. If key stakeholders collectively hold theories with such a contradictory message as the one presented in Figure 1, it is predictable that their reality will be messy. Further
evaluation resources can be saved if an early critical review of causal assumptions improves the intervention. That is why I suggest not to postpone the confrontation of diverse program theories until a final deliberative moment but instead use the discovery of ambiguity in the service of an improved coordination of interventions.

In the case at hand, it was easy to disambiguate the situation. The technical staff were kindly reminded about official legislation, and they were ordered to remove the ashtray from the entrance to a place further away where smoking was allowed. The forces of legislation and management were in position to remove ambiguity. But it happened only because an evaluative mind (your humble author) reacted to the ambiguity that is represented in formalized form in the model above. The attention to ambiguity was both a source of insight and a call for action.

What Are the Types of Ambiguity That Evaluators Encounter?

The ambiguity described in this simple example is not the only form of ambiguity which TBE may encounter. In this section, I offer a typology (based on experiences with evaluation in health care and social services in Denmark). While the above example illustrates only the first type (functional ambiguity), I will briefly mention how some other forms of ambiguity could be illustrated graphically, but space does not allow full-fledged case studies and graphic models of all types. In addition, ambiguity in the sociopolitical context around an evaluation process, which is not part of program theory itself, indirectly influences stakeholder views and commitments to theories. All in all, the types below are meant as heuristic devices, not as exhaustive and mutually exclusive types.

**Functional Ambiguity**

As illustrated in the example above, a Janus variable is functionally ambiguous when it has a positive causal effect on one chain of events and a negative on another. You can also imagine a caretaker who prepares lunch for an elderly person with the best intentions (making sure the person gets something to eat), but thereby undermines an official policy of coproduction, which is intended to help the elderly maintain her own competencies in the kitchen. Another example is antidepressive pills, which create dependency over time. This is more than just a side effect. Addiction is bad for patients, but beneficial for the commercial interests of drug producers. (Remember that their program theory influences elements in your program field theory.) Addiction is a wicked Janus variable. These are all examples of functional ambiguity because of the simultaneous positive and negative causal influences flowing from the Janus variable.

**Means/Ends Ambiguity**

There is a fundamental evaluative distinction between phenomena that have inherent values in themselves versus phenomena that have value only because they help produce other valuable phenomena. We call these ends and means, respectively. There is potential ambiguity when some see the same phenomenon as a means and others see it as an end (thus, a Janus variable). Some may support an initial outcome in a particular program theory not as an end in itself but as a means to another end.

Consider an organizational reform in the hospital sector. According to one theory, a number of initiatives such as training, information technology, coordinators, and new incentives should lead to a new organization. Indications hereof would be a new culture, a new structure, and communication infrastructures that work in order to support new patient journeys efficiently. In turn, this new organization should lead to improved treatment outcomes for patients. Many, including funders,
intuitively support this theory because ultimate outcomes for patients constitute a convincing jus-
tification for the whole program.

According to organizational consultants or change agents, the new organization is a goal in itself
that, if accomplished, allows them to turn to other tasks. Whether the final outcome can be safely
neglected depends on whose point of view is applied but also on which promises were originally
made. Graphically, we would have a theory of A leads to B overlaid with another theory of how A
leads to B which leads to C. B (the new organization) is a Janus variable because it as a means for
some and for others an end in itself. Whenever “outcomes” are linked in a chain, ambiguity lurks,
because all outcomes except the final one operate as both means and ends.

**Contextual Ambiguity**

Context is a broad and slippery notion. Contextual ambiguity includes different views about whether
textual conditions and moderators are in place to make the program successful (Pawson & Tilley,
1997), whether these factors can and should be manipulated so that the program theory works better,
and, if yes, whether these factors should really in essence be seen as part of the program. Can
sufficiently favorable conditions for a particular program be created in a particular context, and
should they? How far does control of context reach? To what extent do people in the context have a
responsibility for taking part in the program?

Consider a package of initiatives to combat stress among teachers developed by a special center.
School principals help implement the initiatives against stress but they do not see themselves as part
of the intervention. It they are understood as “context,” the intervention fails if it does not induce
school principals to do the right thing. If they are enrolled in the “intervention,” and they fail, they
are coresponsible for an implementation failure. Actors and variables in contexts that “help” inter-
ventions work, share important double-edged characteristics with Janus variables. Graphically, the
same phenomenon would be portrayed in a context versus as part of an intervention in two com-
peting theories.

