EDITORS' INTRODUCTORY COMMENTS
TO CHAPTER 1

In this introductory chapter, the volume editors explore the relation-ship between social psychology and the evaluation of programs, poli-cies, and practices. They discuss the historical relationship, includ-ing the role of a set of major figures whose work involves both fields. The current status of the relationship is also examined. Today, the authors suggest, the primary form of intersection involves "program theory." A program theory is a conceptual model of how and why a social intervention is expected to bring about its anticipated bene-fits. In many domains of social and educational interventions, social psychological research and theory underlies program theories.

As Mark, Donaldson, and Campbell indicate, however, there are several other potential benefits for both evaluation and social psy-chology that are far from fully realized. Whether these benefits are achieved will depend on the future of the relationship between social psychology and evaluation. The authors of Chapter 1 consider three alternative futures, clearly preferring one over the others.

In Appendix 1.1, they also review several of the concepts and terms related to research methodology that are used throughout the volume.

The Past, the Present, and Possible Futures of Social Psychology and Evaluation

Melvin M. Mark
Stewart I. Donaldson
Bernadette Campbell

Consider the following four scenarios:

> Staff members of an international assistance agency are developing a new program. It is intended to help women in a traditional culture develop the ability to insist on safe sex practices with their (often unfaithful) partners. In developing the program, the staff members refer to the literature on behavior change and build a plan for the program that draws on concepts such as social norms and self-efficacy.

> Members of a research firm are hired to evaluate the effectiveness of a workplace-based program that was put in place to reduce employees' unhealthy behaviors (and ultimately to improve health and reduce health care costs). In designing the evaluation, the researchers discuss theories of interpersonal influence and behavior change processes to help them identify shorter-term outcomes to measure. The researchers also look at summaries from research on the best ways to get valid self-reports about behavior (such as diet and exercise) that takes place over time.

> A researcher interested in theories of negotiation is asked to help with and study a contentious process in the person's local school district. The school is trying to implement a policy to "mainstream" children

with serious disabilities into regular classrooms. While observing meetings with various interested parties, the researcher realizes that traditional academic research on negotiation has left out an important factor—the degree to which participants identify with multiple interest groups in the negotiation (like the teacher who is also the mother of a special needs student).

The United States Government Accountability Office (GAO) is often asked to study the implementation and effectiveness of government programs and policies. When hiring new staff members, the GAO seeks new PhDs with skills in research design, understanding of behavior change theories, and the ability to measure attitudes and behaviors.

All four of these scenarios, and many others described throughout this book, involve the intersection of social psychology, on the one hand, and the evaluation of social programs, policies, and practices, on the other. The historical linkages between these two areas are noteworthy. For example, Kurt Lewin, one of the most important figures in the history of social psychology, carried out what we would call evaluations. A prominent example involves his work in developing and evaluating efforts to modify food choices during World War II (Lewin, 1947, 1948). Later, in the 1960s and 1970s, the United States saw rapid growth in the practice of evaluation. This expansion largely took place during the Great Society, when the birth of new social and educational programs (such as Head Start) was accompanied by mandates to evaluate them. Among the leading figures during that growth spurt in evaluation were social psychologists such as Don Campbell and Peter Rossi (Shadish, Cook, & Leviton, 1991), whose roots in social psychology and contributions to evaluation are described shortly.

The current volume explores various aspects of the relationship between social psychology and evaluation. This includes discussion and examples of how beneficial the relationship can be. The volume also identifies ways the social psychology–evaluation relationship has not yet fulfilled its promise and ways this relationship might fruitfully be enhanced in the future.

This chapter introduces and overviews the social psychology–evaluation relationship. After providing definitions of both fields, we briefly review how this relationship developed and how it has changed over the years. We offer our views on the relationship as it currently stands. We close the chapter by speculating about alternative futures for the social psychology–evaluation relationship. This includes steps that we believe might be taken to help bring about a more beneficial relationship, both for evaluation and for social psychology. An appendix to the chapter provides a selective review of research methods used in evaluation and of a few key concepts from social psychology; this appendix serves as a kind of glossary for many of the concepts and terms used throughout the volume.

Social Psychology, and Evaluation, Defined

Numerous definitions of social psychology have been offered through the years. Many have described social psychology as the scientific study of the way the thoughts, feeling, or actions of an individual are affected by the real or implied presence of others. Other definitions focus on the interplay between the individual and the group. In part, what the field of social psychology is can be inferred from the topics that its members study. Social psychology has long addressed a wide range of topics, such as person perception, interpersonal attraction, helping, aggression, and prejudice and stereotyping.

Many of the topics addressed by social psychology are relevant to evaluation, as well as to the design of the programs, policies, and practices that are evaluated. Social psychological topics of interest to evaluation include the general processes that affect behavior change; attitude change and persuasion; the effects of perceived norms; biases that can affect human judgment; individual and collective decision making; and interpersonal and intergroup processes.

Social psychology has roots in both sociology and psychology, with most social psychologists trained in psychology departments. Social psychologists are interested in developing and testing theories. They do so both to better understand social psychological phenomena (so-called basic research) and to apply research and theory to address social problems (so-called applied research). In fact, much of social psychology through the years has defied categorization in terms of basic versus applied. For example, researchers have striven both to enhance understanding of fundamental social psychological processes and to address issues of societal import, such as obedience to authority (Milgram, 1974), failure to help others (Darley & Latané, 1968), and prejudice and discrimination (Dovidio, Glick, & Rudman, 2005).

Social psychologists employ a wide range of methods, including an increasing array of complex statistical procedures. However, it appears that psychological social psychologists, at least as represented in the major mainstream journals, tend primarily to use randomized experiments, in which participants are assigned at random, as by the flip of a fair coin or a random number program, to one of two or more conditions (e.g., to a positive or negative mood induction). These experiments are typically conducted in laboratory settings with introductory psychology students as study participants (Cook & Groom, 2004). As noted later in this chapter, an increased relationship between social psychology and evaluation could help overcome some of the limitations that arise from these methodological predispositions.

Like social psychology, evaluation has been defined in various ways. Perhaps the most common definition today, popularized by Michael

Scriven (1991), defines evaluation as the systematic determination of the merit, worth, or significance of something. For example, is a school-based pregnancy prevention program worthwhile? Is it worthy, metaphorically speaking, of a grade of A, or B, or perhaps an F? Another kind of definition, popularized by Peter Rossi and his colleagues (e.g., Rossi, Lipsey, & Freeman, 2004), describes evaluation as the application of social science research methods to the study of social programs or policies, in order to improve understanding and guide action. A typical illustration would be the use of a well-designed experiment to see whether the pregnancy prevention program in fact led to fewer pregnancies, with the hope that the findings would inform policy makers who make choices about school-based programs.

