American Journal of Evaluation

Designed Blindness: An Action Science Perspective on Program Theory Evaluation

Victor J. Friedman

The online version of this article can be found at:

Published by:

$SAGE

http://www.sagepublications.com

On behalf of:

AMERICAN
EVALUATION
ASSOCIATION

American Evaluation Association

Additional services and information for *American Journal of Evaluation* can be found at:

Email Alerts: http://aje.sagepub.com/cgi/alerts

Subscriptions: http://aje.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://aje.sagepub.com/content/22/2/161.refs.html

Designed Blindness: An Action Science Perspective on Program Theory Evaluation

VICTOR J. FRIEDMAN

ABSTRACT

This article is intended to stimulate a dialogue between program theory evaluation and action science for the purposes of cross-fertilization and mutual enrichment. Both program theory evaluation and action science use the concept of implicit "theories of action" as a central construct in the study of social practice. However, an action science approach suggests a wider understanding of program theory that (1) specifies the links between individual reasoning and behavior to program implementation, and (2) accounts for how programs deal with dilemmas, conflict, and error. This paper begins with a systematic, though not exhaustive, comparison of program theory evaluation and action science. It analyzes an exemplar of program theory evaluation from an action science perspective to illustrate a subtheory, "designed blindness," and its impact on both program implementation and the evaluation itself. It then offers a theory for overcoming designed blindness. Finally, this article argues that action science concepts and skills can enable program theory evaluators to be more effective in confronting defensiveness and in facilitating learning among stakeholders when there is a gap between "espoused" program theory and "theory-in-use."

INTRODUCTION

The ultimate objective of this article is to stimulate a dialogue between "program theory evaluation" (Chen, 1990; Rogers, Petrosino, Huebner, & Hacsi, 2000) and "action science" (Argyris, 1993; Argyris, Putnam, & Smith, 1985; Argyris & Schön, 1974, 1996; Friedman, 2000). In a sense, this article is a response to a puzzling lack of dialogue and cross-fertilization between two schools of applied research that employ the concept of implicit "theories of action" as a central construct in the study of social practice. This dialogue should be instrumental in helping both fields overcome current obstacles to achieving their goals of improving social practice (Lipshitz, 2000; Weiss, 1997).

Victor J. Friedman • Department of Behavioral Sciences, Ruppin Institute, Emek Hefer 40250, Israel; Tel: +972-9-8983031; Fax: +972-9-8987604; E-mail: victorf@ruppin.ac.il.

161

The idea of program theory emerged in the evaluation community about 30 years ago (Suchman, 1967), and has gained wide acceptance as an important concept for understanding how programs work and for assessing their effectiveness (Bickman, 1990; Chen, 1990; Lipsey & Pollard, 1989; Patton, 1997; Rogers et al., 2000; Weiss, 1997). It has appeared in the literature under a variety of titles: "theory-driven evaluation" (e.g., Chen, 1990), "theory-based evaluation " (e.g., Weiss, 1997), and "program theory evaluation" (e.g., Rogers et al., 2000; Stufflebeam, 2001). Action science emerged from inquiry into the nature of professional practice (Argyris & Schön, 1974). It aims to integrate theory building into practice in fields characterized by uniqueness, uncertainty, and instability (Argyris et al., 1985, p. 36; Friedman, 2000). Action science, or the "theory of action approach" (Argyris, 1997), has played an important role in theories of organizational learning (Argyris & Schön, 1978, 1996; Senge, 1990) and reflection-in-action (Schön, 1983, 1987).

Program theory evaluation and action science both share the proposition that tacit "theories of action" are embedded within program planning and practices (Argyris & Schön, 1978; Chen, 1990; Schön, 1997). Theories of action are "if. . . then" propositions that specify the causal links between program ends and means as well as the conditions under which these ends are to be achieved. Program theory consists of the goals of program, treatments, and the implementation environment (Chen, 1990). In action science terms, an organization's "instrumental" theory of action is a complex system of goals, norms, action strategies, and assumptions governing task performance (Argyris & Schön, 1978, pp. 14–15). Both program theory evaluation and action science recognize that program stakeholders are frequently unaware of the theories implicit in their action.

Theories of action serve explanatory functions (i.e., explaining events or behavior), predictive functions (i.e., inferring future events or outcomes), and normative/control functions (i.e., identifying actions that should be taken so that outcomes can be made to occur). Chen (1990, p. 53) distinguishes between "normative program theory," or what program designers and practitioners intend, and "causative theory," which reflects the actual causal mechanisms relating treatments to outcomes. Patton (1997, p. 219) built on this distinction to identify three approaches to program theory evaluation: a "deductive approach" that draws on academic theory and research to determine the causative theory, an "inductive approach" that focuses on generating grounded theory through fieldwork, and a "user-focused approach" that attempts to extract the theory of action implicit in the actions of program users.

Much of the current interest in program theory evaluation focuses on developing a model of the logic implicit in program design. Logic models explain how program inputs and activities are intended to create the desired outcomes (Diefendorf, 1999; Funnell, 1997; McLaughlin & Jordan, 1999; Schmitz, 1999). Logic models are a tool for developing program theory. However not all logic models represent valid program theories, which require a deeper, more demanding, empirically-based explication of causal mechanisms and underlying assumptions (Cole, 1999; Stufflebeam, 2001; Weiss, 1997).

Action science makes a fundamental distinction between "espoused theory," what people intend or believe they do, and "theory-in-use," what can be inferred from actual behavior (Argyris & Schön, 1974, pp. 6–7). Both program theory evaluation and action science aim at improving social practice by discovering gaps between espoused theories and theories-in-use and by bringing this information to the attention of stakeholders (e.g., Argyris, 1982, 1993; Bowen, 1999, 2000; Chen, 1970; Della-Piana, 1999; Friedman, 2000; Patton, 1997; Weiss, 1997). Rogers et al. (2000), however, point to a lack of published studies systematically comparing program espoused theory with program theory-in-use,

noting that while "collaboratively building a program theory can be an energizing team activity, exposing this to harsh empirical tests can be less attractive" (pp. 11–12). Well-developed logic models provide a useful framework for identifying measurable short- and long-term outcomes (e.g., Funnell, 2000), but outcomes do not necessarily reveal theories-in-use. Jumping from the espoused logic model to the measurement of outcomes risks seeing and measuring only what is in the logic model, while ignoring aspects of program theory-in-use that have a critical impact on performance.

Donaldson and Chen (2001) cite a large number of published theory-driven evaluations, but they also note that the involvement of stakeholders and the formative, interactive process of program theory development "pose substantial implementation challenges in practice" (p. 15). An action science approach offers both concepts for understanding program theory and a set of evaluation strategies that can help move the field to meet these challenges. For example, action science proposes that organizations possess subtheories for dealing with gaps, inconsistencies, and conflicts in instrumental theories-in-use. These features of theories-in-use result from the need to deal with change as well as with incompatible requirements for performance (Argyris & Schön, 1978, 1996). These subtheories guide adaptation and learning through the detection and correction of error, as well as the discovery and exploitation of opportunities. Other subtheories, such as "defensive routines" (Argyris, 1985), function to maintain a sense of constancy and protection from uncertainty, instability, threat, and embarrassment. Defensive routines powerfully influence a program's implementation (theory-in-use) and limit learning.

Because action science views program theory as serving both instrumental and defensive functions, it suggests that program theory evaluation has not fully accounted for how threatening it is to "open up the black box." It is one thing to present program stakeholders with data that indicate that they have fallen short of desired outcomes. It is another thing, and perhaps even more threatening, to confront them with data that reveal gaps, inconsistencies, and contradictions in their own thinking and behavior (Argyris, 1982; Friedman & Lipshitz, 1992). Program theory evaluation says little about the difficulties of confronting people with gaps between program theory and program practice.