Such an ambiguity can be used in a political perspective to place or displace accountability,
blame, and risk (Rothstein, Huber, & Gaskell, 2006), if the diverse theories are kept apart. However,
it can also lead to joint sense-making and recognition of mutual responsibility, which is more likely
if the competing understandings are first brought together into one model that shows the interaction
of the diverse causal understandings.

**Temporal Ambiguity**

Beliefs, values, and loyalties change for a number of reasons. Sometimes, “stakeholders are only
vaguely aware that their theories have changed” (Funnell & Rogers, 2011, p. 147).

Temporal ambiguity relates to whether people do in fact remember old documents, whether they
keep promises, whether they use earlier arguments to stabilize present beliefs, whether they have
control over their future preferences, whether various stakeholders come and go through the evalua-
tion process, whether dominant cultural beliefs in program theories are stable and solid over time,
and whether learning from the TBE process is genuine.

The grounds for commitment to a particular program theory at a given time are not always
knowable. A person may endorse a program theory because her values are represented in means,
not in ends in the theory, because the presentation of the theory is neat, or because it helps secure
funding. If initial acceptance of a common program theory is functionally necessary for the further
evaluation process, the commitment may not be deeper than that. Some may be on board because
they like the theory, some because they hope it will be tested, and some because they assume it will
not be. Some are on board because they did not or could not foresee the consequences of TBE.
Now, take a second look at the example with the new organization in the hospital sector mentioned above. Although great outcomes for patients have been promised, methodological conditions needed to document them are not in place. The organizational change is comprehensive, thus making it difficult to establish a control group that receives treatment as usual (Ovretveit, 2003). Without a control group, an ultimate treatment effect for patients cannot be demonstrated convincingly, even if it was the crucial element that helped finance the reform in the first place.

If program people deliberately speculated in promising an outcome that they knew could never be supported by evidence, they may have acted in their own self-interest. If they learned along the way about the difficulties of documenting an outcome they hoped and worked for, they may have lacked methodological competence but still have pure intentions. It may remain ambiguous why their commitment to their original promises about outcomes withers away.

Graphically, you could imagine ambiguity in the friction between an old and a new theory overlaid in the same presentation, which would at least make the temporal ambiguity visible.

**Consequential Ambiguity**

This ambiguity relates to how evaluative conclusions can lead to decisions and who can, consequentially, be held accountable for successes and failures. If TBE is adopted mostly as a tool for learning, the initial outcome is hypothetical, and new versions of program theory with new outcomes can be developed over time, apparently without any repercussions (Funnell & Rogers, 2011). In an accountability perspective, the promised outcome may instead be critical for the survival of the program.

Belief in the power to directly or indirectly influence a particular factor for which one is held accountable increases the chances of acceptance of a program theory which includes that factor (Vakkari & Meklin, 2006). The opposite is true if one is held accountable for factors one cannot control. If one actor pushes the risk of failure risk to another level in the system, actors on that other level may figure out what the risk implies, and they may try to push it back (Rothstein et al., 2006).

Consequences also depend on how strictly an accountability regime is imposed, especially when it comes to funding. It may be impossible to know from looking at a program theory which accountability regime will be linked to evaluation results in the future. Consequential ambiguity is eternally paradoxical because if good results have good consequences, they may be deemed good from the perspective of immediate sense-making, even if they are based on a (logically) bad theory, whereas a (logically) good theory cannot be unconditionally good if it leads to undesirable consequences. These speculations are consistent with Kahneman’s (2011) findings: People adopt or reject causal theories in everyday life for a number of reasons that have little to do with the validity of these theories. Ambiguity rests within a program theory at least until its consequences are seen.

Consequential ambiguity makes the significance of phenomena in program theories change depending on the consequences of the TBE process. It is naive to assume that advocates of a program theory do not consider the consequences of that theory for themselves. Nevertheless, even if they are shrewd, they may learn that their depiction of a particular program theory turned out to be not in their best interests, for example, if evaluative findings later indicate an unforeseen failure of their favorite theory. Presumably, the more experienced constructors of program theories are, the more advanced will be their understanding of how TBE resonates with their interests. Sometimes, exploiting temporal ambiguity and simply delaying reactions are ways of handling consequential ambiguity. An evaluation of outcomes of a reform is often postponed until it is too late to reverse the reform. Since purposes and intentions are interpreted retrospectively in sense-making processes (Weick et al., 2005), various stakeholders have some leeway to change their commitments to various versions of program theory over time. Consequential ambiguity around theory-making thus helps us understand why stakeholders reserve their commitment so that ambiguity occurs in theory-making.
The above typology is not exhaustive; rather it offers a first attempt to identify indicative types of ambiguity. Some forms of ambiguity can be directly illustrated as Janus variables in program theory models, as suggested and shown in the example. Janus variables can be identified when there are competing and interacting program theories that share at least one phenomenon. They compete in the sense that they offer two contrasting versions of the meaning of that phenomenon in the program theory model. Other forms of ambiguity, such as consequential ambiguity, are better understood as facets of the evaluative process that indirectly influence theory-making and stakeholder commitments to these various program theories.

