

12

THE NATURE AND PRACTICE OF PROFESSIONAL JUDGMENT IN EVALUATION

CONTENTS

Introduction	478
The Nature of the Evaluation Enterprise	478
Our Stance	479
Reconciling the Diversity in Evaluation Theory With Evaluation Practice	480
Working in the Swamp: The Real World of Evaluation Practice	481
Ethical Foundations of Evaluation Practice	482
Power Relationships and Ethical Practice	485
Ethical Guidelines for Evaluation Practice	486
Evaluation Association-Based Ethical Guidelines	486
Understanding Professional Judgment	490
What Is Good Evaluation Theory and Practice?	490
Tacit Knowledge	492
Balancing Theoretical and Practical Knowledge in Professional Practice	492
Aspects of Professional Judgment	493
The Professional Judgment Process: A Model	495
<i>The Decision Environment</i>	497
<i>Values, Beliefs, and Expectations</i>	497
Cultural Competence in Evaluation Practice	498
Improving Professional Judgment in Evaluation	499
Mindfulness and Reflective Practice	499
Professional Judgment and Evaluation Competencies	501
Education and Training-Related Activities	504
Teamwork and Improving Professional Judgment	505
The Prospects for an Evaluation Profession	506
Summary	509
Discussion Questions	510
Appendix	511
Appendix A: Fiona's Choice: An Ethical Dilemma for a Program Evaluator	511
Your Task	512
References	513

INTRODUCTION

Chapter 12 combines two central themes in this textbook: the importance of defensible methodological cores for evaluations and the importance of professional judgment in evaluation practice. We summarize our stance (we had outlined it in Chapter 3) and point out that the theoretical and methodological richness that now characterizes our field must be understood within the realities of current evaluation practice, where economic, organizational, and political pressures may constrain or misdirect the choices of design, implementation, or reporting. A theme in this textbook is that credible and defensible methodology is our foundation, but that in addition a good evaluator needs to understand the public sector environment and develop navigation tools for his or her evaluation practice.

First, we introduce several ethical lenses relevant to evaluation work and connect them to our view that evaluation practice has a moral and ethical dimension to it. We describe recent work that has been done to bring attention to the issue of ethical space for evaluation in the face of pressures to align with dominant values in public sector organizations and governments. Ethical professional practice requires evaluators to reflect on the idea of agency. The Greek concept of practical wisdom (phronesis) is explored as a concept to guide ethical practice.

We introduce ethical guidelines from several evaluation associations and describe the ethical principles that are discernable in the guidelines. We connect those principles to our discussion of ethical frameworks and to the challenges of applying ethical principles to particular situations.

We then turn to understanding professional judgment in general—how different kinds of judgments are involved in evaluation practice and how those relate to the methodological and ethical dimensions of what evaluators do. We relate professional judgment to evaluator competencies and suggest ways that evaluators can improve their professional judgment by being effective practitioners and by acquiring knowledge, skills, and experience through education, reflection, and practice. Evaluative work, from our point of view, has ethical, societal, and political implications.

The final part of our chapter is our reflections on the prospects for an evaluation profession in the foreseeable future.

THE NATURE OF THE EVALUATION ENTERPRISE

Evaluation is a structured process that creates, synthesizes and communicates information that is intended to reduce the level of uncertainty for stakeholders about the effectiveness of a given program or policy. It is intended to answer questions (see the list of evaluation questions discussed in Chapter 1) or test hypotheses, the results of which are then incorporated into the additional information bases used by those who have a stake creating, implementing, or adjusting programs or policies, ideally for the public good. Evaluative information can be used for program or organizational improvement, or for accountability and budgetary needs. It is a broad field. There can be various uses for evaluations, and thus stakeholders can mean central budget authorities, departmental decision-makers, program managers, program clients, and the public.

Our Stance

This textbook is substantially focused on evaluating the effectiveness of programs and policies. Central to evaluating effectiveness is examining causes and effects. We are not advocating that all program evaluations should be centered on experimental or quasi-experimental research designs. Instead, what we are advocating is that an evaluator needs to understand how these designs are constructed and needs to understand the *logic of causes and effects* that is at the core of experiments and quasi-experiments. In particular, it is important to identify and think through the rival hypotheses that can weaken our efforts to examine program effectiveness. In other words, we are advocating a way of thinking about evaluations that is valuable for a wide range of public sector situations where one of the key questions is whether the program was effective, or how it could become more effective. That includes asking whether the observed outcomes can be *attributed to the program*; our view is that different research designs, including qualitative approaches, can be appropriate to address questions around program effectiveness, depending on the context. In many cases, multiple lines of evidence may be necessary.

Sound methodology is necessary to evaluate the effectiveness of programs, *but it is not sufficient*. Our view is that evaluation practice also entails making judgments—judgments that range in scope and impact but are an intrinsic part of the work that we do. Fundamentally, professional judgments include both “is” and “ought” components; they are grounded in part in the tools and practices of our craft but also grounded in the ethical dimensions of each decision context. Part of what it means to be a professional is to be able to bring to bear the ethics and values that are appropriate for our day-to-day practice.

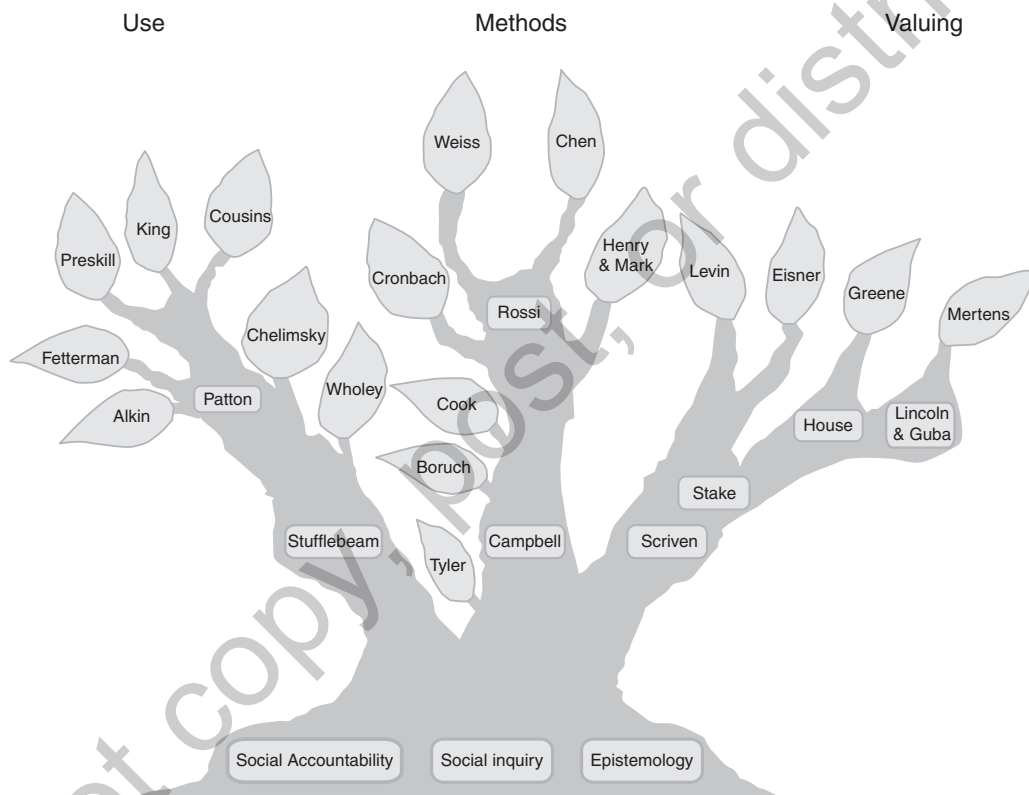
We will explore the nature and practice of making judgments in evaluations, but for now we want to be clear that because of the intrinsically political nature of evaluations, embedded as they are in value-laden environments and power relationships, it is important for evaluators who aspire to becoming professionals to recognize that the *context* for evaluations (interpersonal, organizational, governmental, economic, cultural, and societal) *all influence and are potentially influenced by* the judgments that we make as a part of the work that we do.