From an action science perspective, the key to addressing defensive routines lies in understanding the link between individual theories-in-use and theories-in-use at the organizational level (Argyris, 1985, 1993). Action science research has found that, when confronted with gaps or inconsistencies in the theories-in-use, individuals become defensive and try to distance themselves from responsibility for these errors. Even when actors acknowledge the contradictions, they are frequently unable to behave in ways consistent with their espoused theories (Argyris, 1982; Argyris & Schön, 1974, 1978). These responses are so predictable that Argyris and Schön (1974) hypothesized that individual theories of action are driven by a higher-order mental theory-in-use, which they called "Model I," which aims at maximizing unilateral control, protecting oneself and others, and maintaining rationality.

A key component of Model I theories-in-use, as with any skilled behavior, is unawareness (Argyris, 1986). Unawareness enables people to smoothly execute many complex behaviors—whether it be driving, skiing, teaching, managing, or evaluating—almost automatically and without conscious thought. Model I reasoning can be effective for solving relatively unthreatening, technical problems but not under conditions of uncertainty, ambiguity, and goal conflict. Model I reasoning amplifies ineffectiveness in situations that require people to critically reflect on their perceptions of reality, their reasoning processes, their values, and their own competence (Argyris & Schön, 1974). Furthermore, when people

interact at the group or organizational level, their Model I theories-in-use create a set of dynamics that limit effectiveness and learning (Argyris & Schön, 1978; Argyris & Schön, 1974).

Argyris and Schön (1974) discovered that even highly skilled professionals, including themselves, were unable to overcome Model I defensive routines through insight alone. Therefore they invented and learned "Model II," an alternative theory-in-use that enables people to critically reflect on their reasoning processes under conditions of uncertainty, ambiguity, conflict, and psychological threat. Model II reasoning and behavior is driven by the values of generating valid information, free and informed choice, and internal commitment. Model II reasoning has proven effective in enhancing individual and organizational learning (Argyris, 1993; Argyris & Schön, 1996; Friedman, 2000; Schön, 1987).

Because program theory evaluation focuses mainly on instrumental program theory and not the subtheories and higher-order theories that govern implementation, there is a potential blind spot in the evaluator's field of vision. This article will use action science concepts and analytical tools to analyze an exemplar of program theory evaluation (Bowen, 1999, 2000). It will identify a specific program subtheory and how it prevented program stakeholders from discovering and addressing gaps between program planning and implementation. This subtheory, which will be called "designed blindness," also influenced the program evaluator, leading her to unintentionally act in ways that reinforced stakeholder blindness. Finally, this article will offer a strategy for overcoming designed blindness.

PROGRAM THEORY EVALUATION: A CASE STUDY

A study entitled "Development of local program theory: using theory-oriented evaluation to make a difference" was presented at the 1999 AEA Annual Conference (Bowen 1999) and later published in a slightly revised version (Bowen, 2000). That research was aimed at uncovering implicit theories of action for the purpose of explaining program outcomes. It was consistent with action science in that the research findings were shared with the subjects at both the individual and group levels to test for validity and to inform practice (Bowen, 1999, pp. 10–11). Furthermore, the study's methodological rigor and its use of a wide variety of data collection and analysis methods went beyond what is commonly employed in action science. Nevertheless, as the following analysis illustrates, it provides an example of designed blindness in both program practice and program theory evaluation.

The aim of Bowen's (1999, 2000) study was to describe the program theory of a Teen Age Services Act (TASA) program, that is, an adolescent pregnancy and parenting program established in accordance with a state legislative mandate to provide services that "maintain and strengthen family life and. . . maintain or retain the capability for maximum self-support and personal independence" (Bowen, 1999, p. 4). The evaluation research took place in a county where a community-based case management program had been contracted to carry out the legislation. Teen participation in the program was voluntary. The evaluation itself was initiated by the researcher, for whom it constituted the basis of her doctoral work, and the program agreed to host her.

At the outset of the paper, Bowen (1999, p. 4) pointed out that the outcomes of the program were mixed. On the one hand, program participants were delaying second pregnancies and fewer of them were welfare-dependent, homeless, drug abusers, or infected by sexually transmitted diseases. On the other hand, fewer participants attended and/or gradu-

ated from high school, received child support from the fathers of their children, and improved their independent living skills. During the years studied, the community experienced an increase in the number of pregnant and parenting teens, as well as an increase in foster care placements.

To try to understand these mixed outcomes, Bowen (1999, 2000) analyzed the program theory at four different levels. The first level of analysis involved the logic model of the program planners as inferred from established goals and objectives, financial and budget records, organizational rules and regulations, correspondence to program staff, annual reports, and other documents. The planners' program theory aimed at reducing welfare dependency, homelessness, second pregnancy, and foster care, while increasing child support, living standards, education, and mental/physical health. According to planners' logic, these goals were to be achieved through a "goal planning" process involving on-going needs assessment, counseling, goal setting, and service referral and coordination (Bowen, 1999, p. 8).

The study then analyzed the local program theories as held by the program staff members (case managers) to uncover the logic by which specific implementation strategies (e.g., case management, providing information, home visits, and transportation) were supposed to lead to the established outcomes. Using concept mapping (Trochim, 1989) and one-on-one interviews, Bowen (1999) constructed a detailed logic model that was confirmed as accurate by the staff members themselves (p. 13). This logic model reflected the program staff's belief that the most crucial strategies for achieving program goals were "relationship building" and "encouragement/goal setting" (Bowen, 1999, p. 9).

Relationship building was not an explicit part of the program planners' design. However, the program staff believed that the foundation for the success of all the interventions strategies was the establishment of a relationship based upon trust, respect, and nonjudgmental attitudes. They wanted to achieve program goals by promoting the teens' emotional health and positive self-esteem, focusing on their strengths, and acknowledging "small steps" made in the right direction. In addition, program staff considered the planners' goals unfeasible (p. 22), instead seeing the real goal as "planting seeds for teens to nurture and actualize on their own" (p. 31).

Bowen (1999) took this analysis a step further by comparing the staff's program theory with the logic implicit in actual implementation. The latter was inferred through chart review and observations, as well as concept mapping and interviews with the program recipients themselves (i.e., the teens). In action science terms, this amounted to comparing the staff's espoused theory with its theory-in-use. While most of the teen interviewees cited a trusting/nonjudgmental relationship with their case managers, few were able to identify specific ways in which their decisions were influenced or their behavior changed through the program. As a result, Bowen (1999) concluded the staff's primary objective was not goal achievement and behavioral change, but rather building relationships with the teens (p. 25).

These findings would have been fairly consistent with the staff's espoused theory except that concept mapping revealed that the teens perceived tangible goods and services as the most important aspect of the program. They viewed the program "as a stable, dependable source of advocacy and support as well as a source for tangible items to meet their needs such as diapers, clothing, transportation, and infant formula" (Bowen, 1999, pp. 22–23). On the other hand, interviews with the teens confirmed that the relationship with the case manager was perceived as the most important aspect of the program. This apparent discrepancy was reconciled by the fact that the teens understood "a trusting relationship and unconditional support" to mean that the case managers would help them get what they wanted without

questioning these needs. For example, when the staff did engage in personal goal establishment with the teens, it often involved goals such as getting a car, money for gas, or daycare for the child, which were not reflective of the outcomes envisioned by program planners (Bowen, 1999, p. 22).

In their unofficial role as advocates, case managers frequently found themselves in conflict with other service providers in the Department of Social Services and in the schools. These other service providers accused the case managers of making life too easy for the teens by taking responsibility and making excuses for them. The case managers believed that these people did not understand the plight of the adolescent parent. They viewed these particular service providers as "indifferent, uncaring, and rude" (Bowen, 2000, pp. 36–37).