**Discussion: How Much Ambiguity? Under Which Circumstances?**

Decisions to highlight and confront ambiguity depend on how evaluators perceive of ambiguity as a device in their work and how they see their role as evaluators. Shadish, Cook, and Leviton (1991, p. 62) distinguish between focusing and contingency devices. A focusing device suggests what evaluators should focus on generally across situations. With ambiguity as a focusing device, an evaluator would always seek to find cracks, openings, and controversies in causal theories. Whether or not ambiguity is created as a part of a cunning political strategy or not, it is clarifying to confront it. Limits can only be known by pushing them.

In contradistinction, with ambiguity as a contingency device, evaluators should attend to ambiguity only in contexts where it will be fruitful to do so. To identify those situations, following questions may be useful:

- **Is a program theory model with an ambiguous and illustrative Janus variable easy to draw?** To find this out, draw little sketches of models with Janus variables at an early stage. Good models should emerge with competing theories, depicting the same phenomenon in conflicting ways. The relevance of this endeavor rests with how critical this friction is for program improvement and for wider sociopolitical issues.

- **What are the likely consequences of highlighting ambiguity?** In the case with the cigarette butts, it was relatively easy to disambiguate the situation. Once the friction between the two program theories was acknowledged, the ashtray was removed. In the example, where the caretaker prepares food for the elderly, disambiguation may be more difficult because there is a real tension in roles and expectations. The elderly person has lived a long life and feels she deserves to be taken care of. Official program theories say that the elderly must manage daily activities herself in order to maintain her competencies. The partners involved will have to find a way and negotiate a plan for who does what. Early theorizing helps identify the problem.

Imagine again our intervention against stress among teachers. Some of the initiatives focus on the individual level (awareness raising, time management, learning to respond to feelings of unease, relaxation techniques) and some on an organizational level (management attention, reorganization of work). Such a package is meaningful because earlier interventions on one level only were ineffective as they overspecified the separate responsibilities of either managers or employees.

The evaluation identified a special problem that occurred when individuals under stress contacted managers in order to have their workload reduced or their work processes redefined. Managers “solved” the problem by transferring work from the stressed employees to the not-yet-stressed, thus increasing the pressure on the latter. In other words, the evaluation identified “reorganization of work” as a Janus variable because it worked well in some respects and badly in others. To realize that this is more than just a conventional side effect, it is necessary to see the interventions at the individual level and the organizational level as systemically connected and to see the effects upon
the stressed and the not-yet-stressed as part of the same reality. This is what the evaluators did, and they recommended to weave together the individual and the organizational interventions more cleverly in the next version of the intervention.

In a broader sociopolitical context, one can ask whether a particular program theory is already officially mandated and accepted as such or whether there is a landscape of alternative theories underneath a more or less superficial consensus.

- Have particularly promising alternative theories been ignored? How deep-seated are worldviews and political interests that undergird various theories about interventions in a policy area?
- How high is the level of political conflict? Is there a pressure for change that is likely to enhance the use of evaluation even under conditions of conflict (Ledermann, 2012)?
- Can a willingness to deliberate ambiguity among stakeholders be promoted?
- Is the friction between alternative theories of particular democratic interest so that the issue deserves to be brought to broader democratic arenas?
- Can the program or project at hand serve as an exemplar in policy-making or in evaluation?

Depending on such situational factors, a TBE process that highlights ambiguity may be more or less productive. A complication is that a focus on ambiguity may be productive also under less inviting conditions, but it would take more energy and perhaps conflict to get to a productive result. Sometimes, paradoxically, the more difficult evaluation is to do, the more it is needed.

An evaluator’s stance on ambiguity would also depend on his or her role. Being an evaluator in a collective form of theorizing is like being an advisor. Emanuel and Emanuel (1992) have conceptualized advisor roles. The first one is paternalistic. A paternalistic advisor considers all information available from his or her perspective and recommends a course of action. A paternalistic advisor runs the risk of sweeping ambiguity under the carpet. The conventional TBE process may lean toward a paternalistic role for the evaluator when he or she articulates a unified theory that should be accepted by all as a common framework for the evaluation.