Later in this chapter we outline an approach to understanding and practicing professional judgment that relies in part on understanding professional practice that originated in Greek philosophy some 2500 years ago. The Aristotelian concept of *phronesis* (translated in different ways but often rendered as practical wisdom, practical reasoning, or practical ethics) is now recognized as a component of a balanced approach to professional practice—a way of recognizing and valuing the autonomy of professionals in the work they do, in contradistinction to restricting professional practice with top-down manuals, regulations and such that codify practice and are intended to make practice and its “products” uniform and predictable. This latter approach, evident in licensed human services in particular (Evans & Hardy, 2017; Kinsella & Pitman, 2012) can have the effect of reducing or hampering professional discretion/judgment in interactions with clients. Some argue that professional practice under such conditions is ethically compromised (Evans & Hardy, 2017).

Reconciling the Diversity in Evaluation Theory With Evaluation Practice

Alkin (2013) has illustrated that the field of evaluation, in its relatively short time as a discipline, has evolved into having a wide (and growing) range of theoretical approaches. The Evaluation Theory Tree, depicted in Figure 12.1, suggests the range of approaches in the field, although it is not comprehensive. For example, the Valuing part of the tree has been questioned for not separately representing social justice-related evaluation theories as a distinct (fourth) set of branches on the tree (Mertens and Wilson, 2012).

FIGURE 12.1 THE EVALUATION THEORY TREE



Source: Alkin, 2013, p. 12.

Inarguably, there is a wide range of ways that evaluators approach the field. This theoretical richness has been referenced in different chapters of this textbook. One reason why evaluators are attracted to the field is the opportunity to explore different combinations of philosophical and methodological approaches. But our field is also grounded in practice, and understanding some of the contours of actual evaluation practice is important in our pursuing the nature and practice of professional judgment in the work we do. Public sector evaluations should be designed to address public interest, but there are a number of different views on how to determine choices to be made in the realm of 'public interest'.

Working in the Swamp: The Real World of Evaluation Practice

Most evaluation practice settings continue to struggle with optimizing methodological design in the public sector milieu of “wicked problems.” Typical program evaluation methodologies rely on multiple, independent lines of evidence to bolster research designs that are case studies or implicit designs (diagrammed in Chapter 3 as XO designs, where X is the program and O is the set of observations/data on the outcomes that are expected to be affected by the program). That is, the program has been implemented at some time in the past, and now the evaluator is expected to assess program effectiveness—perhaps even summatively. There is no pre-test and no control group; there are insufficient resources to construct these comparisons, and in most situations, comparison groups are not feasible. Although multiple data sources permit triangulation of findings, that does not change the fact that the basic research design is the same; it is simply *repeated* for each data source (which is a strength since measurement errors would likely be independent) but is still subject to the prospective weaknesses of that design. In sum, typical program evaluations are conducted *after* the program is implemented, in settings where the evaluation team has to rely on evidence about the program group alone (i.e., there is no control group). In most evaluation settings, these designs rely on mixed qualitative and quantitative lines of evidence.

In such situations, some evaluators would advocate not using the evaluation results to make any causal inferences about the program. In other words, it would be argued that such evaluations ought not to be used to try to address the question: “Did the program make a difference, and if so, what difference(s) did it make?” Instead the evaluation should be limited to describing whether intended outcomes were actually achieved, regardless of whether the program itself “produced” those outcomes. That is essentially what performance measurement systems do.

But, many evaluations are commissioned with the need to know whether the *program* worked, and why. Even formative evaluations often include questions about the effectiveness of the program (Chen, 1996; Cronbach, 1980; Weiss, 1998). Answering “why” questions entails looking at *causes and effects*.

In situations where a client wants to know if and why the program was effective, and there is clearly insufficient time, money, and control to construct an evaluation design that meets criteria for answering those questions using an experimental design, evaluators have a choice. They can advise their client that wanting to know whether the program or policy worked—and why—is perhaps not feasible, or they can proceed with the understanding that their work may not be as defensible as some research textbooks (or theoretical approaches) would advocate.

Usually, some variation of the work proceeds. Although RCT comparisons between program and no-program groups are not possible, comparisons among program recipients (grouped by socio-demographic variables or perhaps by how much exposure they have had to the program), comparisons over time for program recipients who have participated in the program, and comparisons with other stakeholders or clients are all possible. We maintain that the way to answer causal questions without research designs that can rule out most rival hypotheses is to acknowledge that in addressing issues such as program effectiveness (which we take to be the central question in most evaluations and one of the distinguishing features of our field) we cannot

offer definitive findings or conclusions. Instead, our findings, conclusions, and our recommendations, supported by the evidence at hand and by our *professional judgment*, will reduce the uncertainty associated with the question.

In this textbook, our point of view is that in *all evaluations*, regardless of how sophisticated they are in terms of research designs, measures, statistical tools, or qualitative analytical methods, evaluators will use one form or another of professional judgment in the decisions that comprise the process of designing and completing an evaluation project. Moreover, rather than focusing exclusively on the judgment of merit and worth, we are saying that judgment calls are reflected in decisions that are made *throughout the process* of providing information during the performance management cycle.

Where research designs are weak in terms of potential threats to their internal validity, as evaluators we introduce to a greater extent our own experience and our own (sometimes subjective) assessments, which in turn are conditioned by ethical considerations and our values, beliefs, and expectations. These become part of the basis on which we interpret the evidence at hand and are also a part of the conclusions and the recommendations. This professional judgment component in every evaluation complements and even supplements the kinds of methodologies we deploy in our work. We believe it is essential to be aware of what professional judgments consist of and learn how to cultivate and practice sound professional judgment.

ETHICAL FOUNDATIONS OF EVALUATION PRACTICE

In this section of Chapter 12, we introduce basic descriptions of ethical frameworks that have guided contemporary public administration. We then introduce the growing body of theory and practice that advocates for including “practical wisdom” as a necessity for an ethical stance in everyday professional practice. Importantly, practical wisdom is intended to create space for professionals to exercise their judgment and to take into account the social context of the decisions they make in the work they do.

We distinguish three different approaches to ethics that are all relevant to public administration and by implication to the work that goes on in or with both public organizations and governments. An understanding of these three approaches gives evaluators a bit of a map to understand the political and organizational context surrounding the evaluation design, implementation, and reporting process.

The “duty” approach to ethics (sometimes called **deontological ethics**) was articulated by Emmanuel Kant in part as a reaction to what he saw as contemporary moral decay (he lived from 1724 to 1804) and is based on being able to identify and act on a set of unchanging ethical principles. For Kant, “situational, or relativistic ethics invited moral decay. Without immutable, eternal, never-changing standards, a person or a society was trapped on a slippery slope where anything was allowed to achieve one’s goals.” (Martinez, 2009, p. xiii). **Duty ethics** has evolved over time but is linked to contemporary administrative systems that have codified and elaborated policies and rules that determine how to respond to a wide range of decision-making situations. In effect this approach relies on nested rules and regulations to guide public officials

in their duties and responsibilities (Langford, 2004). Where the existing rules are found to be short, new rules can be elaborated to cover those (heretofore) unanticipated situations. Over time, duty ethics applications can suffer from accretion. Rules pile on rules to a point where procedures can dominate decision-making, and processing slows down administrative activities and decisions. The rules establish consistency and equality of treatment, but efficiency and effectiveness may be sacrificed because of red tape.

An important criticism of this approach by proponents of New Public Management in the 1980s and 1990s (Osborne & Gaebler, 1992) was that depending on processes to guide decision-making displaces a focus on achieving results; efficiency is reduced and effectiveness is under-valued.

A second approach that is now an important part of contemporary administrative and governmental settings, and is arguably replacing rules-based ethical regimes, is a focus on results-based values. In contemporary administrative settings, this approach has evolved into values-based ethical frameworks wherein sets of core values (desirable qualities or behaviors for individuals or groups) can be identified for public servants, and those values are promulgated as the foundation for making ethical decisions (Langford, 2004). It is a basis for NPM norms of “letting the manager manage” in an environment of performance incentives and alignment with organizational objectives.