During a feedback session with the case managers, the evaluator shared with the case managers the teens' perceptions of their role. They became defensive, justifying their actions and providing explanations of the teens' behavior, and she backed off from the issue. This incident was not mentioned at all in the written report (Bowen, 2000), which was much less explicit about the discrepancy between the case managers' and the teens' perceptions than was the oral presentation. Although the study emphasized that both the staff and the teens perceived a "trusting relationship" as central to the program theory, it did not explicitly address how this relationship might have been interpreted differently by the two groups, though these differing interpretations could be inferred from a careful examination of the information provided in a table (Bowen, 1999, pp. 21–24).

In her published report, Bowen (1999) explained the mixed program outcome data in terms of a gap between the planners' program theory and the realities of the implementation environment:

> This evaluation produced evidence that suggested (that) the establishment of context free program goals by external planners may very well have been a recipe for the *illusion of program failure* (my italics). This illusion might very well lead to the reality of program discontinuation despite the positive impacts the program is also making on adolescents lives. This program, like many others, had a paucity of indicator data that could accurately document the effects of each intentional staff intervention. Sorely missing was the tracking and measurement of interim milestones that could show reliably that staff's focus on relationship building did indeed impact positively on program participants, assisting them reach program goals as established at some point in the future . . . . The people who planned the program did not share the same logic as people implementing the program. Because external planners were context free, their plans (while well meaning) for the program reflected generic preventive or ameliorative strategies drawn from the literature and/or personal experiences rather than selecting responsive outcomes that reflected the exquisitely specific context of the community in which this program existed. Additionally, the political environments and personal agendas of politician/decision makers may have driven the establishment of outcomes that were not necessarily aimed at reality for this community and this program staff.
> . . . in the process of building relationships and dealing with contingencies, staff sometimes lost sight of long-term program goals. . . unbeknownst to staff, the process of program delivery had overtaken concerns about and focus upon the end product (Bowen, 1999, p. 28, 32).

These conclusions suggest that a better understanding of the specific implementation environment would have enabled program planners to design a more appropriate set of goals and interventions or a training program for staff to help them avoid "losing sight" of formal project original goals despite obstacles.

AN ACTION SCIENCE CRITIQUE: "DESIGNED BLINDNESS" AS A PART OF PROGRAM THEORY

Bowen's concluding remarks pointed to the "illusion of program failure" and traced it to the establishment of "context free" goal setting. However, a careful reading of her study revealed a chain of blindness at a number of different levels of the program. She indicated that the decision makers who legislated the program were blind to the reality of the situation they were trying to address, that the program planners were blind to key features of the context, and that the staff who implemented the program became blind to its long-term goals. This chain of blindness among the program stakeholders was reinforced by the formal system for monitoring program performance, which itself was "blind" to potentially important interventions and achievements.

The evaluator's confrontation with the program staff revealed two additional links in this chain of blindness. One link was the staff's blindness to the teens' perception of the program and the meaning of relationship building. When the evaluator herself tried to bring these findings to the staff's attention, she encountered a defensive routine that rebuffed this potentially disturbing and embarrassing information. The other link involves the evaluator herself. She not only backed off on this issue, but also downplayed it in her written report. When I asked the author whether she might have been protecting the staff from embarrassment, she wrote:

> I have seriously thought about the possibility of my inadvertent protection of the program staff and, unfortunately, I think to a certain extent you are right. . . I do believe that the data provided sound evidence that this program was effective in many respects. . . The program staff was providing all of these adolescents with unconditional support, advocacy, and relationships built upon trust. My mistake, I think, was not realizing that there may have been a significant number of teens using these attributes to get what they needed and wanted (this was an oversight I really regret). . . .

Dr. Bowen's candid response showed that she herself was unintentionally "infected" by blindness and had been unaware of it.

This chain of blindness may sound familiar to experienced evaluators and program people. However, Bowen's (1999), 2000) study did not offer an answer as to why it occurred—nor did it even raise the question itself. Rather, blindness was treated as a series of errors or oversights that simply occurred, but could be avoided if planners were less context free, if staff kept their focus on long-term goals, and if appropriate short-term evaluation measurements were devised.

From an action science standpoint, *the blindness is neither the result of unfortunate circumstance nor of human folly but rather an integral part of the program theory itself.* Although the blindness is not the result of conscious planning, action science suggests that designed blindness should be considered an implicit component, or subtheory, of the instrumental program theory. Furthermore, action science assumes that there is an inherent logic to the blindness, which explains why it may be so widespread and resistant to good advice for avoiding it. Blindness is goal-oriented and dictates strategies that make good sense given the Model I reasoning of program stakeholders and the Model I behavioral world in which they function.

The lack of attention to designed blindness reflects an important gap in the theory and practice of program theory evaluation. Because program theory evaluation treats the blind-
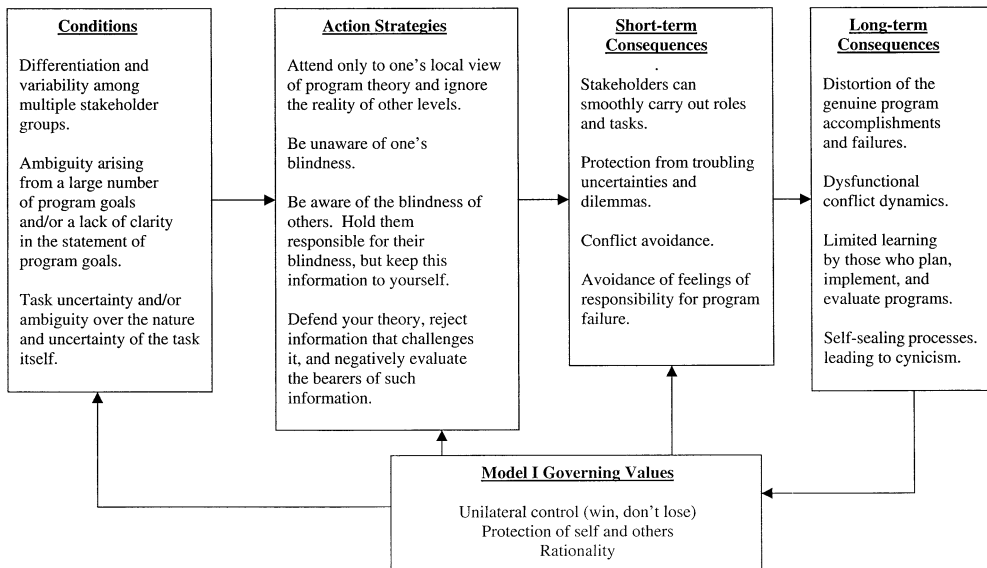
Figure 1.   A Theory of Designed Blindness.

ness as exogenous to program theory, it fails to provide an adequate theory to explain the "illusion of program failure." Although this component of program theory is seldom made explicit, it can have a significant impact on implementation and its outcomes. To the extent that program theory evaluation fails to take this kind of component of program theory into account, it is fundamentally incomplete.

To illustrate how action science can fill this gap, this article will attempt to identify the subtheory implicit in designed blindness to explain the phenomenon described in Bowen's study (Bowen, 1999, 2000). The theory will be presented through two action science "maps" (Argyris, 1993; Argyris et al., 1985; Friedman, 2000, Friedman & Lipshitz, 1994; Weick & Bougon, 1986). One map (Figure 1) presents the designed blindness subtheory inferred from the data in this case. It illustrates the conditions under which designed blindness is likely to occur as a part of program theory, the implicit goals of designed blindness (i.e., what it is meant to achieve in the short-term), the behavioral strategies through which these outcomes are achieved, and the unintended long-term consequences of these strategies. The second map (Figure 2) focuses more specifically on the theory-in-use of the case managers, the group of stakeholders in the TASA case about which Bowen's study (Bowen, 1999, 2000) provides the most information.