A wide range of methods is used in evaluation. The choice of methods should depend in part on the purpose of the evaluation (as well as other considerations, such as the pragmatics of budget and timeframes). Two of the common purposes of evaluation are formative—that is, the evaluation is intended to help improve the thing being evaluated—and summative— that is, the evaluation is intended to generate a bottom-line judgment about the thing being evaluated (Scriven, 1967). For example, a summative evaluation might be undertaken to help program staff improve the program, perhaps by identifying ways of doing a better job of implementing it and by getting more of those who are eligible to participate. The evaluator in this example might engage in direct observation of the program, along with interviews of program staff, clients, and prospective clients. Alternatively, a summative evaluation might be conducted in order to inform public officials about whether to fund the expansion of a pilot program that currently is operating at only a few sites. In this case, the evaluator might carry out an experiment or other kind of study to see what effect, if any, the program has on the outcomes of interest. Many other possibilities exist in terms of evaluation methods, as is illustrated throughout this volume.

Before turning to a brief history of the relationship between social psychology and evaluation, a note about terminology is needed. To this point, we have often referred to "the evaluation of program, policies and practices," rather than simply to "evaluation." There are two reasons we have used this long phrase, with one reason coming from social psychology and the other from evaluation. First, for social psychologists, the concept of attitude refers to the degree an individual likes or dislikes something—that is, to an individual's evaluative judgment about some object. In social psychology, the term *evaluation* is sometimes used to refer to these individual-level attitudes or evaluative judgments. To be clear, we are not focused on these individual-level evaluations, but rather on systematic research studies that contribute to sensible evaluative judgments about policies, programs, and practices. Second, evaluators recognize that in principle it is possible to evaluate anything. Product evaluation and personnel evaluation are two

other broad areas of evaluation practice, for example. By referring to the evaluation of programs, policies, and practices, we explicitly restrict the focus of the book. In particular, we are referring to systematic evaluation of programs, policies, and practices. For the most part, we ignore the evaluation of personnel, products, or anything other than programs, policies, or practices. Stylistically, to continually repeat this trilogy, "programs, policies, and practices," would soon be redundant, if not outright irritating. Usually, then, for the sake of simplicity we and the other contributors to the volume refer simply to "evaluation," even though we mean the systematic evaluation of programs, policies, and practices. Likewise, for simplicity's sake we will often refer to the "program" being evaluated, rather than the "program, policy, or practice."

A Brief and Selective History[1]

Even a brief review of noteworthy historical connections between social psychology and evaluation cannot ignore the role of Kurt Lewin. Lewin is widely recognized as a central figure in the development of social psychology. Indeed, he is considered by many to be the founder, or father, of modern social psychology. Rooted in Gestalt psychology, which maintained that the whole is different than the sum of its parts, Lewin's "field theory" emphasized the dynamic interaction between the individual and the social context within which the individual's behavior occurs. In social psychology, Lewin is credited, among other things, for initiating the study of group dynamics, for stimulating the work that led to balance theory, dissonance, and attribution theory, and for highlighting the value of integrating so-called basic and applied research.

Lewin had an unwavering commitment to blending theory and practice—he once said, "Research that produces nothing but books will not suffice" (Lewin, 1946). Lewin's contributions to the field of evaluation are not as widely recognized as his contributions to social psychology, but they are important. Indeed, Lewin's work foreshadowed many developments that took years to take root in evaluation. For example, Lewin's focus on using theory in any applied research was an intellectual predecessor to the attention to program theory in much of evaluation practice today (see Chen, 1990, and Donaldson & Crano, Chapter 5, this volume). Lewin's attention to social context presaged growing concern for context in contemporary evaluation (Rog, 2009), including attention to identifying for whom and in which settings a program is effective (e.g., Pawson & Tilley, 1997).

Perhaps the contribution to evaluation for which Lewin is most widely credited centers on what he named "action research." Defined broadly, action research is the process by which practitioners (perhaps aided by a researcher) attempt to study their problems scientifically in order to guide,

correct, and evaluate their decisions and actions (Corey, 1953). Lewin's legacy of action research has persisted more in related fields other than in (psychological) social psychology. These include organizational development, community psychology, and education. Drawing on these fields and on Lewin and his colleagues' original work, action research has historically been and increasingly today is a key influence for some approaches to evaluation. In terms of evaluation, action research is often cited and sometimes used by those interested in *improving* programs and practices, that is, formative evaluation, rather than conducting a summative, bottom-line test of their effectiveness (e.g., Rogers & Williams, 2006). Several contemporary evaluation scholars acknowledge Lewin's direct influence in their theorizing. Without going into detail here, the heritage of action research can be seen in several approaches to evaluation, including Fetterman's (1998) empowerment evaluation, Rothman's (1997) action evaluation, practical participatory evaluation (Cousins & Whitmore, 1998), and developmental evaluation (Patton, 2008).

Lewin's action research spanned a broad range of topics, several of which are relevant to evaluation. Take but one prominent example. During World War II, the U.S. Defense Department commissioned Lewin, along with Margaret Mead and dozens of other prominent social scientists in the United States, to help solve a potential food shortage. Much of the domestic meat supply was being shipped abroad to feed soldiers and allies, creating concern about a shortage of protein for domestic consumption. Government officials thought they could solve this problem by getting Americans to eat so-called variety meats or organ meats, such as hearts, brains, intestines, and heads. Lewin and others were asked to figure out how to get Americans to eat organ meat and enjoy it. In his successful studies, Lewin found that the key to changing the eating habits of individuals was first to change the perspectives and practices of the group (Lewin, 1947, 1948). Lewin subsequently applied this mechanism to efforts to address other social problems, including prejudice, criminal behavior, and work productivity (Lewin, 1948).

In this work, Lewin foreshadowed theory-driven evaluation and its cousins. In short, theory-driven evaluation involves identifying the processes by which a program should have its effects, and using this model to guide evaluation activities. Theory-driven evaluation is described in more detail in chapters that follow, especially Donaldson and Crano (Chapter 5). This relationship between Lewin and theory-driven evaluation should not be surprising. After all, Lewin (1951) was the source of the oft-repeated quote, "There's nothing as practical as a good theory."

In sum, Kurt Lewin not only founded modern social psychology but also had a formative impact on the field of evaluation. Indeed, we would argue that his impact on evaluation was greater than is generally recognized. In part, this is because his contributions largely occurred before evaluation

existed as a distinct area of practice. In addition, Lewin's contributions to evaluation are underestimated because many of them have been indirect. That is, they have taken place by way of Lewin's influence on others who in turn influenced evaluation.