Langford (2004), in a trenchant critique of the Canadian federal government’s values-based ethical framework, points out that statements of core values are hard to pin down and hard to translate into guidance for particular decision-making situations. His comments on “value shopping” suggest that this approach to ethics engenders organizational conflicts:

Beyond the inherent silliness of valuing anything and everything, lies the spectre of endless value conflict. For the cynical, a long list of core values affords an opportunity to “value shop.” The longer the list, the more likely it is that a federal public servant, facing a hard choice or questions from superiors about an action taken, could rationalize any position or rule interpretation by adhering to one core value rather than to another. What is an opportunity for the cynical is a nightmare for more responsible public servants. Where one sees the obligation to advance the value of service in a particular situation, another might see the value of accountability as dominant, and another might feel compelled by the demands of fairness. Value conflict is the inevitable result of large core-value sets. (p. 439)

A third approach is **consequentialism**—an approach to ethical decision making that focuses on making choices based on a weighing of the social cost and benefit consequences of a decision. Although different formulations of this approach have been articulated, they generally have in common some kind of formal or informal process wherein decision makers weigh the “benefits” and the “costs” of different courses of action and make a choice that, on balance, has the most positive (or least negative) results for society or key stakeholders. Langford, in his critique of the values-based ethics regime put into place by the federal government of Canada in the 1990s, argues that public servants are inherently more likely to be consequentialists “While undoubtedly removed from contemporary philosophical debates about consequentialism, virtually all public servants intuitively resort to the premium attached in all democratic societies to being able to defend actions or rules in terms of their impacts on all affected stakeholders in specific situations.” (Langford, 2004, p. 444).

Consequentialism, based as it is on a philosophical tradition that emphasizes weighing ethical decisions in terms of “benefits versus costs,” has commonalities with utilitarianism (Mill, Bentham, Ryan, & Bentham, 1987). However, the consequentialist approach has been criticized for being incapable of taking into account human rights (an equality- and fairness-based duty ethics perspective).

For example, in a recent evaluation of an ongoing program in New York City that focused on providing wrap-around services to those who were at risk of being homeless (Rolston, Geyer, Locke, Metraux, & Treglia, 2013), a sample of homeless or near homeless families were given a choice to participate in a random assignment process (half of those agreeing would receive the program and the other half would be denied the program for two years—the duration of the experiment). Families *not* choosing to be randomly assigned were *denied the service* for up to four months while sufficient families were recruited to run the two-year RCT.

A consequentialist or even values-based ethical perspective *could* be used to defend the experiment; the inconvenience/costs to those families who were denied the program would have to be weighed against the benefits to all those families who receive the program *into the future*, if the program showed success (consequentialist). It could be seen as an innovative, efficient way to test a program’s effectiveness (values-based). Indeed, the evaluation did go forward, and demonstrated that the program reduced homelessness, so there was a commitment to continue funding it on that basis. However, from a human rights perspective (duty ethics) the informed consent process was arguably flawed. The at-risk families who were asked to participate were vulnerable, and expecting them to provide their “free and informed consent” (Government of Canada, 2010, p. 1) in a situation where the experimenters enjoyed a clear power-over relationship appeared to be unethical.

Another approach has re-emerged as a way to guide contemporary professional practice (Evans and Hardy, 2017; Flyvbjerg, 2004; Melé, 2005). The views of Aristotle, among the ancient Greek thinkers, have provided ideas for how to situate ethics into the practical day-to-day lives of his (and our) contemporaries. For Aristotle, five different kinds of knowledge were intended to cover all human endeavors: *episteme* (context-independent/universal knowledge); *nous* (intuition or intellect); *sophia* (wisdom); *techne* (context-dependent knowledge used to produce things); and *phronesis* (**practical wisdom**, practical reasoning, or practical ethics) (Mejlgaard et al., 2018).

Phronesis has been defined as: “Deliberation about values with reference to praxis. Pragmatic, variable, context-dependent. Oriented toward action. Based on practical value-rationality.” (Flyvbjerg, 2004, p. 287). Flyvbjerg adds, “Phronesis concerns values and goes beyond analytical, scientific knowledge (episteme) and technical knowledge or know how (techne) and involves what Vickers (1995) calls “*the art of judgment*” (Flyvbjerg, 2004, p. 285, emphasis added).

Mejlgaard et al., (2018), in referring to previous work Flyvbjerg published, suggest five questions that comprise a framework for making practical ethical decisions: *Where are we going? Is this desirable? What should be done? Who gains and who loses? And by what mechanisms?* (p. 6). Schwandt (2018) uses these questions to challenge contemporary evaluation practice. He highlights the tensions that can occur between one’s beliefs about ethical conduct, one’s political stance, and one’s professional obligations.

Professional practitioners (social workers, teachers, and healthcare workers are examples) sometimes find themselves being constrained by organizational and governmental expectations to behave in ways that are consistent with organizational objectives (efficiency and cost-cutting, for example), over client-focused program objectives or overall social good (Evans & Hardy, 2017). Situating a practical wisdom perspective on ethical decision-making, Evans and Hardy (2017) suggest that this fusion of ancient and modern opens up possibilities for seeing ethical decision-making in pragmatic terms:

An alternative approach is “ethics” that sees ethical theories as resources to help us think about these fundamental issues. Concern for consequences, rights, procedural consistency, individual ethical creativity and virtue are not mutually exclusive; they do not reflect different schools but are necessary tools that can be drawn on to analyse the nature of the ethical problem and identify an ethical response. For O’Neil (1986, p. 27), ethical thinking “. . . will require us to listen to other appraisals and to reflect on and modify our own . . . Reflective judgment so understood is an indispensable preliminary or background to ethical decisions about any actual case” (p. 951).

This is a subtle point, but worth highlighting: There is not necessarily one “best” model of ethics; professional judgment entails being aware of the various types of ethical pressures that may be in play in a given context, and being able to reflectively navigate the situation.

Similarly, Melé (2005), in his discussion of ethical education in the accounting profession, highlights the importance of cultivating the (Aristotelian) virtues-grounded capacity to make moral judgments:

In contrast to modern moral philosophy, the Aristotelian view argues that moral judgment “is not merely an intellectual exercise of subsuming a particular under rules or hyper-norms. Judgment is an activity of perceiving while simultaneously perfecting the capacity to judge actions and choices and to perceive being” (Koehn, 2000, p. 17). (p. 100).

In a nutshell, as part of one’s professional judgment as an evaluator, ethical reflection is necessary because it is practically inevitable that an evaluator, at some point, will find herself or himself in a situation that requires an ethical decision and response. An evaluator’s personal “agency” can be challenged by power relationships. We explore that topic next.

Power Relationships and Ethical Practice

Flyvbjerg (2004) acknowledges that Aristotle and other proponents of this ethical approach (Gadamer, 1975) did not include power relationships in their formulations. The current interest in practical wisdom is coupled with a growing concern that professionals, working in organizations that operate under the aegis of neo-liberal principles that prioritize effective and efficient administration (Emslie & Watts, 2017; Evans & Hardy, 2017; House, 2015; Petersen & Olsson, 2015) are subject to pressures that can cause ethical tension: The “ethical ‘turn’ in the social work academy over the past few years has occurred partly in response to concerns that contemporary practice, occurring with a framework of neo-liberal managerialism, is actually unethical.” (Evans & Hardy, 2017, p. 948).

Sandra Mathison (2017), in a keynote speech to the Australasian Evaluation Society, draws a connection between the dominant sociopolitical ideologies that have paralleled the development of the evaluation field, and the normative focus of the evaluation field itself: social democracy (1960 to roughly 1980), neo-liberalism (1980 to the present day) and populism (present day into the future). Her concern is that, notwithstanding some evaluators' continued focus on the goal of improving social justice (e.g., Astbury, 2016; Donaldson & Picciotto, 2016; House, 2015; Mertens & Wilson, 2012), "by most accounts, evaluators' work isn't contributing enough to poverty-reduction, human rights, and access to food, water, education and health care." (p. 1). In summary, her view is that the field, and evaluation practice in particular, is "not contributing enough to the public good." (p. 2). Mathison (2017) argues that we are still in the neo-liberal era, notwithstanding the recent emergence of populism and the uncertainties that it brings. The dominant view of evaluation (and policy analysis) is that "evaluation has become a tool of the state . . . constantly monitoring and assessing public policies, the conduct of organizations, agencies and individuals, even serving as the final evaluator" (p. 4).

Proponents of practical wisdom as an ethical stance are asserting that valuing more robust professional autonomy for practitioners is a way to push back against the pressures to which Mathison and others point. In effect, advocates for incorporating practical wisdom into the ethical foundations for practice are saying that by acknowledging the moral dimensions of professional practice, and fostering the development of moral dispositions in those who practice, it is more likely that practitioners will be able and willing to reflect on the consequences of their decisions for their clients and for other stakeholders, and have ethical considerations impact their actual practice. This is more than consequentialism; instead, it is about taking a critical stance on the importance of improving social justice by addressing the power-related implications of professional practice.