Short-term Goals/Consequences

The implicit goal of designed blindness is to facilitate smooth program implementation when key constituents hold very different views of its ends and means. It enables each stakeholder group to play its role, to attempt to achieve its own goals as it sees fit, and to ignore the likelihood that they are working along different lines or even at cross-purposes. By limiting the framing of a problem and its solution to a local perspective, blindness reduces perceived complexity.
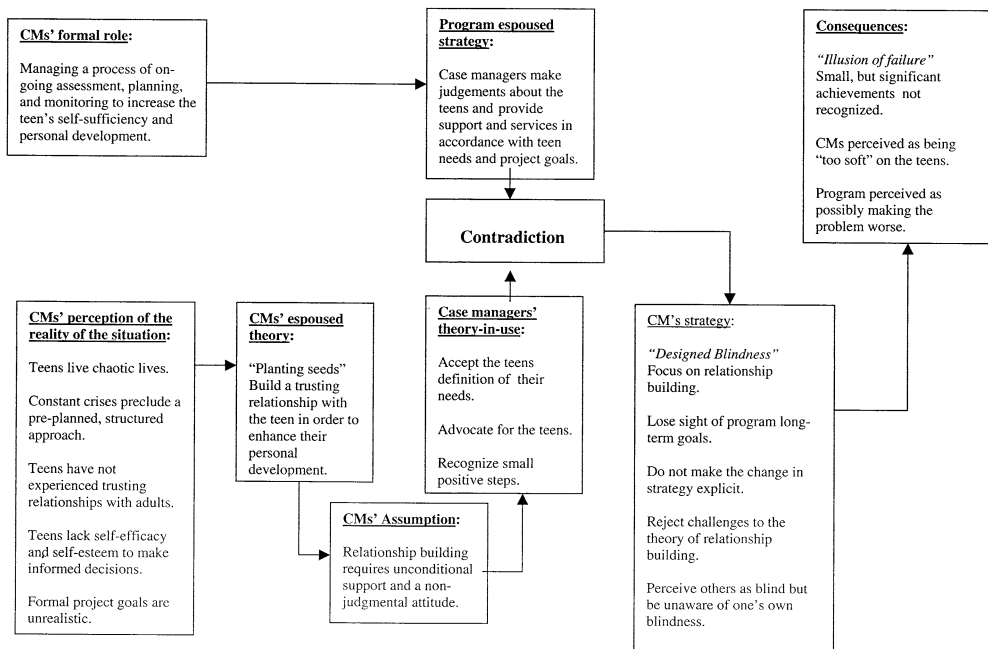
Figure 2. The Case Managers' (CMs') Theory-In-Use.

Blindness also facilitates implementation by protecting stakeholders from having to confront troubling uncertainties, dilemmas, internal contradictions, and gaps in program design. The TASA program addressed a very complex social problem for which there is no reliable, technical solution. Blindness, however, enabled the policy makers and designers to assume the existence of a well defined set of resources, activities, and techniques that, if faithfully executed, would produce the desired outcomes. In practice, however, services were delivered through an interpersonal relationship between the case manager and the teen. It would be difficult, if not impossible, to reduce this relationship to a technical process that could be standardized or measured.

Blindness enabled the case managers to manage a complex dilemma at the very center of their practice (see Figure 2). On the one hand, their formal role called for them to continually assess the teens and guide them through a structured process of goal setting, planning, and action. However, the caseworkers believed that the teens' chaotic and crisis-ridden life and lack of self-esteem made this process unfeasible. They also perceived the teens' lack of supportive relationships with adults and believed that meeting this need would "plant the seeds" for long-term personal development and goal attainment. They also believed that, to connect with the teens, they had to be unconditionally supportive and nonjudgmental. This strategy, however, contradicted their formal role, which required them to make judgments and provide resources in accordance with program goals.

To manage this conflict, the case managers focused on relationship building and lost sight of the official program goals. The case managers were aware of the planners' blindness, but there is no indication in the study that the case managers attempted to share this knowledge with the program planners or administrators. Thus, the blindness enabled the staff

to manage its conflict by simply ignoring one of their role demands while acting as if they were behaving consistently with their role.

By keeping themselves blind to each other's program theories and to the inconsistencies in their own theories, stakeholder groups avoid intergroup conflict (e.g., between planners and program staff). Conflict avoidance among stakeholders is a well-known strategy and is often considered essential for getting programs off the ground (Chen, 1990). The evaluator also became partially blind to avoid a conflict with the program staff.

Finally, blindness enables stakeholders to distance themselves from responsibility for program failure by placing blame elsewhere. For example, the case managers blamed the planners for being unrealistic and blamed other service providers for their lack of sensitivity to the true needs of the teens. The program director of TASA was probably in the best position to see the gap between the planners' and the staff's program theories. However, she blamed the poor outcomes on "the impact. . . welfare reform was having on educational attainment of participants and the unrealistic expectation that teens should enter the work-force by age 18" (Bowen, 1999, p. 29).

The goals of designed blindness are consistent with and informed by Model I theories-in-use at both the individual and the group levels (Argyris & Schön, 1974, 1978, 1996). By keeping away awareness of contradictory program theories, designed blindness enabled each stakeholder group to maintain unilateral control, or at least the illusion of control, over its task environment. By limiting the perceived complexity and uncertainty surrounding the program, blindness enabled each group to see itself as acting purposefully and rationally. For example, if the program staff seriously considered the possibility that they were being used or manipulated by the teens, they might reasonably have experienced strong feelings of anger towards their clients—the antithesis of nonjudgmental, unconditional support.

Action Strategies

Program stakeholders employ a number of strategies to develop selective perception to the point of blindness. The central strategy in the TASA case was attending only to one's own part of the program and to ignore the other levels. In this way each stakeholder group could focus on its own world view, needs, and interests without having to take into account those of others. Another strategy is for actors to be unaware of their own blindness. Unawareness enabled the TASA stakeholders to carry out their roles smoothly despite the gaps and inner contradictions. It prevented them from stopping to critically reflect on their theories-in-use and discovering their responsibility for unintended consequences.

An additional strategy is defending one's version of program theory against challenges. In the TASA case, other service providers and community members claimed that the case managers were being too lenient and making the situation worse. Even though the outcome data could support such an interpretation, the case managers not only rejected these claims but also attributed to these people a lack of understanding and caring. Similarly, when the evaluator confronted the case managers with information that could challenge their beliefs, they reacted negatively enough to cause the evaluator to back off.

Conditions for Designed Blindness

A theory of designed blindness must designate those conditions under which the phenomenon is likely to develop as a significant feature of a program's theory-in-use. The

case of TASA enables us to formulate some hypotheses as to what these conditions might be. For example, the program was legislated, planned, and carried out by three very different groups who appeared to be highly differentiated both organizationally and professionally. Thus, it is possible to hypothesize that the greater the extent to which these different functions are carried out by separate groups of people with different professional roles, organizational affiliations, physical locations, and cultures, the higher the likelihood of selective perception and blindness (Dearborn & Simon, 1958).

The case suggests that goal ambiguity might be another condition for designed blindness. In TASA, the ambiguity stemmed from its multiple goals. The program logic of TASA, as envisioned by the planners, contained 10 separate goals (Bowen, 1999, p. 6). My assertion here is that the greater number of goals, the greater the potential for different stakeholders to focus on certain goals and ignore others (March & Olsen, 1976). Furthermore, the greater the ambiguity in the statement of goals themselves, the greater the potential for interpreting them differently and not recognizing alternative interpretations.

A third condition may be the degree of uncertainty in and ambiguity of the program treatment. Perrow (1970) defined task uncertainty as a function of the variability of inputs and of the degree to which there exist clearly defined techniques or processes for transforming inputs to outputs ("analyzability"). High task uncertainty creates conditions under which different stakeholders can conceive of the same task in different ways. In the case of TASA, the degree of task uncertainty was itself ambiguous. The planners envisioned a fairly standardized, structured process of assessment, planning, and monitoring implementation. The program staff saw exactly the same task as a highly variable and unpredictable process that depended upon establishing a relationship with each unique individual.