More evident to many evaluators is the influential role of Donald Campbell as well as various of his associates at the Northwestern University (Oral History Project Team; Miller, King, Mark, & Stockdill, 2003). Although Campbell is known to most evaluators for his methodological contributions, such as the detailing of quasi-experimental designs and the taxonomy of kinds of validity (see Appendix 1.1), he was a notable social psychologist. His wide-ranging work included research and theory on prejudice, ethnocentrism, and intergroup relations (e.g., Campbell, 1965, 1967), interpersonal perception, including the biasing effects of one's own attitudes (Miller, Campbell, Twedt, & O'Connell, 1966), the intergenerational transmission of norms (Jacobs & Campbell, 1961), and attitude measurement (e.g., Campbell, 1950). In evaluation, in addition to being a persuasive advocate of experimental and quasi-experimental methods, Campbell was also instrumental in bringing a number of colleagues and students (and some of their colleagues and students) into the field of evaluation. For those interested in more detail, Shadish, Cook, and Leviton (1991) offer a thorough analysis of Campbell's theory of and contributions to evaluation.

Campbell was among the people who indirectly brought some of Lewin's influence to evaluation. While Campbell was a graduate student at Berkeley, he took three courses from Lewin (who was visiting professor), including a small seminar on Lewin's theories (Campbell, 1988). Like Lewin, Campbell was drawn to the integration of theory and application. However, the theory Campbell brought to evaluation was not program-specific theory about how behavior change can be achieved in a particular case. Rather, Campbell brought to evaluation a broader theory of human knowledge processes, of the biases that can affect the informal evaluations people make of a program, and of methodological theory that could be brought to bear to reduce biases in evaluation. It was a theory that supported the use of experimental and quasi-experimental research designs to evaluate programs, so as to avoid the erroneous conclusions that could otherwise arise when attempting to estimate the effects of a program.

Another important kind of intersection between social psychology and evaluation is personified by Peter Rossi, who represented the link between more sociological social psychology and evaluation. A sociologist, Rossi came from the "social problems" approach to social psychology. The issues that Rossi studied during his highly productive career include homelessness (e.g., Rossi, 1989), social welfare (e.g., Rossi, 1998), public perceptions of criminal sentences (Rossi & Berk, 1997), crime, public subsidies, and recidivism (Rossi, Berk, & Lenihan, 1980), and gun control (Wright & Rossi, 1994). For Rossi, attention to evaluation was a relatively simple extension

of his other work as an applied social researcher. Among Rossi's many contributions to evaluation are his best-selling textbook (now Rossi, Lipsey, & Freeman, 2004) and his advocacy of theory-driven evaluation in articles and chapters with Huey Chen (which preceded Chen's widely cited, 1990, book on the topic). Shadish, Cook, and Leviton (1991) provide a detailed discussion of Rossi's theory of and contributions to evaluation, and they credit him with developing a model of the different types of evaluation that should be undertaken under varying circumstances.[2]

Although it would be possible to examine the intersection of social psychology and evaluation through many individuals and institutions beyond Lewin, Campbell, and Rossi, we will turn our brief historical review to another question. That is, was it purely historical accident, or was there something more meaningful that led to the interplay that occurred between social psychology and evaluation?

Certainly, it appears that aspects of happenstance were involved. For example, the nature of food shortages on the home front in World War II led to Lewin being asked to develop and evaluate interventions to change eating patterns. In the case of Campbell, both he (Campbell, 1984) and Shadish et al. (1991) have told the story of how Campbell came to be involved in evaluation—by accident. E. A. Suchman (1967) wrote a seminal book on evaluation that highlighted Campbell and Stanley's (1966) work on experimental and quasi-experimental design and on internal and external validity. After Suchman's book, Campbell was widely sought as an expert in evaluation. Suchman, his book, and its aftermath essentially pulled Campbell into the field of evaluation.

If happenstance sometimes contributed to a *particular* connection between social psychology and evaluation, there were also sound reasons that made the *general* connection compelling. For example, Suchman's book may have pulled Campbell into the field, but Campbell did not drag his heels and resist. Rather, he quickly became a major, senior figure in evaluation. Perhaps this social psychologist entered the field of evaluation because there are many sensible reasons for an intersection between social psychology and evaluation.

- There is a good fit between the methodological and measurement skills of social psychologists and many of the practice needs of evaluation (with Suchman's attention to Campbell and his experimental and quasi-experimental design work being but one important example).

- Another kind of fit exists, in terms of the motivation for joining a research community. Many people in social psychology were drawn to the field because of applied interests and the concern for social betterment. This motivation was quite compatible with doing evalu-

ation, with its potential consequences including program improvement and the selection of more effective programs.

- Social psychological change processes underlie many if not most social programs and policies. As a result, social psychologists have a head start in thinking about program theory—that is, the underlying rationale as to how and why the program is expected to operate.

- Pragmatically, career options were needed at a time period when graduate training in social psychology was growing beyond the capacity of the traditional, mostly academic job market to offer positions to newly trained professionals.

Given the compatibility between social psychology and evaluation, and perhaps especially given the desirability of a new employment pathway for growing numbers of PhDs, many social psychologists and other applied social scientists moved into evaluation during the 1960s and 1970s. Despite the forces that attracted social psychologists to evaluation, however, the link between the two declined sharply in the 1980s. This decline appears to be attributable largely to two factors. First, budget cuts of the early Reagan administration resulted not only in the elimination or reduction of many social programs. These cuts also seriously set back the then-growing field of evaluation (Shadish et al., 1991, p. 27). At the very least, the Reagan era budget cuts probably stemmed the inflow of young social psychologists (and others) into the previously growing field of evaluation.

We believe a second factor also contributed to a weakened relationship between evaluation and social psychology: the so-called cognitive revolution in social psychology (e.g., Fiske & Taylor, 1984). The social cognition movement had several features that increased the distance between social psychology and evaluation. Specifically, the social cognitive movement served to decrease the status and frequency of applied research in mainstream social psychology; focus social psychologists on "cold" cognitive processes, which in general may be less relevant to social issues and social programs than are "warmer" motivational and affective processes; increase use of methods from cognitive psychology, such as response time and recall measures, which are not as applicable to evaluation as is the measurement of attitudes and behaviors; and lead to corresponding changes in graduate training in ways that meant newly trained social psychologists had skills that were less appropriate for work in evaluation (skills in measuring response time not being too relevant for evaluation, for instance). The budget cuts under Reagan can probably be credited for disrupting the early relationship between social psychology and evaluation. However, the social cognitive revolution may have helped preclude a speedy resumption of the relationship from occurring later on, when funding for social programs and evaluation picked up again.