ETHICAL GUIDELINES FOR EVALUATION PRACTICE

As the field of evaluation grows and diversifies internationally (Stockmann & Meyer, 2016), and as evaluation practitioners encounter a wider range of political, social and economic contexts, there is a growing concern that the field needs to come to grips with the implications of practicing in a wide range of political and cultural contexts, some of which challenge evaluators to take into account power imbalances and inequalities (House, 2015; Mathison, 2017; Picciotto, 2015; Schwandt, 2017). What, so far, have evaluation societies established to address norms for ethical practice?

Evaluation Association-Based Ethical Guidelines

The evaluation guidelines, standards, and principles that have been developed by various evaluation associations all address, in different ways, ethical practice. Although evaluation practice is not guided by a set of professional norms that are enforceable (Rossi, Lipsey, & Freeman, 2004), ethical guidelines are an initial normative reference point for evaluators. Increasingly, organizations that involve people (e.g., clients or employees) in research are expected to take into account

the rights of their participants across the stages of the evaluation. In universities, for example, human research ethics committees routinely scrutinize research plans to ensure that they do not violate the rights of participants. In both the United States and Canada, there are national policies or regulations that are intended to protect the rights of persons who are participants in research (Government of Canada, 2014; U.S. Department of Health and Human Services, 2009).

The past quarter century has witnessed significant developments in the domain of evaluation ethics guidelines. These include publication of the original and revised versions of the *Guiding Principles for Evaluators* (AEA, 1995, 2004, 2018), and the second and third editions of the *Program Evaluation Standards* (Sanders, 1994; Yarbrough, Shulha, Hopson, & Caruthers, 2011). The 2011 version of the *Program Evaluation Standards* has been adopted by the Canadian Evaluation Society (CES, 2012b). Two examples of books devoted to program evaluation ethics (Morris, 2008; Newman & Brown, 1996) as well as chapters on ethics in handbooks in the field (Seiber, 2009; Simons, 2006) are additional resources. More recently, Schwandt (2007, 2015, 2017) and Scriven (2016) have made contributions to discussions about both evaluation ethics and professionalization.

The AEA is active in promoting evaluation ethics with the creation of the Ethical Challenges section of the *American Journal of Evaluation* (Morris, 1998), now a rotating feature of issues of the journal. Morris (2011) has followed the development of evaluation ethics over the past quarter century and notes that there are few empirical studies that focus on evaluation ethics to date. Additionally, he argues that “most of what we know (or think we know) about evaluation ethics comes from the testimonies and reflections of evaluators”—leaving out the crucial perspectives of other stakeholders in the evaluation process (p. 145). Textbooks on the topic of evaluation range in the amount of attention that is paid to evaluation ethics; in some textbooks, it is the first topic of discussion on which the rest of the chapters rest, as in, for example, *Qualitative Researching* by Jennifer Mason (2002) and Mertens and Wilson (2012). In others, the topic arises later, or in some cases it is left out entirely.

Table 12.5 summarizes some of the ethical principles that can be discerned in the AEA’s *Guiding Principles for Evaluators* (AEA, 2018) and the Canadian Evaluation Society (CES) *Guidelines for Ethical Conduct* (CES, 2012a).

The ethical principles summarized in the right-hand column of Table 12.5 are similar to lists of principles/values that have been articulated by other professions. For example, Melé (2005) identifies these values in the *Code of the American Institute of Chartered Professional Accountants* (AICPA): service to others or public interest; competency; integrity; objectivity; independence; professionalism; and accountability to the profession (p. 101). Langford (2004), lists these core values for the Canadian federal public service: integrity; fairness; accountability; loyalty, excellence; respect; honesty and probity (p. 438).

These words or phrases identify desirable behaviors but do so in general terms. Recalling Langford’s (2004) assessment of the values-based ethical framework put into place in the Canadian federal government in the 1990s, a significant challenge is how these values would be *applied* in specific situations. Multiple values that could apply could easily put practitioners into situations where choices among conflicting values have to be made.

For example, the “keeping promises” principle in Table 12.5 suggests that contracts, once made, are to be honored by evaluators. But consider the following example: An evaluator makes an agreement with the executive director of a nonprofit agency to conduct an evaluation of a major program that is delivered by the agency. The contract specifies that the evaluator will deliver three interim progress reports to the executive director, in addition to a final report. As the evaluator begins her work, she learns from several agency managers that the executive director has been redirecting money from the project budget for office furniture, equipment, and her own travel expenses—none of these being connected with the program that is being evaluated. In her first interim report, the evaluator brings these concerns to the attention of the executive director, who denies any wrongdoings and reminds the evaluator that the interim reports are not to be shared with anyone else—in fact threatens to terminate the contract if the evaluator does not comply. The evaluator discusses this situation with her colleagues in the firm in which she is employed and decides to inform the chair of the board of directors for the agency. She has broken her contractual agreement and in doing so is calling on another ethical principle. At the same time, the outcome of this decision (a deliberative judgement decision) could have consequences for the evaluation engagement and possibly for future evaluation work for that group of professionals.

Of note, the frameworks in Table 12.5 include guidelines aimed at outlining responsibilities for the common good and equity (AEA, 2018). While the AEA’s (2004) fifth general guiding principle was “Responsibilities for general and public welfare”, the updated version of this principle is “Common good and equity”. It states: “Evaluators strive to contribute to the common good and advancement of an equitable and just society” (AEA, 2018, p. 3).

Our earlier discussion of practical wisdom as an attribute of professional practice goes beyond current ethical guidelines in that respect. It suggests that in particular situations, different mixes of ethical principles (and stakeholder viewpoints) can be in play, and evaluators who aspire to be ethical practitioners need to have *practice* making ethical decisions using exemplars, the experiences of other practitioners, observation, discussions with peers, and case studies. Learning from one’s own experiences is key. Fundamentally, cultivating practical wisdom is about being able to acquire virtues (permanent dispositions) “that favor ethical behavior” (Melé, 2005, p. 101). Virtues can be demonstrated, but learning them is a subjective process (Melé, 2005). In Appendix A, we have included a case that provides you with an opportunity to grapple with an example of the ethical choices that confront an evaluator who works in a government department. We discussed internal evaluation in Chapter 11, and this case illustrates the tensions that can occur for internal evaluators. The evaluator is in a difficult situation and has to decide what decision she should make, balancing ethical principles and her own well-being as the manager of an evaluation branch in that department. There is no right answer to this case. Instead, it gives you an opportunity to see how challenging ethical choice making can be, and it gives you an opportunity to make a choice and build a rationale for your choice.

The case is a good example of what is involved in exercising **deliberative judgment**—at least in a simulated setting. Flyvbjerg (2004) comments on the value of case-based curricula for schools of business administration, “In the field of business administration and management, some of the best schools, such as Harvard Business School, have understood the importance of cases over rules and emphasize case-based and practical teaching. Schools like this may be called Aristotelian” (p. 288).

TABLE 12.1 **ETHICAL PRINCIPLES IN THE AMERICAN EVALUATION ASSOCIATION (AEA) GUIDING PRINCIPLES AND THE CANADIAN EVALUATION SOCIETY (CES) GUIDELINES FOR ETHICAL CONDUCT**

AEA Guiding Principles	CES Guidelines for Ethical Conduct	Ethical Principles for Evaluators
Systematic inquiry		
Evaluators conduct databased inquiries that are thorough, methodical, and contextually relevant	Evaluators should apply systematic methods of inquiry appropriate to the evaluation	<ol style="list-style-type: none"> 1. Commitment to technical competence 2. Openness and transparency in communicating strengths and weaknesses of evaluation approach
Competence		
Evaluators provide skilled professional services to stakeholders	Evaluators are to be competent in their provision of service	<ol style="list-style-type: none"> 1. Commitment to the technical competence of the evaluation team 2. Commitment to the cultural competence of the team
Integrity		
Evaluators behave with honesty and transparency in order to ensure the integrity of the evaluation	Evaluators are to act with integrity in their relationships with all stakeholders	<ol style="list-style-type: none"> 1. Being honest 2. Keeping promises 3. No conflicts of interest—disclose any roles, relationships or other factors that could bias the evaluation engagement 4. Commitment to integrity
Respect for people		
Evaluators honor the dignity, well-being, and self-worth of individuals and acknowledge the influence of culture within and across groups	Evaluators should be sensitive to the cultural and social environment of all stakeholders and conduct themselves in a manner appropriate to the environment	<ol style="list-style-type: none"> 1. Free and informed consent 2. Privacy and confidentiality 3. Respect the dignity and self-worth of all stakeholders 4. When feasible, foster social equity so that those that have given to the evaluation may benefit from it 5. Understand, respect and take into account social and cultural differences among stakeholders 6. Maximize the benefits and reduce unnecessary harms