Long-term Consequences

Designed blindness represents a form of "skilled incompetence" (Argyris, 1986) which facilitates smooth functioning in the short-term but incurs high long-term costs. Bowen cited the illusion of a failure, which could lead to the unjustified termination of a project. Whether it creates an illusion of failure or an illusion of success, designed blindness makes it very hard to see the real impacts of a program. In the TASA case, the case managers believed that their strategy was achieving small but significant gains, but these purported gains were on outcomes that were never formally assessed nor brought to the attention of other stakeholders. Instead, by adhering to their strategy of unconditional support and advocacy, the case managers were perceived as being too soft on the teens and as possibly making the problem worse.

Designed blindness also contributes to dysfunctional conflict dynamics. As the TASA case illustrates, when confronted with conflict the case managers rejected the other side's claims and judged them as lacking understanding and caring. This approach to conflict not only harmed the relationships, but also inhibited effectiveness and learning. In the short-term, blindness helped the case managers function smoothly in the face of an inner contradiction in their practice. Their strategy, however, raised the risk of being taken advantage of by the teens and exploited in ways that were counterproductive to the teens' development. By strongly rejecting the views of others, the case managers protected themselves from discovering the potential negative consequences of their strategy. The more they invested in self-protection, the more threatening it would be to discover that they were being manipu-

lated or used by the teens. Thus, the more the case managers reinforced their blindness by rejecting any information that could alert them to possible errors.

A final long-term consequence of designed blindness is that it reinforces both the initial conditions and the Model I behavioral world. The lack of clarity regarding outcomes and the dysfunctional conflict dynamics both feed back on the quality of the information and perceptions that make up the environment of similar programs. If designed blindness prevented the caseworkers' context-specific knowledge from reaching the planners, the planners could reasonably continue planning similar programs. Similarly, the persistence of Model I reasoning and the consistency with which it produces these negative long-term consequences reinforces the belief that such outcomes are inevitable. Over time the gap between planning and implementation may become accepted as part of the "reality" of social programs (McLaughlin, 1985; Pressman & Wildavsky, 1984). This belief fosters cynicism, feelings of helplessness, and a distancing from responsibility at all levels of the program.

OVERCOMING DESIGNED BLINDNESS

So far this article has attempted to illustrate how program theory evaluation can improve its descriptive and explanatory power by taking into account subtheories such as designed blindness. However, advocates of action science would not be satisfied producing a theory that simply describes and explains the phenomenon. Rather, the aim is to produce "action-able" knowledge, that is theories that enable people to produce intended consequences (Argyris, 1993, p. 1). From an action science perspective, the best way of testing the theory, and indeed of understanding designed blindness, would be changing the situation so that the problem is prevented, diminished, or solved (Argyris et al., 1985; Lewin, 1951). Thus, a theory of designed blindness would be incomplete without a theory for overcoming it.

A theory for overcoming designed blindness would assume the existence of the same initial conditions, specify the intended long- and short-term goals/consequences, and describe the action strategies that ought to produce them. A comprehensive theory for overcoming designed blindness is beyond the scope of this article, so the following discussion will focus on a strategy of "formative action evaluation" (Rothman & Friedman, 1999, 2000) that can guide the redesign of the case managers' theory-in-use (Figure 3).

Action Strategies

Having generated an empirically based map of the case managers' theory-in-use (Figure 2), action evaluators first test the validity of the theory with the actors who provided the data by presenting the map in Figure 2 and asking whether it accurately portrays their perceptions, reasoning, and action. It is important that this map be general enough to provide a comprehensive picture of the problem and specific enough that all participants see their reasoning and behavior accurately reflected in it.

If program participants dispute parts or all of the map, action evaluators make their reasoning as transparent as possible, tracing the logical steps which connect the data to the theory. They share the data upon which their inferences were drawn and test the logic of their inferences, asking the other participants whether these inferences makes sense, whether there are more plausible interpretations, and whether there are important data which may have been
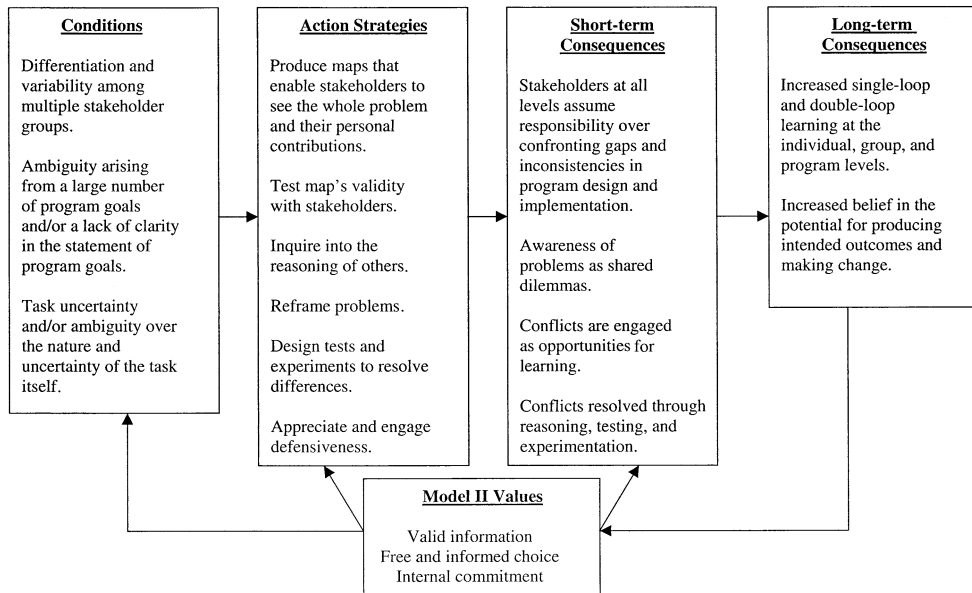
Conditions	**Action Strategies**	**Short-term Consequences**	**Long-term Consequences**
Differentiation and variability among multiple stakeholder groups. Ambiguity arising from a large number of program goals and/or a lack of clarity in the statement of program goals. Task uncertainty and/or ambiguity over the nature and uncertainty of the task itself.	Produce maps that enable stakeholders to see the whole problem and their personal contributions. Test map's validity with stakeholders. Inquire into the reasoning of others. Reframe problems. Design tests and experiments to resolve differences. Appreciate and engage defensiveness.	Stakeholders at all levels assume responsibility over confronting gaps and inconsistencies in program design and implementation. Awareness of problems as shared dilemmas. Conflicts are engaged as opportunities for learning. Conflicts resolved through reasoning, testing, and experimentation.	Increased single-loop and double-loop learning at the individual, group, and program levels. Increased belief in the potential for producing intended outcomes and making change.

Model II Values

Valid information
Free and informed choice
Internal commitment

Figure 3. A Theory for Overcoming Designed Blindness.

missed. A useful action science tool for guiding this process is the "ladder of inference" (Argyris, 1982, p. 181; Senge et al., 1994, p. 243).

Action evaluators actively seek data that disconfirms their theories and try to genuinely understand what leads others to see the same situation differently. They must also be willing to admit their errors and change their interpretations when they are convinced by the kind of testing process described above. At the same time, they encourage others to make their reasoning explicit, to be open to testing, and to have their own views disconfirmed.

Once agreement is reached on the validity of the theory presented in the map, they inquire into the reasoning behind the theory-in-use. Inquiry begins with an indeterminate, problematic situation in which uncertainty, an inherent conflict, or confusion block effective action. The inquirer probes the case managers' perception of the situation to help them make better sense of it (Argyris & Schön, 1996, pp. 31–32; Dewey, 1938; Preskill & Torres, 1999; Schön, 1983, 1987). Inquiry focuses on the case managers' construction of reality, their behavioral strategies for dealing with it, and/or their goals. It also attempts to reveal the "framing" of the problem (assisting teen mothers) and the solution (Schön, 1983; Schön & Rein, 1994).