The split between social psychology and evaluation was of sufficient magnitude that some people left the field in which they had been trained. For example, in his chapter with Manuel Riemer (Chapter 4), Leonard Bickman describes his separation from social psychology during many years of doing evaluation. Others drifted from mainstream social psychology, doing applied research and evaluation in a content area such as health or criminal justice. Today seems ripe for a strengthening of the relationship. As an example, despite his having left social psychology earlier, Bickman and his co-author Riemer detail their use of social psychological theory and research findings in constructing an intervention. In Chapter 13, Johnson, Smoak, and Boynton describe the benefits of a stronger integration of the applied research and evaluation in health psychology with the theory building and testing of more basic social psychology. Before examining the promise of the future, however, let's consider the present a bit more.

The Present

In principle, the relationship between social psychology and evaluation can be bidirectional. Each side can profit from the other, perhaps equally. However, our strong sense is that today evaluation draws substantially more from social psychology than it gives back. Most notably, social psychology provides a set of theories and concepts that translate into theories of the program, policy, or practice evaluated. Other kinds of exchanges, in either direction, are not as frequent.

Consider the "program theory" connection. Since the early heyday of evaluation in the United States, most evaluators have become far more attentive to program theory, that is, to the assumptions that underlie the program, policy, or practice being evaluated. In contrast, the Great Society days provided many examples of what are now called "black box" evaluations. That is, the program was treated as a black box into which the evaluator did not peek. A Campbell-styled evaluation would tell you, for example, *whether* Head Start resulted in increased test scores, but not *why*. Nor would a black box evaluation even tell you what the Head Start program actually consisted of, in terms of the specific activities that children experienced. As a result, when a program did not work, the evaluation gave no information as to why it was ineffective. Was some aspect of the underlying rationale simply wrong? Was the underlying rationale sound, but implementation poor? Black box evaluations didn't give guidance as to what to try next when the program didn't work. When programs did work, there was no learning about underlying processes that might guide action in other settings. Indeed, without understanding more about what the program actually consisted of and how it worked, it was dubious to generalize evaluation findings to other settings.

Over time, evaluators have learned to give considerable attention to the (usually implicit) model specifying why the program is expected to make a difference. There are several, somewhat overlapping ways that evaluators have tried to capture the rationale that underlies a program, policy, or practice. These are sometimes combined under the broad umbrella of "theories of change."

One approach is to build a "logic model." In practice, logic models consist of listings of inputs (the resources that go into a program), activities (the things a program does, such as offer certain services), outputs (how many people receive which services), and short-term, intermediate, and long-term outcomes. A well-constructed logic model represents a string of if-then relationships that, if they prove to be true, will result in program success. Consider a simplified (and theoretically not very compelling example). If the resources are available, then the program activities will occur. If services are offered, many people will receive them. If people receive the services, then their attitudes will change (a short-term outcome). If attitudes change, then the targeted behavior will change (an intermediate outcome). And if their behavior changes, then their health will improve (a long-term outcome). Although this kind of if-then logic should underlie a logic model, in practice these models are typically presented as lists of resources, activities, and so on, with little if any attention to the expected causal linkages

As the term is generally used, a "program theory" lays out the expected causal connections between program activities, intermediate outcomes, and long-term outcomes. Program theories are typically presented graphically, with arrows showing the expected causal pathways from certain activities to specific intermediate to long-term outcomes. This volume shows many examples of program theories, including Figures 2.2, 3.1, and 4.3.

Evaluators often begin their work by helping to uncover or make explicit the program theory. Often this involves working with program staff and other interested parties (stakeholders), as well as reviewing program documents and observing program operations. Increasingly, evaluators are taking a more active role, playing a part in the initial development and the reformulation of explicit program theories (Coryn, Noakes, Westine, & Schröter, in press; Donaldson, 2007; Patton, 2008; Renger & Titcomb, 2002). Whether the evaluator's role is to uncover the program theories implicit in stakeholders' minds or to contribute more actively to program theory, concepts from social psychology and related disciplines play a large role in many program theories. This should be not surprising, for theories of the kind associated with Al Bandura and Icek Ajzen were designed as general theories of behavior change. The use of these theories for program design and evaluation is described and illustrated in the next two chapters, by Bandura and Ajzen, respectively. The use of social psychological theory to inform program theory is discussed in many other chapters in this book, including all of the chapters in Part II and several in Part IV.

This particular connection, with social psychological theories informing program theory, is a major contribution of social psychology to evaluation. This is not to say that social psychology is the only field whose theories can inform program theory—but it is one major source. To trained social psychologists, this contribution of their theories to the applied work of evaluation may seem rather natural and unsurprising. On the other hand, to those unfamiliar with social psychology, the way that social psychological theory can inform program design and program theory is fairly impressive. In our experience, in many program and policy areas, social psychology will largely be foreign territory to program designers, program staff, and most stakeholders—who will have an everyday understanding of many concepts of the field, such as attitudes and norms, but often will not have formal training in the relevant theory and research. In such circumstances, having an evaluator who is well versed with social psychological theory and research can lead to program theories that are better grounded in the research base of social psychology.

In short, the behavior change theories from social psychology can have considerable benefit for program design and for theory-driven program evaluation. We hasten to add, however, that social psychological theory is probably not sufficient in and of itself. Indeed, several of this book's early chapters both describe the value of social psychological theory and point to the need for other kinds of theories or models for guidance. The need commonly exists, for example, for guidance that can aid in the translation of abstract social psychological concepts into concrete program activities. Bandura illustrates this need nicely in his discussion of how to translate the general concepts from his theory into specific scenes for the characters in a radio or TV serial. Need also often exists for guidance about such things as how to integrate a new program or practice into the ongoing routines of an organization, how to ensure high-quality implementation, and how to facilitate sustainability or maintenance of the program over time. Reimer and Bickman illustrate several of these considerations. In short, then, social psychology is likely to provide a good general map as to how a program is supposed to change behavior and other outcomes, while other kinds of theories and knowledge might be useful for such matters as making the program appropriate to the context in which it is being implemented and making it sustainable.

In summary, social psychological theory can play an important role in the major evaluation tasks for articulating and testing sound program theory. Notably, there are other important ways that social psychology can enhance evaluation. Practicing evaluators face a range of challenges, and social psychology potentially can contribute solutions to many of these. For instance, a classic concern in evaluation involves "evaluation use" (e.g., whether evaluation findings are used by policy makers in their decisions or affect stakeholders' understandings of the problem or its potential solu-

tions). The large social psychological literature on attitude change and persuasion is quite relevant to the topic of evaluation use, as discussed by Fleming in Chapter 8. As that chapter suggests, applying the literature on persuasion to evaluation holds potential both for understanding and for increasing use (see also Mark & Henry, 2004).