(Continued)

TABLE 12.1 ● (CONTINUED)

AEA Guiding Principles	CES Guidelines for Ethical Conduct	Ethical Principles for Evaluators
Common good and equity		
Evaluators strive to contribute to the common good and advancement of an equitable and just society	See above (under 'Respect for people'), AND: Evaluators are to be accountable for their performance and their product	<ol style="list-style-type: none"> 1. Take into account the public interest and consider the welfare of society as a whole 2. Balance the needs and interests of clients and other stakeholders 3. Communicate results in a respectful manner 4. Honor commitments made during the evaluation process 5. Commitment to full and fair communications of evaluation results

UNDERSTANDING PROFESSIONAL JUDGMENT

The competent practitioner uses his or her learned, experiential, and intuitive knowledge to assess a situation and offer a diagnosis (in the health field, for example) or a decision in other professions (Eraut, 1994; Cox & Pyakuryal, 2013). Although theoretical knowledge is a part of what competent practitioners rely on in their work, practice is seen as more than applying theoretical knowledge. It includes a substantial component that is learned through practice itself. Although some of this knowledge can be codified and shared (Schön, 1987; Tripp, 1993), part of it is **tacit**—that is, known to individual practitioners, but not shareable in the same ways that we share the knowledge in textbooks, lectures, or other publicly accessible learning and teaching modalities (Schwandt, 2008; Cox & Pyakuryal, 2013). Evaluation context is *dynamic*, and evaluators need to know how to navigate the waves of economic, organizational, political, and societal change. We explore these ideas in this section.

What Is Good Evaluation Theory and Practice?

Views of evaluation theory and practice, and in particular about what they ought to be, vary widely (Alkin, 2013). At one end of the spectrum, advocates of a highly structured (typically quantitative) approach to evaluations tend to emphasize the use of research designs that ensure sufficient internal and statistical conclusions validity that the key causal relationships between the program and outcomes can be tested. According to this view, experimental designs—typically randomized controlled trials—are the benchmark of sound evaluation designs, and departures from this ideal can be associated with problems that either require specifically designed (and usually more complex) methodologies to resolve limitations, or are simply not resolvable—at least to a point where plausible threats to internal validity are controlled. The emphasis on scientific methodology has waxed and waned in evaluation over the years.

Robert Picciotto (2015) has suggested that there have been four waves in “the big tent of evaluation” (p. 152), each reflecting the dominant political ideology of the time (Vedung, 2010). The first wave was Donald Campbell’s “experimenting society” approach to evaluation wherein programs were conceptualized as disseminable packages that would be rigorously evaluated at the pilot stage and then, depending on the success of the program, either rolled out more broadly or set aside. An important feature of Campbell’s approach was the belief that programs could be more or less effective, but conferring effectiveness did not “blame or shame” those who operated the program. Evaluations were ways of systematically learning “what worked”.

The second wave was a reaction to this positivist or post-positivist view of what was sound evaluation. This second wave was “dialogue-oriented, constructivist, participatory and pluralistic” (Picciotto, 2015, p. 152). We have outlined ontological, epistemological and methodological elements of this second wave in Chapter 5 of the textbook, where we discussed qualitative evaluation.

The third wave, which generally supplanted the second, paralleled the ideological neo-liberal, new public management shift that happened in the 1980s and beyond. That shift “swelled and engulfed the evaluation discipline: it was called upon to promote free markets; public-private partnerships and results-based incentives in the public sector” (p. 152). An important feature of this wave was a shift from governments valuing program evaluation to valuing performance measurement systems. The field of evaluation, after initially resisting performance measurement and performance management (Perrin, 1998), has generally accepted that performance measurement is “here to stay” (Feller, 2002 p. 438). An accountability and compliance-focused “what works” emphasis often dominates both program evaluation and performance measurement systems. Picciotto sees our current fourth wave as “a technocratic, positivist, utilization-focused evaluation model highly reliant on impact assessments” (p. 153). While acknowledging that “scientific concepts are precious assets for the evaluation discipline”, he argues:

We are now surfing a fourth wave. It has carried experimental evaluation to the top of the methodological pyramid. It is evidence based and it takes neo-liberalism for granted. The scientific aura of randomization steers clear of stakeholders’ values. By emphasizing a particular notion of impact evaluation that clinically verifies “what works” it has restored experimentalism as the privileged approach to the evaluation enterprise. By doing so it has implicitly helped to set aside democratic politics from the purview of evaluation—the hallmark of the prior dialogical wave. (p. 153)

An example of the enduring influence of “results-based” neo-liberalism on government policies is the recent changes made to the evaluation policy in the Canadian federal government. In 2016, the *Policy on Results* (Treasury Board, 2016a) was implemented, rescinding the earlier *Policy on Evaluation* (Treasury Board, 2009). The main thrust now is a focus on measuring and reporting performance—in particular implementing policies and programs and then measuring and reporting their outcomes. This approach is a version of “deliverology”—an approach to performance management that was adopted by the British government with the guidance of Sir Michael Barber (Barber, 2007). Program evaluation, still required for many federal departments and agencies, is not featured in the *Policy on Results*. Instead it is outlined in the *Directive on Results* that is intended to detail the implementation of the policy (Treasury Board, 2016b). It is

arguable that program evaluation has to some extent been supplanted by this focus on performance measurement (Shepherd, 2018).

Tacit Knowledge

Polanyi (1958) described **tacit knowledge** as the capacity we have as human beings to integrate “facts” (data and perceptions) into patterns. He defined tacit knowledge in terms of the process of discovering theory: “This act of integration, which we can identify both in the visual perception of objects and in the discovery of scientific theories, is the tacit power we have been looking for. I shall call it tacit knowing” (Polanyi & Grene, 1969, p. 140). Pitman (2012) defines tacit knowledge this way, “Tacit knowledge carries all of the individual characteristics of personal experience, framed within the epistemic structures of the knowledge discipline that is utilized in the professional’s practice” (p. 141).

For Polanyi, tacit knowledge cannot be communicated directly. It has to be learned through one’s own experiences—it is by definition personal knowledge. Knowing how to ride a bicycle, for example, is in part tacit. We can describe to others the physics and the mechanics of getting onto a bicycle and riding it, but the experience of getting onto the bicycle, pedaling, and getting it to stay up is quite different from being told how to do so.

Ethical decision making has been described as tacit (Mejlgaard et al., 2018; Pitman, 2012). This suggests that experience is an important factor in cultivating sound ethical decision-making (Flyvbjerg, 2004; Mejlgaard et al., 2018).

One implication of acknowledging that what we know is in part personal is that we cannot teach everything that is needed to learn a skill. The learner can be guided with textbooks, examples, and demonstrations, but that knowledge (Polanyi calls it impersonal knowledge) must be combined with the learner’s own capacity to tacitly know—to *experience* the realization (or a series of them) that he or she understands/intuits how to use the skill.

Clearly, from this point of view, *practice* is an essential part of learning. One’s own experience is essential for fully integrating impersonal knowledge into working/personal knowledge. But because the skill that has been learned is in part tacit, when the learner tries to communicate it, he or she will discover that, at some point, the best advice is to suggest that the new learner try it and “learn by doing.” This is a key part of craftsmanship.

Balancing Theoretical and Practical Knowledge in Professional Practice

The difference between the applied theory and the practical know-how views of professional knowledge has been characterized as the difference between *knowing that* (publicly accessible, propositional knowledge and skills) and *knowing how* (practical, intuitive, experientially grounded knowledge that involves wisdom, or what Aristotle called *praxis*) (Eraut, 1994; Fish & Coles, 1998; Flyvbjerg, 2004; Kemmis, 2012; Schwandt, 2008).

These two views of professional knowledge highlight different views of what professional practice is and indeed ought to be. The first view can be illustrated with an example. In the field of

medicine, the technical/rational view of professional knowledge and professional practice continues to support efforts to construct and use expert systems—software systems that can offer a diagnosis based on a logic model that links combinations of symptoms in a probabilistic tree to possible diagnoses (Fish & Coles, 1998). By inputting the symptoms that are either observed or reported by the patient, the expert system (embodying the public knowledge that is presumably available to competent practitioners) can treat the diagnosis as a problem to solve. Clinical decision making employs algorithms that produce a probabilistic assessment of the likelihood that symptoms and other technical information will support one or another alternative diagnoses. More recently, Arsene, Dumitrache and Mihiu (2015) describe an expert system for medical diagnoses that incorporates expert sub-systems for different parts (systems) of the body—circulatory system, for example, that each work with information inputs, and communicate with their counterpart sub-systems to produce an overall diagnosis. The growing importance of artificial intelligence (AI) systems suggests that there will be more applications of this approach in medicine in the future.