Inquiry uncovers points of "leverage" at which relatively small changes can open opportunities for improvement (Della-Piana, 1999; Friedman, 2001; Senge, 1990). For example, inquiring into what the case managers mean by "nonjudgmental" or "unconditionally supportive" could lead to them to discover that they do make judgments but attempt to suspend them so as to give the teens a feeling of self-worth and acceptance for who they are. Furthermore, the case managers may realize that they themselves doubt whether what the teens want is really to their benefit. If so, action evaluators would suggest "reframing" the problem as how to communicate genuine acceptance of the teen while also setting limits. However, the action evaluator must also help the case managers (and other stakeholders)

define their definitions of success, design strategies for putting this strategy into action, and test these strategies out in practice (Rothman & Friedman, 1999, 2000).

The case managers strongly disagreed with other service providers' perception of their relationship-building strategy and its consequences. Action evaluators, however, would ask them to consider (1) whether their strategy has unintended negative consequences, and (2) what data might convince them that their strategy was, in fact, leading to such outcomes. On the basis of this inquiry, evaluators would design "action experiments" for generating these kinds of data (Argyris et al., 1985; Schön, 1983; Schön et al., 1984), determining the extent to which the case managers' strategy was counterproductive. In any case, this approach provides a richer, more finely grained picture of the relationship between interventions and outcomes.

If case managers dispute the description of their strategy as "blindness," saying they told the program director about the gap between the formal program and the realities of the situation, the action evaluation would ask them to illustrate how they tried to bring their knowledge to the program director's attention. This line of inquiry might reveal that they had communicated double messages which led the program director to assume that the problem was not very serious (Argyris, 1986, 1994). Here too the action evaluator would inquire into what led them to use this strategy, and help them design more effective ways of communicating what they know.

Given Bowen's experience, it is quite possible that the case managers would respond to the map by becoming defensive towards the evaluator. They might accuse the evaluator of misunderstanding them, being uncaring, or not appreciating their efforts. When program stakeholders become defensive in this way, it is critical that evaluators not be intimidated but respond in ways that increase openness to learning. For example, action evaluators might ask what they have said or done that has been interpreted as misunderstanding, uncaring, or unappreciative. In addition, they would want to make their own reasoning behind their own behavior explicit and open to testing. For example, they might say:

> It is not my intention to communicate that I do not care, but I would like to be clear about how I understand the meaning of "caring." As I see it, the most caring thing I can do is help people discover the how their own behavior is producing unintended negative consequences.
>
> I feel that I do appreciate your efforts, but my research has uncovered data that also lead me to believe that these efforts may be unintentionally creating consequences which you yourselves would find undesirable. If you are willing, I would like to discuss these findings with you to test their validity. Perhaps I have misinterpreted or overlooked important data. However, if you discover that the data support these findings, it may give you an opportunity to devise more effective strategies for achieving your goals.

These responses are informed by an "appreciation" of defensiveness as a healthy, natural reaction to being confronted with potentially threatening information. Action evaluators generally choose to engage defensiveness rather than avoid, circumvent, or overpower it. They appreciate defensiveness because it can help them discover their own errors. Action evaluators must hold themselves to the same standards as they hold the stakeholders, being willing to admit error on the basis of valid information. While making their intentions and meanings explicit, action evaluators must express a genuine desire to know how stakeholders make sense of the situation.

Short-term Consequences

This theory for overcoming designed blindness aims at increasing the quality of information upon which stakeholders evaluate success or failure. As a result of the action strategies, case managers should increasingly value challenges to their own theories as opportunities for learning. They should see problems as shared dilemmas, gaining insight into their own dilemmas and those of other stakeholders. As a result, they should be less likely to blame others and more likely to frame conflicts in terms of the substantive issues rather than in terms personality, politics, or nasty motives. These strategies should lead the case workers to assume more responsibility for confronting the gaps and inconsistencies in program design and implementation, and for attempting to resolve conflicts through inquiry, testing, and experimentation (Rothman, 1997; Rothman & Friedman, 2000, 2001).

Long-term Consequences

The desired long-term consequence of this theory for overcoming designed blindness is an enhanced ability to learn from experience so as to increase program effectiveness. Further, this approach should increase the caseworkers' belief in the potential for producing intended outcomes. From an action science perspective, the desired outcome would be both "single-loop learning," changes in action strategies, and "double-loop learning," changes in the perceptions, assumptions, goals, and values underlying those strategies (Argyris & Schön, 1974, 1978, 1996). Double-loop learning requires Model II reasoning or its equivalent and demands a significantly greater investment of time, commitment, and skill (Argyris, 1993; Argyris & Schön, 1974, 1996). A third long-term goal would be adding to general knowledge about designed blindness so that other planners, professionals, and evaluators can learn from the experience of this program.

The theory for overcoming designed blindness can be held as valid only to the extent that actors are able to produce these strategies, blindness is prevented, diminished, or eliminated, and no better causal explanations can be found. However, theory building and testing is an iterative process. Failures or partial successes in overcoming blindness are to be expected and used as the basis for ongoing reflection, refinements of the theory, and further experimentation.

DISCUSSION

One purpose of this article has been to initiate a conversation between theory-driven evaluation and action science by illustrating how action science might critique and extend a specific program theory evaluation. Toward this end, this article has focused on a program subtheory that is usually not addressed by program theory evaluation, but that can strongly influence program implementation and effectiveness. Furthermore, investigation of a particular case suggests that the evaluator herself was subject to this kind of blindness.

These findings, however, need to be qualified in a number of ways. First, the analysis in this article is based entirely upon a secondary source (Bowen, 1999, 2000). None of the findings of the original study were empirically checked nor were any additional data collected from the field. While the evaluator candidly confirmed her own blindness, its causes, and its

consequences, all of the other propositions must be regarded as hypotheses deduced from the existing data and other literature.

The finding that the evaluator unintentionally reinforced designed blindness also needs to be qualified by the fact that she was a doctoral student conducting the study as the basis for her doctoral dissertation (Bowen 1999, 2000). As she put it, her evaluation "was not asked for but rather kindly tolerated." Therefore, it could be argued that evaluator collusion in designed blindness is likely to occur mostly in situations in which the evaluator has relatively low status or in which there is relatively low commitment to the evaluation on the part of the program stakeholders.

To what extent can designed blindness be generalized to other programs and other evaluation studies? Actually this question can easily be turned around into the claim that, aside from the terminology, there is nothing new about the phenomenon of blindness in the fields of policy, implementation, and program evaluation. Evaluators and social policy researchers have long been aware of the gaps between planning and implementation. Schneider (1982) notes that converting policies into viable interventions that correspond to the original intentions of the sponsors is the most difficult problem in the area of social programs. Wholey, Scanlon, Duffy, Fukumoto, and Vogt (1971) found that federal programs typically have multiple objectives and that interpretations of both ends and means varied widely among implementors. McLaughlin (1985) argues that the notion of holistic goals is an illusion because programs are transformed as they pass through the hands of multiple actors in the "implementing system" (p. 103). According to this view, the deciding influence over policy is usually exerted by "street-level bureaucrats" (Lipskey, 1980), who deliver programs in the face of intractable dilemmas but are rarely consulted in the design or use of evaluation. Pressman and Wildavsky (1984) argue that no amount of specificity or top-down control would, or should, ensure fidelity to original goals because implementation itself is fundamentally a process of exploration. They allude to evaluator blindness in their claim that evaluators who were insensitive to this process were being "obtuse" and "only accountants" (p. 255).

The action science approach to overcoming designed blindness is also congruent with recommendations already made by a number of evaluation approaches. For example, the idea of testing a program theory-in-use, or evaluation findings in general, with both program implementers is advocated by program theory evaluation (e.g., Chen, 1990; Huebner, 2000), utilization-focused evaluation (Patton, 1997), empowerment evaluation (Fetterman, 1994), evaluative inquiry (Preskill & Torres, 1999), and learning evaluation (Pressman & Wildavsky, 1984). All of these approaches maintain that ongoing process of data collection and communication between program people and evaluators is instrumental to helping programs, organizations, or stakeholders both achieve their goals and learn.