To take another example, social psychological literatures such as those on negotiation and accountability can offer considerable insight into stakeholder processes and dialogue in evaluation. See Campbell and McGrath in Chapter 13 for more discussion of these topics. Similarly, Donaldson, Gooler, and Scriven (2002), Taut and Brauns (2003), and Stevahn and King (2009) have drawn on the social psychological literature on evaluation apprehension and test anxiety to illuminate the anxiety that stakeholders often have about evaluation and to discuss ways to alleviate this anxiety. Tindale and Posavac in Chapter 7 offer a related discussion about increasing trust among those involved with evaluation, drawing on the literature in interpersonal relations. Sanna, Panter, Cohen, and Kennedy (Chapter 6) draw on the social psychological literature on judgment biases, specifically time-related or temporal biases, to address several potential challenges in program design and evaluation. As another example, Schwarz and Oyserman (2001; see also Chapter 9, this volume) describe the implications for evaluation of research on the cognitive and social psychological processes that are involved in answering survey and self-report questions. Our view is that, although several articles and chapters have mined the social psychological literature for its insights for practical challenges in evaluation, much more could be done. Part III of this book includes several efforts to extend social psychology's contributions to specific challenges of evaluation practice.

Contributions can and should also take place in the other direction. That is, evaluations from many areas of practice should have implications for social psychological research and theory. In some instances, such as evaluations of a variety of health-related interventions, relevant findings have made their way back into social psychology and related literatures. Indeed, in some areas of public health and health psychology, little if any distinction is made between evaluation and other forms of research. For the most part, however, lessons from evaluation have not been consolidated and fed back into social psychology. Each chapter in Part IV focuses on a particular area of practice in which evaluation findings have, or could, feed back into social psychology. For example, Johnson, Johnson, and Stevahn (Chapter 11) detail how evaluations of cooperative learning programs and other educational interventions have had important implications for the social psychological literature on cooperation and competition. Cialdini, Goldstein, and Griskevicius (Chapter 10) discuss what social psychologists can learn from evaluations of environmental interventions. Johnson, Dove, and Boynton (Chapter 12) contend that evaluation and other applied

research can profitably be tied back more strongly to mainstream social psychology.

Lessons from evaluation need not come solely from evaluation findings. Sometimes the lessons can be conceptual, taking place even before evaluation findings are in. For instance, attempts to apply social psychology to evaluation can demonstrate gaps in social psychological theory and research. In one example, Campbell and Mark (2006) applied the concept of accountability to interactions between evaluation stakeholders with different interests. Application to this aspect of evaluation revealed a gap in the accountability literature. Specifically, the accountability literature in social psychology had focused on whether or not a person perceived he or she was accountable to others. In considering the evaluation process, Campbell and Mark saw that the practical issue instead often was *to whom* a stakeholder participating in evaluation processes felt he or she was accountable. As this example shows, social psychology can benefit from considering evaluation other than by actual evaluation findings. That is, Campbell and Mark conducted an experiment apart from any program evaluation, to test the effect of different forms of accountability on the quality of stakeholder interactions. We believe that increased involvement in research on evaluation may benefit social psychology as much as or more than evaluation findings per se.

In summary, the current relationship between social psychology and evaluation can be characterized as follows. First, social psychological research and theory is often drawn on when evaluators attempt to discover or build a program theory. This is a noteworthy contribution. Second, social psychology is occasionally used as a source of potential answers to the challenges evaluators face in practice, such as how to deal with the anxiety that the stakeholders in an evaluation may feel. We believe that social psychology could be applied far more extensively to the varied practice challenges of evaluation. Third, in principle social psychology should be able to benefit from evaluation, but this direction of contribution has largely been limited to a few areas such as cooperation and competition. In some instances, evaluation findings themselves should be of interest, while in other cases social psychology might be integrated with research on evaluation.

Alternative Futures
for the Social Psychology–Evaluation Intersection

Going beyond the past and the present, we want to briefly explore alternative possible futures for the relationship between social psychology and evaluation. Of course, one possibility is that the connection between social psychology and evaluation will dissipate in the future. Perhaps it is wishful thinking, but we judge this to be the least likely future direction. The value

of social psychological theories to program theory seems strong and likely at least to persist. A variant on this scenario could take place, though. That is, the presence of people in evaluation who were trained in social psychology could decline by attrition, even as others draw on social psychological theories as a source of sound program theory.

In another, quite feasible future, the relationship between social psychology and evaluation will remain much as it is. That is, social psychology will continue to contribute substantially to the development of program theory. Occasionally evaluators will draw on specific aspects of the social psychological literature as a guide to a particular challenge in evaluation practice (e.g., reducing anxiety about being evaluated). In a few select areas, findings from evaluation may feed back into the social psychological literature, identifying gaps or enhancing the validity of findings by examining a social psychological question in a real-world setting. There may even be occasional emigrants, individuals trained as social psychologists who move to a career in evaluation. But in this possible future, most of the fertile ground at the intersection of these two fields, as at present, will go unexplored.

In yet another alternative future, the relationship between these two fields will be enhanced. This would have several potentially important benefits, several of which we have alluded to already.

• Theory-driven evaluation would benefit from the continued and increased importation of theory and findings from social psychology for the purpose of developing sound program theory.

• In return, social psychology's theory and evidence base would be strengthened by (a) assessing its fit and its modification when translated into program theory and (b) the findings from real-world evaluations that test aspects of social psychological theory.

• Evaluation practice would be strengthened by the importation of social psychological research and theory that is applicable to a wide array of challenges that arise in the conduct of evaluation. Examples include challenges in guiding interactions among stakeholders who vary in power, developing and maintaining trust, and measuring behaviors that take place repeatedly over time, to name but a few. Social psychology should be able to help address such practice-based challenges, as illustrated in several of the chapters that follow.

• Social psychologists interested in evaluation could help meet the need for more systematic study of evaluation (Mark, 2008). For instance, social psychologists might be well positioned to study evaluation influence and use (Mark & Henry, 2004), leading to a better knowledge base to guide evaluators' actions. As another example, rather than simply import lessons from social psychology about how to engender trust, an experiment could be conducted to study the presumed lesson in an evaluation context.

- This future would allow tests of selected social psychological hypotheses in the real-world settings of evaluations. For example, most of the past research on empathy and power has been conducted in artificial laboratory settings. Dialogue between members of different stakeholder groups offers a more realistic setting for such research, and is important in its own right. The often artificial settings and restricted participant populations of social psychology studies are a cause for criticism (Cook & Groom, 2004). The criticism involves both construct validity (i.e., concern about whether the lab experiment is studying the real phenomenon, such as cooperation) and external validity (i.e., concern about whether the findings can be generalized elsewhere). These criticisms would largely be muted for social psychological research coming from evaluation contexts.