Alternatively, the view of professional knowledge as practical know-how embraces the perspective of professional practice as craftsmanship or even artistry. Although it highlights the importance of experience in becoming a competent practitioner, it also complicates our efforts to understand the nature of professional evaluation practice. If practitioners know things that they cannot share and their knowledge is an essential part of sound practice, how do professions find ways of ensuring that their members are competent?

Schwandt (2008) recognizes the importance of *balancing* applied theory and practical knowledge in evaluation. His concern is with the tendency, particularly in performance management systems where practice is circumscribed by a focus on outputs and outcomes, to force “good practice” to conform to some set of performance measures and performance results:

The fundamental distinction between instrumental reason as the hallmark of technical knowledge and judgment as the defining characteristic of practical knowledge is instinctively recognizable to many practitioners (Dunne & Pendlebury, 2003). Yet the idea that “good” practice depends in a significant way on the experiential, existential knowledge we speak of as perceptivity, insightfulness, and deliberative judgment is always in danger of being overrun by (or at least regarded as inferior to) an ideal of “good” practice grounded in notions of objectivity, control, predictability, generalizability beyond specific circumstances, and unambiguous criteria for establishing accountability and success. This danger seems to be particularly acute of late, as notions of auditable performance, output measurement, and quality assurance have come to dominate the ways in which human services are defined and evaluated. (p. 37)

The idea of balance is further explored in the section below, where we discuss various aspects of professional judgment.

Aspects of Professional Judgment

What are the different kinds of professional judgment? How does professional judgment impact the range of decisions that evaluators make? Can we construct a model of how professional judgment relates to evaluation-related decisions?

Fish and Coles (1998) have constructed a typology of four kinds of professional judgment in the health care field. We believe that these are useful for understanding professional judgment in evaluation. Each builds on the previous one; the kinds of judgment differ across the four kinds. At one end of the continuum, practitioners apply **technical judgments** that are about specific issues involving routine tasks. Typical questions would include the following: What do I do now? How do I apply my existing knowledge and skills to do this routine task? In an evaluation, an example of this kind of judgment would be how to select a random sample from a population of case files in a social service agency.

The next level is **procedural judgment**, which focuses on procedural questions and involves the practitioner comparing the skills/tools that he or she has available to accomplish a task. Practitioners ask questions such as “What are my choices to do this task?” “From among the tools/knowledge/skills available to me, which combination works best for this task?” An example from an evaluation would be deciding how to include clients in an evaluation of a social service agency program—whether to use a survey (and if so, internet, mailing, telephone, interview format, or some combination) or use focus groups (and if so, how many, where, how many participants in each, how to gather them).

The third level of professional judgment is **reflective**. It again assumes that the task or the problem is a given, but now the practitioner is asking the following questions: How do I tackle this problem? Given what I know, what are the ways that I could proceed? Are the tools that are easily within reach adequate, or instead, should I be trying some new combination or perhaps developing some new ways of dealing with this problem? A defining characteristic of this third level of professional judgment is that the practitioner is reflecting on his or her practice/experience and is seeking ways to enhance his or her practical knowledge and skills and perhaps innovate to address a given situation.

The fourth level of professional judgment is **deliberative**. The example earlier in this chapter described an evaluation of a homelessness prevention program in New York City (Rolston et al., 2013) wherein families were selected to participate through a process where at least some arguably did not have the capacity to offer them free and informed consent. Members of the evaluation team decided to implement a research design (an RCT) that was intended to maximize internal validity and privilege that over the personal circumstances and the needs of the families facing homelessness. What contextual and ethical factors should the evaluators have considered in that situation? No longer are the ends or the tasks fixed, but instead the professional is taking a broader view that includes the possibility that the task or problem may or may not be an appropriate one to pursue. Professionals at this level are asking questions about the nature of their practice and connecting what they do as professionals with ethical and moral considerations. The case study in Appendix A of this chapter is an example of a situation that involves deliberative judgment.

It is important to keep in mind that evaluation practice typically involves some compromises. We are often “fitting round pegs into square holes.” In some settings, even routine technical decisions (e.g., should we use significance tests where the response rate to our survey was 15 percent?) can have a significant “what should I do?” question attached to them. As we move from routine to more complex decisions, “what should I do?” becomes more important. Addressing

this question involves calling on one's experience and it is important to keep in mind that our experiences are a reflection of our values, beliefs, and expectations, and our ethical stance. Ethics are an important part of what comprises our judgments as professionals. What was being said earlier in this chapter is that professional practice is intrinsically tied to ethics; developing professional judgment involves developing practical wisdom.

The Professional Judgment Process: A Model

Since professional judgment spans the evaluation process, it will influence a wide range of decisions that evaluators make in their practice. The four types of professional judgment that Fish and Coles (1998) describe suggest decisions of increasing complexity from discrete technical decisions to deliberative decisions. Figure 12.2 displays a more detailed model of the way that professional judgment is involved in evaluator decision making. The model focuses on single decisions—a typical evaluation would involve many such decisions of varying complexity. In the model, evaluator ethics, values, beliefs, and expectations, together with both shareable and practical (tacit) knowledge combine to create a fund of experience that is the foundation for professional judgments. In turn, professional judgments influence the decision at hand.

There is a feedback loop that connects the decision environment to the evaluator via her/his shareable knowledge. There are also feedback loops that connect decision consequences with shareable knowledge and ethics, as well as practical know-how (tacit knowledge) and the evaluator's values, beliefs and expectations.

This model is dynamic: the factors in the model interact over time in such ways that changes can occur in professional judgment antecedents, summed up in evaluator experience. Later in this chapter we will discuss reflective practice.

The model can be unpacked by discussing the constructs in it. Some constructs have been elaborated in this chapter already (ethics, shareable knowledge, practical know-how, and professional judgment), but it is worthwhile to define each one explicitly in one table. Table 12.2 summarizes the constructs in Figure 12.2 and offers a short definition of each. Several of the constructs will then be discussed further to help us understand what roles they play in the process of forming and applying professional judgment.