What, then, is the additional contribution of an action science description and analysis of designed blindness? First and foremost, action science attempts to trace the link between the phenomenon at the program level and the individual reasoning and behavior in which they are rooted. In doing so, action science assumes that a phenomenon like designed blindness is "law driven" and that an approximation of these laws (what action science calls a "theory") can be inferred from a single instance or case (Lewin, 1931). Given the conditions and goals outlined in the map, we should be able to predict that people using Model I reasoning will *always* produce strategies that lead to designed blindness.

The goal of action science, however, is not simply to describe the status quo but to change it. From an action science point of view, when lawfulness is rooted in the reasoning

and behavior upon which actors can reflect and exercise volition, it does not imply inevitability or determinism. Theories-in-use are designs for achieving intended consequences given a complex, uncertain world and limited human information-processing capacity. The errors they produce, such as blindness, are a consequence of the design, but they are not intended. Theories-in-use function just below the level of awareness but, through reflection, can be brought into conscious awareness and control. The questions one needs to ask oneself is "What led me to act as I did at that point? What outcome did I intend to produce?" In retrospect, it is not unusual for people to experience surprise at not having seen the "obvious" gaps and contradictions in their own reasoning and behavior.

The focus on individual theories of action also sets the stage for an important difference in how program theory evaluation and action science view the context in which a program takes place. According to Chen (1990, p. 118), evaluators should first determine the "normative" theory for the "implementation environment" under which the program treatment can be appropriately implemented and then compare this with the actual environment. Action science, on the other hand, views the implementation environment as partially an artifact of individual and program theories themselves (Argyris & Schön, 1974, p. 23). Individual theories of action determine people's actions, which help determine the characteristics of the environment, which then influence individual theories of action. Thus, the implementation environment is constructed through a circular process of interaction between the program and its environment (Dahler-Larsen, 2000). Program actors and evaluators may not create the constraints that impinge upon their effectiveness, but they are responsible for the features of their reasoning and behavior that reinforce and shape them.

The theory for overcoming designed blindness is driven by Model II governing values of valid information, free and informed choice, and internal commitment (Argyris & Schön, 1974, 1978, 1996; Friedman, 2000). It is also based on the assumption that most people want to do their best, but often feel constrained by both internal and external obstacles. The point of action science inquiry is to help people and groups recognize that at least some of what they consider to be "reality" is actually the product of choices about how to perceive and manage their world. The more they discover these choices, the more they create potential leverage points for themselves and control over the behavioral world they construct (Friedman, 2000).

Action science offers program theory evaluators sorely needed concepts and skills for helping both stakeholders and themselves discover the theories implicit in individual and organizational behavior and to alter those that limit effectiveness and learning (Donaldson & Chen, 2001). They include tools for overcoming defensive routines (e.g., Putnam, 1993), using "productive reasoning" (e.g., Argyris, 1991), combining advocacy with inquiry (e.g., Senge et al., 1994, pp. 242–259), examining dilemmas (Argyris & Schön, 1974, pp. 99–102), engaging conflict (e.g., Rothman, 1997; Rothman & Friedman, 2001; Schön & Rein, 1994), and dealing with the defensiveness that occurs when people's mental models are challenged (Friedman & Lipshitz, 1991). There are also other related approaches, such as dialogue (Isaacs, 1993; Schein, 1993) and systems thinking (Rogers, 2000; Senge, 1990; Senge et al., 1994), which provide alternative strategies for achieving similar ends.

Program theory evaluation has been criticized for being so demanding that it cannot feasibly be done "right" in most cases (Stufflebeam, 2001, p. 39), and it can be argued that an action science approach would only increase the difficulty (Edmondson, 1996). However, as illustrated above, an action science approach aims at enhancing the effectiveness of current practice and not necessarily adding complex new procedures. There is no question that

learning to skillfully integrate action science tools into existing practice requires a high degree of commitment as well as an investment of time and energy, but the same is true for any form of professional skill.

Because action science seeks to integrate theory building into areas of social practice that are characterized by uncertainty, instability, and uniqueness (Friedman, 2000), it would probably not prove useful when a well-developed, validated program theory already exists. Nor would it be appropriate for purely summative evaluations. Action science attempts to build theories that explain social phenomena, inform practice, *and* adhere to the fundamental criteria of a science. In this respect, it offers an alternative to, though not a replacement for, a "normal science" approach to program theory evaluation (Argyris et al., 1985, p. 105; Schön et al., 1984).

There are additional areas in which a dialogue between program theory evaluation and action science could be fruitful. For example, both approaches share a concern with defining causality and how it can be determined in a complex social world (Argyris, 1993, 1997; Bickman, 2000; Cook, 2000; Davidson, 2000; Rogers, 2000; Schön & Rein, 1994). Another area of common concern involves the role of the researcher and the process of involving practitioners as active participants in the research process (Argyris et al., 1985; Huebner, 2000). Finally, action science offers concepts and methods for ensuring rigor in testing theories of practice under real world conditions (Argyris et al., 1985; Argyris, 1993; Friedman, 2000).

Both program theory evaluation and action science adhere to Kurt Lewin's (1951, p. 163) saying that "there is nothing so practical as a good theory." Hopefully this shared commitment to the usefulness of theory in everyday life will stimulate such a dialogue and enrich the study of social practice.

ACKNOWLEDGMENTS

REFERENCES

Argyris, C. (1982). *Reasoning, learning, and action: Individual and organizational.* San Francisco: Jossey-Bass.

Argyris, C. (1985). *Strategy, change, and defensive routines*. New York: Harper Business.

Argyris, C. (1986). Skilled incompetence. *Harvard Business Review*, *64*(5), 74–79.

Argyris, C. (1991). Teaching smart people how to learn. *Harvard Business Review, 69*(3), 99–109.

Argyris, C. (1993). *Knowledge for action: A guide to overcoming barriers to organizational change*. San Francisco: Jossey-Bass.

Argyris, C. (1994). Good communication that blocks learning. *Harvard Business Review, 72*(4), 77–85.

Argyris, C. (1997). Field theory as a basis for scholarly consulting. *Journal of Social Issues*, *53*(4), 809–824.

Argryis, C., Putnam, R., & Smith, D. (1985). *Action science: Concepts, methods, and skills for research and intervention*. San Francisco: Jossey-Bass.

Argyris, C., & Schön, D. A. (1974). *Theories in practice: Increasing professional effectiveness.* San Francisco: Jossey-Bass.

Argyris, C., & Schön, D. A. (1978). *Organizational learning: A theory of action perspective*. Reading, MA: Addison Wesley.

Argyris, C., & Schön, D. A. (1996). *Organizational learning II: Theory, method, and practice*. Reading, MA: Addison Wesley.

Bickman, L. (2000). Summing up program theory. *New Directions for Evaluation, 87*, 103–112.

Bowen, K. (1999). *Development of local program theory: Using theory-oriented evaluation to make a difference.* Paper presented at the Annual Conference of the American Evaluation Association, Orlando, Fl.

Bowen, K. (2000). Development of local program theory: Using theory-oriented methods to make a difference, *Field Methods, 12*(2), 129–152.

Chen, H. T. (1990). *Theory-driven evaluations.* Newbury Park, CA: Sage.

Cole, G. E. (1999). Advancing the development and application of theory-based evaluation in the practice of public health. *American Journal of Evaluation, 20*(3), 453–470.

Cook, T. (2000). The false choice between theory-driven evaluation and experimentation. *New Directions for Evaluation*, *87*, 27–34.

Dahler-Larsen, P. (2000). *From program theory to constructivism: On tragic, magic, and competing programs.* Paper presented at the annual conference of the American Evaluation Association, Honolulu, HI.