- Both fields would be strengthened by the opportunities for joint training.

- Recently trained social psychologists, especially those with a good grounding in evaluation, can help meet what appears to be an expanding market for evaluation (American Psychological Association, 2005).

- Theory generation in social psychology would be improved. In the Campbell and Mark (2006) example, it became apparent to these researchers that the social psychological literature had generally treated accountability as an either/or, with the person either being accountable or not. In contrast, within stakeholder situations in evaluation, the issue instead often seemed to be *to whom* the stakeholder felt accountable. As this example suggests, the real-life crucible of evaluation practice should help in the development of better, more comprehensive theories.

Fortunately, we believe that conditions may be conducive to a resurgence of interest in evaluation among social psychologists. Relative to the early days of the social cognitive revolution, there has been a return to the historically strong interest in applied work within social psychology, which would seem to bode well for involvement with evaluation. In addition, the continued growth of theory-driven evaluation has reduced the distance between social psychology and evaluation. As we noted previously, much program theory is social psychological in nature. Moreover, the theory-driven approach to evaluation blurs the distinction between basic and applied research because basic theory is tested in the context of the applied work of evaluation. This should increase the overlap between social psychology and at least those evaluations that involve tests of program theory based in social psychology. Perhaps less powerful than the program theory connection but still potentially important is the continuing need for guidance regarding the challenges that arise in the practice of evaluation—and many of these are fundamentally social psychological in nature. For example, challenges involving power, stereotyping, and negotiation arise when evaluators work with representatives from multiple stakeholder

groups to plan an evaluation. Testing hypotheses from social psychology in the context of these practice challenges should not only provide better guidance to evaluators, but also help quell criticisms about the artificiality and questionable validity of much social psychological research. Finally, the pragmatics of employment trends—with limited job opportunities in traditional academic areas of basic social psychology, but a strong market in evaluation and applied social psychology (Donaldson & Berger, 2006; Donaldson & Christie, 2006)—should make the strengthened connection between social psychology and evaluation more appealing to many, including graduate students in social psychology.

We return in the final chapter to some ideas about strengthening the future relationship between social psychology and the evaluation of programs, policies, and practices. For the moment, suffice it to say that the future of this relationship may in fact depend on some of the readers of this book.

APPENDIX 1.1.
Overview of Selected Methods and Concepts

This appendix is provided to serve as an introduction to and review of several concepts and terms that are used elsewhere in this volume. Our focus here is on research methods. Numerous other concepts and terms are described in the chapters and in the introductions to the chapters.

We begin with a question. How does one evaluate the merit and worth of a program? Scriven (1991) has offered a kind of general logic for evaluation. In short, it involves (1) identifying criteria of merit, (2) defining standards of performance on those criteria, (3) measuring performance, and (4) synthesizing across the multiple criteria to generate a bottom-line judgment. This logic applies widely. For example, in evaluating midsize cars, one would initially identify criteria of merit, such as gas mileage, safety in a head-on collision, legroom, and so on. You would then determine standards for each of the criteria (e.g., you might rate gas mileage relative to other midsize cars, but for safety, you might set a minimum performance for a crash test). You would measure performance on the various criteria, such as road tests to calculate gas mileage and crash tests to assess safety. And you could then combine across criteria, leading you to recommend some cars over others.

In the case of programs, one of Campbell's insights is that the things people care about, the criteria of merit, typically involve the program's effects on certain valued outcomes. In evaluating Head Start or other preschool programs, people care about such things as whether the

program causes improvements in children's readiness for school, their social skills, and so on. In evaluating HIV/AIDS prevention program, people care about whether the program causes a reduction in unsafe sex practices and in the number of new HIV/AIDS cases. In short, the criteria of merit that people generally care about for programs, policies, and practices are their outcomes, the effects they have. Further, although the topic is worthy of more discussion than we will give it here, the performance standards of interest are usually relative. For example, how much better is the school readiness of children who have participated in Head Start than those who did not? How effective is one preschool program relative to others?

Ah, then, how can performance be measured? Imagine that you were going to evaluate Head Start decades ago, when there were few preschool programs and nearly all other children had home-based care, commonly with their mothers. Imagine too that you had a set of measures, such as knowledge of the alphabet, ability to count, and so on. Perhaps you might consider measuring children's performance before and after they participated in Head Start. Of course, this would be problematic. The children would be older after a year of Head Start, and simply because they were older they would probably do better even if Head Start was completely ineffective. (Campbell labeled this sort of problem as the "validity threat" of **maturation**.) You might think instead, let me find a group of kids the same age as my Head Start graduates and measure them at the same point in time. If Head Start was effective, you might think, the Head Start graduates should outperform the comparison group that did not participate in the program. Again, there's a problem. The kind of kids that ended up in Head Start may differ from those who did not. If so, these preexisting differences, not Head Start, might be responsible for any group differences on the measures (Campbell used the term **selection** to refer to problems like this). In fact, Head Start was restricted to children from families with low incomes, so the comparison group is likely to differ on income and other indicators of socioeconomic status. Hmm, you might say, let me try **matching**, that is to say, to find for each Head Start participant a child who did not attend Head Start but whose family has a similar income. The problem is that there may still be other differences that your matching didn't take care of. For example, for two families with similar incomes, perhaps the family with the Head Start participant is more interested in education or has better connections in the community.

Problems such as maturation and selection are known as threats to **internal validity**. In a program evaluation context, internal validity refers to the accuracy of inferences about the effect of the program on the

outcome variables of interest. If you simply measured kids before and after Head Start, your evaluation would have serious internal validity problems. Concerns about internal validity threats, and the inadequacy of simple approaches such as matching, is why evaluators such as Campbell and his followers (Campbell Collaboration, 2009) advocate the use of **randomized experiments**. In a randomized experiment, individuals (or other units) are assigned at random, essentially by the flip of a coin or some other random process, to groups. For example, you might use a computer program that generates random numbers to assign children either to Head Start or to a comparison group. Because the groups that result are equivalent (within certain statistical limits), randomized experiments provide a fair test of the performance of a program—though other problems can arise. In evaluation, randomized experiments are often referred to as **randomized controlled trials**, or **RCTs**.