FIGURE 12.2 THE PROFESSIONAL JUDGMENT PROCESS

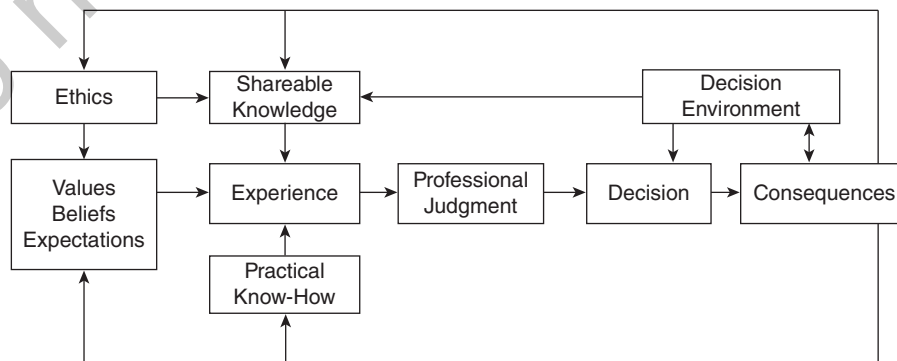


TABLE 12.2 DEFINITIONS OF CONSTRUCTS IN THE MODEL OF THE PROFESSIONAL JUDGMENT PROCESS

Constructs in the Model	Definitions
Ethics	Moral principles that are intended to guide a person's decisions about "right" and "wrong," and typically distinguish between acceptable and unacceptable behaviors. For evaluators, professional guidelines, standards or ethical frameworks are part of the ethical influences on decisions, either directly or indirectly through professional associations (for example). However, there is more to one's ethical decision-making than what is found in the guidelines.
Values	Values are statements about what is desirable, what ought to be, in a given situation. Values can be personal or more general. Values can be a part of ethical frameworks. They can be about choices, but not necessarily about right and wrong.
Beliefs	Beliefs are about what we take to be true, for example, our assumptions about how we know what we know (our epistemologies are examples of our beliefs).
Expectations	Expectations are assumptions that are typically based on what we have learned and what we have come to accept as normal. Expectations can limit what we are able to "see" in particular situations.
Shareable knowledge	Knowledge that is typically found in textbooks or other such media; knowledge that can be communicated and typically forms the core of the formal training and education of professionals in a field.
Practical know-how	Practical know-how is the knowledge that is gained through practice. It complements shareable knowledge and is <i>tacit</i> —that is, acquired from one's professional practice and is not directly shareable.
Experience	Experience is the subjective amalgam of our knowledge, ethics, values, beliefs, expectations, and practical know-how at a given point in time. For a given decision, we have a "fund" of experience that we can draw from. We can augment or change that fund with learning from the consequences of the decisions we make as professionals and from the (changing) environments in which our practice decisions occur.
Professional judgment	Professional judgment is a subjective process that relies on our experience and ranges from technical judgments to deliberative judgments.
Decision	In a typical evaluation, evaluators make hundreds of decisions that collectively define the entire evaluation process. Decisions are choices—a choice made by an evaluator about everything from discrete methodological issues to global values-based decisions that affect the whole evaluation (and perhaps future evaluations) or even the evaluator's career.
Consequences	Each decision has consequences—for the evaluator and for the evaluation process. Consequences can range from discrete to global, commensurate with the scope and implications of the decision. Consequences both influence and are influenced by the decision environment.
Decision environment	The decision environment is the set of contextual factors that influences the decision-making process, and the stock of knowledge that is available to the evaluator. Among the factors that could impact an evaluator decision are client expectations, future funding opportunities, resources (including time and data), power relationships, and constraints (legal, institutional, and regulatory requirements that specify the ways that evaluator decisions are to fit a decision environment). Evaluator decisions can also influence the decision environment—the basic idea of "speaking truth to power" is that evaluator decisions will be conveyed to organizational/political decision-makers. Mathison (2017) suggests that evaluators should "speak truth to the powerless" (p. 7) as a way of improving social justice, as an evaluation goal.

The Decision Environment

The particular situation or problem at hand, *and its context*, influence how a program evaluator's professional judgment will be exercised. Each opportunity for professional judgment will have unique characteristics that will demand that it be approached in particular ways. For example, a methodological issue will typically require a different kind of judgment from one that centers on an ethical issue. Even two cases involving a similar question of methodological choice will have facts about each of them that will influence the professional judgment process. We would agree with evaluators who argue that methodologies need to be situationally appropriate, avoiding a one-size-fits-all approach (Patton, 2008). The extent to which the relevant information about a particular situation is known or understood by the evaluator will affect the professional judgment process—professional judgments are typically made under conditions of uncertainty.

The decision environment includes constraints and incentives both real and perceived that affect professional judgment. Some examples include the expectations of the client, the professional's lines of accountability, tight deadlines, complex and conflicting objectives, organizational environment, political context, cultural considerations, and financial constraints. For people working within an organization—for example, internal evaluators—the organization also presents a significant set of decision-related factors, in that its particular culture, goals, and objectives will have an impact on the way the professional judgment process unfolds.

Values, Beliefs, and Expectations

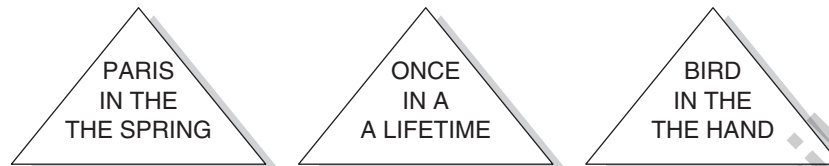
Professional judgment is influenced by personal characteristics of the person exercising it. It must always be kept in mind that “judgment is a human process, with logical, psychological, social, legal, and even political overtones” (Gibbins & Mason, 1988, p. 18). Each of us has a unique combination of values, beliefs, and expectations that make us who we are, and each of us has internalized a set of professional norms that make us the kind of practitioner that we are (at a given point in time). These personal factors can lead two professionals to make quite different professional judgments about the same situation (Tripp, 1993).

Among the personal characteristics that can influence one's professional judgment, expectations are among the most important. Expectations have been linked to paradigms; perceptual and theoretical structures that function as frameworks for organizing one's perspectives, even one's beliefs about what is real and what is taken to be factual. Kuhn (1962) has suggested that paradigms are formed through our education and training. Eraut (1994) has suggested that the process of learning to become a professional is akin to absorbing an ideology.

Our past experiences (including the consequences of previous decisions we have made in our practice) predispose us to understand or even expect some things and not others, to interpret situations, and consequently to behave in certain ways rather than in others. As Abercrombie (1960) argues, “We never come to an act of perception with an entirely blank mind but are always in a state of preparedness or expectancy, because of our past experiences” (p. 53). Thus, when we are confronted with a new situation, we perceive and interpret it in whatever way makes it most consistent with our existing understanding of the world, with our existing paradigms. For the most part, we perform this act unconsciously. We are often not even aware of how our particular worldview influences how we interpret and judge the information we receive on a daily basis in the course of our work, or how it affects our subsequent behavior.

How does this relate to our professional judgment? Our expectations can lead us to see things we are expecting to see, even if they are not actually there, and to not see things we are not expecting, even if they are there. Abercrombie (1960) calls our worldview our “schemata” and illustrates its power over our judgment process with the following figure (Figure 12.3).

FIGURE 12.3 THE THREE TRIANGLES



Source: Abercrombie, 1960.

In most cases, when we first read the phrases contained in the triangles, we do not see the extra words. As Abercrombie (1960) points out, “it’s as though the phrase ‘Paris in the Spring,’ if seen often enough, leaves a kind of imprint on the mind’s eye, into which the phrase in the triangle must be made to fit” (p. 35). She argues that “if [one’s] schemata are not sufficiently ‘living and flexible,’ they hinder instead of help [one] to see” (p. 29). Our tendency is to ignore or reject what does not fit our expectations. Thus, similar to the way we assume the phrases in the triangles make sense and therefore unconsciously ignore the extra words, our professional judgments are based in part on our preconceptions and thus may not be appropriate for the situation. Later in this chapter we will discuss reflective practice.

Cultural Competence in Evaluation Practice

The globalization of evaluation (Stockmann & Meyer, 2016) and the growth of national evaluation associations point have evidenced that evaluation practice has components which reflect the culture(s) in which it is embedded. Schwandt (2007), speaking of the AEA case, notes that “the Guiding Principles (as well as most of the ethical guidelines of academic and professional associations in North America) have been developed largely against the foreground of a Western framework of moral understandings” (p. 400) and are often framed in terms of individual behaviors, largely ignoring the normative influences of social practices and institutions.

The American Evaluation Association (AEA, 2011) produced a cultural competence statement that is not intended to be generalized beyond the United States and describes cultural competence this way:

Cultural competence is not a state at which one arrives; rather, it is a process of learning, unlearning and relearning. It is a sensibility cultivated throughout a lifetime. Cultural competence requires awareness of self, reflection on one’s own cultural position, awareness of others’ positions, and the ability to interact genuinely and respectfully with others (AEA, 2011, p. 3).

The same document defines culture: “Culture can be defined as the shared experiences of people, including their languages, values, customs, beliefs, and mores. It also includes worldviews, ways of knowing and ways of communicating.” (p. 2). Although work is being done to update the evaluator competencies (King and Stevahn, 2015), the cultural competencies document (AEA, 2011) continues to stand apart from the competency framework.

One issue that stands out in reflecting on cultural competencies is power relationships (Chouinard & Cousins, 2007; Lowell, Kildea, Liddle, Cox & Paterson, 2015). Chouinard and Cousins, in their synthesis of Indigenous evaluation-related publications, connect the creation of knowledge in cross-cultural evaluations with a post-modern view of the relationship between knowledge and power, “To move cultural competence in evaluation beyond the more legitimate and accepted vocabulary, beyond mere words, we must appreciate that there is no resonant universal social science methodologies and no neutral knowledge generation. Knowledge, as Foucault (1980) suggests, is not infused with power, it is an effect of power” (p. 46). This view accords with the perspective taken by those who are critical of professional practice for having been perhaps “captured” by neo-liberal values (see Donaldson & Picciotto, 2016; Evans & Hardy, 2017; House, 2015; Picciotto, 2015; Schwandt, 2017). An essential part of incorporating practical wisdom as a way to approach practice is to acknowledge the moral nature of professional practice and the importance of keeping in view the power relationships in which practitioners are always embedded (Mejlgaard et al., 2018). Schwandt (2018) in a discussion of what it means for us to be evaluation practitioners suggests:

Because boundaries are not given, we have to “do” something about boundaries when we make judgments of how to act in the world. Thus, ‘what should we do?’ is a practical, situated, time- and place-bound question. Developing good answers to that question is what practical reasoning in evaluation is all about—a commitment to examining assumptions, values, and facts entailed in the questions: ‘What do we want to achieve/Where are we going?’ ‘Who gains and who loses by our actions, and by which mechanisms of power?’ ‘Is this development desirable?’ ‘What, if anything, should we do about it?’ (p. 134)

With this in mind, we move on to examine how to go about improving one’s professional judgment.