An informational advisor makes information available and lets people decide for themselves. An interpretive advisor, however, puts various pieces of information together, shares his or her thoughts with stakeholders, and explores the implications together with stakeholders. An interpretive TBE evaluator calls forward ambiguity and stays with stakeholders until sufficient sense has been made of it. Finally, a deliberative advisor is one who explores worldviews and value differences undergirding the ambiguities that surface in TBE. In this capacity, the evaluator does not only help stakeholders understand the meaning of ambiguity but also explores conflicting value frameworks and alternative courses of action. Therefore, it is important to pay attention to ambiguities in the sociopolitical context around the theory-making process. I see the two latter roles, the interpretive advisor and the deliberative advisor, as most compatible with theorizing in TBE as a collective and deliberative form of sense-making. At the same time, the roles of interpretive and deliberative advisor are by definition interactive.

We have come full circle when we recognize that the TBE evaluator is thrown into concrete and ongoing sense-making together with others. This is not surprising. Remember, this condition generally characterizes the human construction of meaning under conditions of ambiguity (Best, 2008).

**Conclusion and Implications**

The socially constructed nature of program theories reveals itself as a logical consequence of contemporary changes in TBE itself. So it may be less fruitful just to deny or neglect ambiguity.
Ambiguity may help us understand why TBE evaluators already have found theory-making to be a rough social process, why unified and official program theories are sometimes forgotten in later parts of the evaluation process, and why TBE does not always live up to its own ambitions (Coryn et al., 2011). The types of ambiguity identified in this article (functional, means–ends, contextual, temporal, and consequential) can be used as an agenda for deliberation among evaluators and stakeholders. Which ambiguity can we sort out? Which ambiguity can we have on board and still get organized?

The typology is not exhaustive. I could have added, for example, evidential ambiguity that relates to whether it is possible, under given circumstances, to convince a set of stakeholders about some evaluative claim on the basis of data and arguments, relative to the perceived novelty of findings and the perceived quality of the evaluation (Ledermann, 2012). Not enough attention was paid to value ambiguity where no agreement or clarity exists about whether a phenomenon (such as an outcome) or aspects hereof should be evaluated and if yes, positively or negatively, but it might be fruitful to do so. Given recent research on the structure of values and valuing (Julnes, 2012), TBE evaluators and theorizers may identify more forms of ambiguity in the future.

If ambiguity is acknowledged, TBE will gain in sophistication as a social and politically relevant practice. TBE might develop a vocabulary that emphasizes the interactive and fragile nature of theory construction and distances itself from an air of technical hypothesis testing, especially if TBE in fact no longer recommends testing of theory as the exclusive end point, but instead aims more broadly at learning, enlightenment, and deliberation. A paradigm shift may help move focus from consensual program theory toward theorizing as a socially embedded sensemaking process. We should encourage experiments with intelligent formats for developing local program theories and making sense of the ambiguity and friction in their interaction. My specific proposal is to juxtapose disparate program theories in one graphical model. This concrete and low-tech tool is innovative but builds on existing trends in TBE and is therefore feasible.

Ambiguity is a source of insight. Janus variables made visible in early stages of TBE serve as triggers for sense-making. They call for deliberation and a conversation about potentially better coordination since they show that theories interfere with each other in the same sociopolitical reality. Janus variables are particularly relevant when there are competing, interacting, and even mutually counterproductive program theories.

Unpacking ambiguity takes place bit by bit and is not equally pressing, relevant, nor possible in all situations. Situational characteristics should be considered, such as whether genuine contributions can be made to new knowledge among stakeholders, and how susceptible they are to deliberation. On the other hand, in the spirit of ongoing collective sense-making embedded in a democratic polity, not to unpack ambiguity is also a choice and it is not neutral.

If students of TBE learn about ambiguity at an early stage, they are more likely to understand the frictions they will encounter in TBE in practice and better prepared to reinvent TBE for the future. TBE evaluators can become better pathfinders and analytically stronger advisors if they can typify and deal with ambiguities they encounter. These include types of ambiguity which can be portrayed directly in program theory models as well as ambiguities in the sociopolitical context which indirectly influence theory-making.

The evaluator who plays an interpretive or deliberative role vis-à-vis ambiguity should not pretend that all aspects are under control. As TBE departs from the conventionally designed process and the evaluator role changes, the responsibility for the process should be shared. Partial program theories may be developed separately, but it is bringing them together in a collective, argumentative, and reflexive encounter that realizes their democratic potential. This is an important reason why it is fruitful to pay more attention to ambiguity in TBE. A good, operational way of doing so is to recognize and address Janus variables in play.
Declaration of Conflicting Interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

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

References