Davidson, J. (2000). Ascertaining causality in theory-based evaluation. *New Directions for Evaluation, 87*, 17–26.

Dearborn, D., & Simon, H. (1958). Selective perception: A note on the Department Identifications of Executives. *Sociometry*, 284–295.

Dell-Piana, G. (1999). Toward a context-based and dilemma-based model of program implementation theory to guide program evaluation. *Mechanisms*, *3,* 12–21.

Dewey, J. (1938). *Logic: The theory of inquiry*. New York: Holt, Rhinehart and Winston.

Diefendorf, M. (1999). *The NECTAS logic model: Its development, theory, and uses.* Poster presented at the American Association Annual Conference, Orlando, Fl.

Donaldson, S. I., & Chen, H. T. (2001). *Theory-driven program evaluation: State-of-the-art at the turn of the century.* Unpublished manuscript.

Edmondson, A. (1996). Three faces of Eden: The persistence of competing theories and multiple diagnoses in organizational intervention research. *Human Relations, 49*(5), 571–595.

Fetterman, D. (1994). Empowerment evaluation. *Evaluation Practice*, *15*(1), 1–15.

Friedman, V. (2000). Action science: creating communities of inquiry in communities of practice. In H. Bradbury & P. Reason (Eds.), *The handbook of action research* (pp. 159–170). Thousand Oaks, CA: Sage.

Friedman, V. (2001). Individual as agent of organizational learning. In D. Meinholf, A. Berthoin-Antal, J. Child, & Y. Nonaka (Eds.), *The handbook of organizational learning* (pp. 398–414). Oxford: Oxford University Press.

Friedman, V. J., & Lipshitz, R. (1992). Shifting cognitive gears: Overcoming obstacles on the road to Model 2. *Journal of Applied Behavioral Science*, *28*(1), 118–137.

Friedman, V. J., & Lipshitz, R. (1994). Human resources or politics: Framing the problem of appointing managers in an organizational democracy. *Journal of Applied Behavioral Science*, *30*(4), 438–457.

Funnel, S. (1997). Program logic: An adaptable tool for designing and evaluating programs. *Evaluation New and Comments*, 5–17.

Funnel, S. (2000). Developing and using a program theory matrix for program evaluation and performance monitoring. *New Directions in Evaluation*, *87*, 91–102.

Huebner, T. (2000). Theory-based evaluation: Gaining a shared understanding between school staff and evaluators. *New Directions for Evaluation*, *87*, 79–90.

Isaacs, W. (1993). Dialogue, collective thinking, and organizational learning. *Organizational Dynamics*, *22*(2), 24–39.

Lewin, K. (1931). The conflict between Aristotelian and Galileian modes of thought in contemporary psychology. Republished in M. Gold (Ed.), *The complete social scientist: A Kurt Lewin reader* (pp. 37–66). Washington, D.C.: The American Psychological Association, 1999.

Lewin, K. (1951). *Field theory in social science.* New York: Harper & Row.

Lipsey, M., & Pollard, J. (1989). Driving toward theory in program evaluation: More models to choose from. *Evaluation and Program Planning*, *12*, 317–328.

Lipskey, M. (1980). *Street-level bureaucracy: Dilemmas of the individual in public service.* New York: Russell Sage Foundation.

Lipshitz, R. (2000). Chic, mystique, and misconception: Argyris and Schön and the rhetoric of organizational learning. *Journal of Applied Behavioral Science, 36*, 456–473.

March, J., & Olsen, J. (1976). *Ambiguity and choice in organizations*. Bergen, Norway: Universitetsforlaget.

McLaughlin, J., & Jordan, G. (1999). Logic models: A tool for telling your program's performance story. *Evaluation and Program Planning, 27*(1).

McLaughlin, M. W.(1985). Implementation realities and evaluation design. In R. L. Shotland & M. M. Mark (Eds.), *Social science and social policy* (pp. 96–120). Beverly Hills, CA: Sage.

Patton, M. (1997). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Perrow, C. (1970) *Organizational analysis: A sociological view*. Monterey, CA: Brooks Cole.

Preskill, H., & Torres, R. T. (1999) *Evaluative inquiry for learning in organizations*. Thousand Oaks, CA: Sage.

Pressman, J. L., & Wildavsky, A. (1984). *Implementation: How great expectations in Washington are dashed in Oakland*. Berkeley, CA: University of California Press.

Putnam, R. (1993). Unlocking organizational routines that prevent learning. *The Systems Thinker,* (August), 1–4.

Rogers, P. (2000). Causal models in program theory evaluation. *New Directions for Evaluation, 87*, 47–56.

Rogers, P., Petrosino, A., Huebner, T., & Hasci, T. (2000). Program theory evaluation: practice, promise, and problems. *New Directions for Evaluation*, *87*, 5–13.

Rothman, J. (1997*). Resolving identity-based conflict in nations, organizations and communities*. San Francisco: Jossey-Bass.

Rothman, J., & Friedman, V. (2001). Conflict, identity, and organizational learning. In D. Meinholf, A. Berthoin-Antal, J. Child, & Y. Nonaka (Eds.), *The handbook of organizational learning* (pp. 582–597). Oxford: Oxford University Press.

Rothman, J., & Friedman, V. (1999), *Action evaluation: Helping to define, assess and achieve organizational goals.* Presented at the American Association Conference in Orlando, Florida, and published in the Action Evaluation web site http://www.aepro.org/inprint/papers/aedayton.html.

Rothman, J., & Friedman, V. (2000). *Action evaluation as a means of knowledge creation in social-educational programs.* Paper presented at the Annual Conference of the American Evaluation Association, Honolulu, HI.

Schein, E. H. (1993). On dialogue, culture, and organizational learning. *Organizational Dynamics*, *22*(2), 40–51.

Schmitz, C. (1999). *Everything you wanted to know about logic models but were afraid to ask.* Paper presented at the American Association Annual Conference, Orlando, FL.

Schneider, A. L. (1982). Studying policy implementation. *Evaluation Review 6*(6), 715–730.

Schön, D. A. (1983). *The reflective practitioner*. New York: Basic Books.

Schön, D. A. (1987). *Educating the reflective practitioner: Toward a new design for teaching and learning in the professions*. New York: Basic Books.

Schön, D. A. (1997). *Theory-of-action evaluation.* Paper presented to the Harvard Evaluation Task Force.

Schön, D., Drake, W., & Miller, R. (1984). Social experimentation as reflection-in-action. *Knowledge creation, diffusion, and utilization, 6*(1), 5–36.

Schön, D. A., & Rein, M. (1994). *Frame reflection: Toward the resolution of intractable policy controversies*. New York: Basic Books.

Senge, P. (1990). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday Currency.

Senge, P., Kleiner, A., Roberts, C., Ross, R., & Smith, B. (1994). *The fifth discipline fieldbook.* New York: Currency Doubleday.

Stufflebeam, D. (2001). Evaluation models. *New Directions for Evaluation, 89*, 37–39.

Suchman, E. A. (1967). *Evaluative research: Principles and practice in public service and social action programs.* New York: Russell Sage Foundation.

Trochim, W. (1989). An introduction to concept mapping for planning and evaluation. *Evaluation and Program Planning*, *12*, 1–16.

Weick, K., & Bougon, M. (1986). Organizations as cognitive maps: Charting ways to success and failure. In H. P. Sims & D. A. Gioia (Eds.), *The thinking organization: Dynamics of organization social cognition*. San Francisco: Jossey-Bass.

Weiss, C. (1997). How can theory-based evaluation make greater headway? *Evaluation Review*, *21*, 501–524.

Wholely, J., Scanlon, J., Duffy, H. Fukumoto, J., & Vogt, L. (1971). *Federal evaluation policy: Analyzing the effects of public programs*. Washington, D.C.: The Urban Institute.