Randomized experiments are not always feasible. It may be impractical or unethical to assign children at random to a program, for example. In such circumstances, a **quasi-experiment**—Campbell's term for an approximation of an experiment—may be the best alternative. There exists a wide range of quasi-experiments, some close to the randomized experiment in terms of internal validity, others far weaker. We will not cover the entire array here. Rather, we note that there has been a trend toward integrating quasi-experimental design with increasingly sophisticated statistical analysis. One development along these lines, referred to in several chapters of this volume, is the use of **propensity scores**. In essence, propensity scores are a form of matching on steroids. Multiple matching variables are combined statistically into a single index. (More technically, this index is created by conducting a logistic regression using the variables to predict membership in the program rather than the comparison group. The index thus captures the propensity to be in the program group, and it is used to match similar treatment and comparison group members).

As noted previously, internal validity refers to the accuracy of inferences about the effect of the program on outcome variables of interest. As Campbell and his colleagues pointed out, internal validity is not the only kind of validity one worries about in experiments and other studies that are used to estimate the effects of an intervention. **External validity** refers to the accuracy of inferences about other persons, places, and settings than those observed in the evaluation. Put differently, external validity involves the generalizability of the evaluation findings, including generalizability to the future. **Construct validity** refers to the accuracy of the abstract labels that are used to identify the treatment and outcomes. For example, there is a construct validity

problem if you refer to "Head Start" as the causal variable, when the real reason that participants in the program condition outperformed the comparison group is that they received two nutritious meals a day while at Head Start. In this instance, perhaps the treatment should instead be labeled as "nutritional supplementation."

Experiments do not in and of themselves solve external and construct validity problems. To the contrary, in some instances conducting an RCT may create external validity problems. For example, perhaps only very unusual kinds of families will consent to having their children randomly assigned to conditions in an evaluation. If so, there would be a tradeoff between internal and external validity. Skillful evaluation practice involves efforts to choose the best methods in the face of tradeoffs. Often the best approach, if resources allow, is to use different methods over time.

Increasingly, both experimental and quasi-experimental evaluations include procedures that can help strengthen external and construct validity in a single evaluation study. Tests of **moderation** assess whether a program is equally effective for different types of clients or across different circumstances. For example, you might test to see whether Head Start works better for boys or girls, or whether it is equally beneficial in urban, suburban, and rural settings. (For those trained in an analysis of variance framework, tests of moderation involve a statistical interaction.) With tests of **mediation**, one traces the steps in a program theory to see whether, for example, change in the hypothesized short-term outcomes appear to be responsible for change in the longer-term outcomes. For instance, in an evaluation of Head Start, program theory might suggest that initial improvement in both prereading skills and attention span will mediate longer-term academic performance. Various kinds of statistical procedures, such as structural equation modeling, can be used to estimate meditational models. Tests of moderation can contribute to external validity by showing the conditions under which the program is effective. Tests of mediation can enhance construct validity by clarifying what the processes are by which the program has its effects.

Even so, to the extent possible it is desirable not to rely on a single evaluation when making high-stakes decisions. **Meta-analysis** is a way of putting together the findings from multiple evaluations. Meta-analytic procedures allow you, for example, to combine statistically the estimated effects of Head Start from dozens of Head Start evaluations. Meta-analysis can also allow you to test for moderation, for example, by comparing the effects of Head Start from evaluations in urban areas, relative to the effects in rural areas, thus strengthening external validity.

Meta-analysis requires the computation of an **effect size esti-mate** for each of the individual studies that is included. Head Start evaluations might use slightly different measures of reading readiness, for example, and these would have to be put into a common metric to be combined and compared. A common and relatively simple effect size estimate is computed by subtracting the average comparison group score from the average treatment group score (e.g., how much better does the Head Start group do on reading readiness?); this figure is then divided by the pooled standard deviation (a measure of how spread out scores are within the treatment and within the comparison group). Meta-analyses typically give an average effect size combining across the multiple studies. There are general guidelines for whether an effect size is large or small, but it is generally advisable to make a judgment of size in the context of the specific program area.

There are many methods other than those reviewed in this brief Appendix. This includes a wide range of quasi-experimental designs, a similarly wide range of qualitative methods, descriptive methods such as sample surveys, and mixed method designs that integrate qualitative and quantitative methods. All of these methods have a proper role in evaluation, depending on the purpose of an evaluation, the stage and other characteristics of the program to be evaluated, the resources available and the time and other pragmatic constraints, and how much is already known. This Appendix has focused primarily on the methods that are presented, sometimes without much explanation, in other chapters in this volume. We hope the Appendix will help readers who needed general background to these methods.

NOTES

1. Portions of this and some subsequent sections in this chapter draw on Mark, Gamble, and Mills (2005).

2. In light of Lewin's seminal contributions, it is interesting to note that both Campbell and Rossi did work early in their careers on topics that echoed those of Lewin's earlier studies. Like Lewin, Rossi early on did research for the armed forces on changing food attitudes (Armed Forces Food and Container Institute, 1958). Leadership was another area in which Lewin did what we would consider to be evaluation, particularly in the evaluation of different types of leadership practices (Lewin, Lippitt, & White, 1939). Campbell also did research on leadership, specifically leadership in the Navy (described in Campbell, 1984). And the dots connect: Campbell apparently generated some of his ideas about quasi-experimentation while thinking about the effects of a given leader. The rotation of officers in and out of particular military units in theory offered a way to try to estimate a leader's effects on unit performance.

REFERENCES

American Psychological Association. (2005). *Non-academic careers for scientific psychologists: Interesting careers in psychology.* Retrieved October 15, 2010, from *www.apa.org/science/nonacad_careers.html.*

Armed Forces Food and Container Institute. (1958). *Study of the bases for changing food attitudes* (Official investigator, Peter H. Rossi). Chicago: Quartermaster Food and Container Institute for the Armed Forces, Research and Engineering Command, Quartermaster Corps, U.S. Army.

Campbell, B., & Mark, M. M. (2006). Toward more effective stakeholder dialogue: Applying theories of negotiation to policy and program evaluation. *Journal of Applied Social Psychology*, 2834–2863.

Campbell Collaboration. (2009). Available at *www.campbellcollaboration.org.*

Campbell D. T. (1950). The indirect assessment of social attitudes. *Psychological Bulletin*, *47*, 15–38.

Campbell, D. T. (1965). Ethnocentrism and other altruistic motives. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (pp. 283–311). Lincoln: University of Nebraska Press.

Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American Psychologist*, *22*, 817–829.

Campbell, D. T. (1969). Ethnocentrism of disciplines and the fishscale model of omniscience. In M. Sherif (Ed.), *Problems of interdisciplinary relationships in the social sciences* (pp. 328–348). Chicago: Aldine.

Campbell, D. T. (1984). Can we be scientific in applied social science? *Evaluation Studies Review Annual*, 26–48.

Campbell, D. T. (1988). Perspectives on a scholarly career. In D. T. Campbell & E. S. Overman (Eds.), *Methodology and epistemology for social science: Selected papers* (pp. 1–26). Chicago: University of Chicago Press.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

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

Cook, T. D., & Groom, C. (2004). The methodological assumptions of social psychology: The mutual dependence of substantive theory and method choice. In C. Sansone, C. C. Morf, & A. T. Panter (Eds.), *The Sage handbook of methods in social psychology* (pp. 19–44). Thousand Oaks, CA: Sage.

Corey, S. (1953). *Action research to improve school practice.* New York: Teachers College.

Coryn, C., Noakes, L. A., Westine, C. D., & Schröter, D. C. (in press). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation.*

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *New directions for evaluation: No. 80. Understanding and practicing participatory evaluation* (pp. 5–23). San Francisco: Jossey-Bass.

Darley, J. M., & Latané, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*, 377–383.

Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications.* Mahwah, NJ: Erlbaum.

Donaldson, S. I., & Berger, D. E. (2006). *The rise and promise of applied psychology in the 21st century.* In S. I. Donaldson, D. E. Berger, & K. Pezdek (Eds.), *Applied psychology: New frontiers and rewarding careers* (pp. 3–23). Mahwah, NJ: Erlbaum.

Donaldson, S. I., & Christie, C. A. (2006). *Emerging career opportunities in the transdiscipline of evaluation science* (pp. 243–259). In S. I. Donaldson, D. E. Berger, & K. Pezdek (Eds.), *Applied psychology: New frontiers and rewarding careers.* Mahwah, NJ: Erlbaum.

Donaldson, S. I., Gooler, L. E., & Scriven, M. (2002). Strategies for managing evaluation anxiety: Toward a psychology of program evaluation. *American Journal of Evaluation, 23,* 261–273.

Dovidio, J. F., Glick, P. G., & Rudman, L. (Eds.). (2005). *On the nature of prejudice: Fifty years after Allport.* Malden, MA: Blackwell.

Fetterman, D. (1998). Empowerment evaluation: Collaboration, action research, and a case example. Available at *www.aepro.org.*

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition.* New York: Random House.

Jacobs, R. C., & Campbell, D. T. (1961). The perpetuation of an arbitrary tradition through several generations of a laboratory microculture. *Journal of Abnormal and Social Psychology, 62,* 649–658.

Lewin, K. (1946). Action research and minority problems. *Journal of Social Issues, 2,* 34–46.

Lewin, K. (1947). Frontiers in group dynamics: II. Channels of group life; social planning and action research. *Human Relations, 1,* 143–153.

Lewin, K. (1948). Group decision and social change. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology* (pp. 330–341). New York: Henry Holt. [Reprinted in M. Gold (Ed.), *The complete social scientist: A Kurt Lewin reader* (pp. 265–284). Washington, DC: American Psychological Association, 1999.

Lewin, K. (1951). *Field theory in social science: Selected theoretical papers* (D. Cartwright, Ed.). New York: Harper & Row.

Lewin, K., Lippitt, R., & White, R. K. (1939). Patterns of aggressive behavior in experimentally created "social climates." *Journal of Social Psychology, 10,* 271–299.

Mark, M. M. (2008). Building a better evidence-base for evaluation theory. In P. R. Brandon & N. L. Smith (Eds.), *Fundamental issues in evaluation* (pp. 111–134). New York: Guilford Press.

Mark, M. M., Gamble, C., & Mills, J. (2005). *Social psychology and evaluation: The past, the present, and alternative possible futures.* Paper presented at the joint meeting of the American Evaluation Association and the Canadian Evaluation Society, Toronto.

Mark, M. M., & Henry, G. T. (2004). The mechanisms and outcomes of evaluation influence. *Evaluation, 10,* 35–57.

Milgram, S. (1974). *Obedience to authority: An experimental view.* New York: HarperCollins.

Miller, N., Campbell, D. T., Twedt, H., & O'Connell, E. J. (1966). Similarity, contrast and complementarity in friendship choice. *Journal of Personality and Social Psychology, 3,* 3–12.

Miller, R., King, J., Mark, M., & Stockdill, S. (2003). The oral history of evalua-

tion, Part I. Reflections on the chance to work with great people: An interview with William Shadish. *American Journal of Evaluation*, 24, 261–272.

Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Beverly Hills, CA: Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

Renger, R., & Titcomb, A. (2002). A three-step approach to teaching logic models. *American Journal of Evaluation*, 23(4), 493–503.

Rog, D. (2009, November). *Toward context-sensitive evaluation practice*. Presidential Address presented at the American Evaluation Association, Orlando, Florida.

Rogers, P., & Williams, R. (2006). Evaluation for practice improvement and organizational learning. In I. Shaw, J. Greene, & M. M. Mark (Eds.), *Handbook of evaluation* (pp. 76–97). London: Sage.

Rossi, P. H. (1989). *Down and out in America: The origins of homelessness*. Chicago: University of Chicago Press.

Rossi, P. H. (1998). *Feeding the poor: Assessing federal food aid*. Washington, DC: AEI Press.

Rossi, P. H., & Berk, R. A. (1997). *Just punishments: Federal guidelines and public views compared*. New York: De Gruyter.

Rossi, P. H., Berk, R. A., & Lenihan, K. J. (1980). *Money, work, and crime: Experimental evidence*. New York: Academic Press.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Newbury Park, PA: Sage.

Rothman, J. (1997). Action evaluation and conflict resolution: In theory and practice. *Mediation Quarterly*, *15*, 119–131.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22*, 127–160.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives on curriculum evaluation* (pp. 39–83). Chicago: Rand McNally.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Thousand Oaks, CA: Sage.

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation: Theories of practice*. Newbury Park, CA: Sage.

Stevahn, L., & King, J. (2009). *Needs assessment phase 3: Taking action for change*. Newbury Park, CA: Sage.

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

Taut, S., & Brauns, D. (2003). Resistance to evaluation: A psychological perspective. *Evaluation*, *9*, 247–264.

Wright, J. D., & Rossi, P. H. (1994). *Armed and considered dangerous: A survey of felons and their firearms*. New York: de Gruyter.

EDITORS' CONCLUDING COMMENTS TO CHAPTER 1

In the rest of this book, the editors provide comments before and after each chapter. In general, the introductory comments will highlight key points of the upcoming chapter. In some cases, the introduction will also explain new terms or concepts that are used but not explained by chapter authors. The postchapter comments, for the most part, pose questions the editors would like readers to consider, or they reflect on selected themes of the chapter, or both.