IMPROVING PROFESSIONAL JUDGMENT IN EVALUATION

Having reviewed the ways that professional judgment is woven through the fabric of evaluation practice and having shown how professional judgment plays a part in our decisions as evaluation practitioners, we can turn to discussing ways of self-consciously improving our professional judgment. Key to this process is becoming aware of one’s own decision-making processes. Mowen (1993) notes that our experience, if used reflectively and analytically to inform our decisions, can be a positive factor contributing to good professional judgment. Indeed, he goes so far as to argue that “one cannot become a peerless decision maker without that well-worn coat of experience . . . the bumps and bruises received from making decisions and seeing their outcomes, both good or bad, are the hallmark of peerless decision makers” (p. 243).

Mindfulness and Reflective Practice

Self-consciously challenging the routines of our practice is an effective way to begin to develop a more mindful stance. In our professional practice, each of us will have developed routines for addressing situations that occur frequently. As Tripp (1993) points out, although routines

... may originally have been consciously planned and practiced, they will have become habitual, and so unconscious, as expertise is gained over time. Indeed, our routines often become such well-established habits that we often cannot say why we did one thing rather than another, but tend to put it down to some kind of mystery such as "professional intuition." (p. 17)

Mindfulness as an approach to improving professional practice is becoming more appreciated and understood (Dobkin & Hutchinson, 2013; Epstein, 2017; Riskin, 2011). Dobkin and Hutchinson (2013) report that 14 medical schools in Canada and the United States teach mindfulness to their medical and dental students and residents (p.768). More generally, it is now seen as a way to prevent "compassion fatigue and burnout" in health practitioners (Dobkin & Hutchinson, 2013, p. 768).

Mindfulness is aimed at improving our capacity to become more aware of our values and morals, expectations, beliefs, assumptions, and even what is tacit in our practice.

Epstein (2003) characterizes a mindful practitioner as one who has cultivated the art of self-observation (cultivating the compassionate observer). The objective of mindfulness is to see what is rather than what one wants to see or even expects to see. Mindful self-monitoring involves several things: "access to internal and external data; lowered reactivity [less self-judging] to inner experiences such as thoughts and emotions; active and attentive observation of sensations, images, feelings, and thoughts; curiosity; adopting a nonjudgmental stance; presence, [that is] acting with awareness . . . ; openness to possibility; adopting more than one perspective; [and] ability to describe one's inner experience" (Epstein, Siegel, & Silberman, 2008, p. 10).

Epstein (1999) suggests that there are at least three ways of nurturing mindfulness: (1) mentorships with practitioners who are themselves well regarded in the profession; (2) reviewing one's own work, taking a nonjudgmental stance; and (3) meditation to cultivate a capacity to observe one's self. He goes further (Epstein, 2017) to suggest that cultivating mindfulness is not just for individual practitioners but is also for work teams and organizations.

Professionals should consistently reflect on what they have done in the course of their work and then investigate the issues that arise from this review. Reflection should involve articulating and defining the underlying principles and rationale behind our professional actions and should focus on discovering the "intuitive knowing implicit in the action" (Schön, 1988, p. 69).

Tripp (1993) suggests that this process of reflection can be accomplished by selecting and then analyzing **critical incidents** that have occurred during our professional practice in the past (critical incident analysis). This approach is used to assess and improve the quality of human services (Arora, Johnson, Lovinger, Humphrey, & Meltzer, 2005; Davies & Kinloch, 2000). A critical incident can be any incident that occurred in the course of our practice that sticks in our mind and hence, provides an opportunity to learn. What makes it critical is the reflection and analysis that we bring to it. Through the process of critical incident analysis, we can gain an increasingly better understanding of the factors that have influenced our professional judgments. For it is only in retrospect, in analyzing our past decisions, that we can see the complexities underlying what at the time may have appeared to be a straightforward, intuitive professional judgment. "By uncovering our judgments . . . and reflecting upon them," Fish and Coles (1998) maintain, "we believe that it is possible to develop our judgments because we understand more about them and about how we as individuals come to them" (p. 285).

Another key way to critically reflect on our professional practice and understand what factors influence the formation of our professional judgments is to discuss our practice with our colleagues (Epstein, 2017). Colleagues, especially those who are removed from the situation at hand or under discussion, can act as “critical friends” and can help in the work of analyzing and critiquing our professional judgments with an eye to improving them. With different education, training, and experience, our professional peers often have different perspectives from us. Consequently, involving colleagues in the process of analyzing and critiquing our professional practice allows us to compare with other professionals our ways of interpreting situations and choosing alternatives for action. Moreover, the simple act of describing and summarizing an issue so that our colleagues can understand it can reveal and provide much insight into the professional judgments we have incorporated.

Professional Judgment and Evaluation Competencies

There is continuing interest in the evaluation field in specifying the competencies that define sound evaluation practice (King & Stevahn, 2015). Building on previous work (Ghere, King, Stevahn, & Minnema, 2006; King, Stevahn, Ghere, & Minnema, 2001; Stevahn, King, Ghere, & Minnema, 2005a; Stevahn, King, Ghere, & Minnema, 2005b; Wilcox & King, 2013), King and Stevahn (2015) say:

The time has come at last for the field of program evaluation in the United States to address head-on an issue that scholars and leaders of professional evaluation associations have discussed periodically over 30 years: What is the set of competencies that an individual must have to conduct high-quality program evaluations? (p. 21)

This push is part of a broader international effort to develop evaluation competencies and link those to professionalization of the evaluation discipline (King & Stevahn, 2015; Stockmann & Meyer, 2016; Wilcox & King, 2013).

In an earlier study that included the views of 31 evaluation professionals in the United States, they were asked to rate the importance of 49 evaluator competencies and then try to come to a consensus about the ratings, given feedback on how their peers had rated each item (King et al., 2001). The 49 items were grouped into four broad clusters of competencies: (1) *systematic inquiry* (most items were about methodological knowledge and skills), (2) *competent evaluation practice* (most items focused on organizational and project management skills), (3) *general skills for evaluation practice* (most items were on communication, teamwork, and negotiation skills), and (4) *evaluation professionalism* (most items focused on self-development and training, ethics and standards, and involvement in the evaluation profession).

Among the 49 competencies, one was “making judgments” and referred to making an overall evaluative *judgment*, as opposed to a number of recommendations, at the end of an evaluation (King et al., 2001, p. 233). Interestingly, it was rated the *second lowest* on average among all the competencies. This finding suggests that judgment, comparatively, is not rated to be that important (although the item average was still 74.68 out of 100 possible points). King et al. (2001) suggested that “some evaluators agreed with Michael Scriven that to evaluate is to judge; others did not” (p. 245). The “reflects on practice” item, however, was given an average rating of 93.23—a *ranking of 17* among the 49 items. For both of these items, there was substantial

variation among the practitioners about their ratings, with individual ratings ranging from 100 (highest possible score) to 20. The discrepancy between the low overall score for “making judgments” and the higher score for “reflects on practice” may be related to the difference between *making* a judgment, as an action, and *reflecting* on practice, as a personal quality.

If we look at linkages between types of professional judgment and the range of activities that comprise evaluation practice, we can see that some kinds of professional judgment are more important for some clusters of activities than others. But for many evaluation activities, several different kinds of professional judgment can be relevant. Table 12.3 summarizes the steps we introduced in Chapter 1 to design and implement a program evaluation. For each step, we have offered a (subjective) assessment of what kinds of professional judgment are involved. You can see that for all the steps, there are multiple kinds of professional judgments involved and many of the steps involve deliberative judgments—these are the ones that are most directly related to developing a morally-grounded evaluation practice.

Table 12.4 displays the steps involved in designing and implementing a performance *measurement* system (taken from Chapter 9). What you can see is that for all the steps there are multiple kinds of professional judgment involved and for nearly all of them, deliberative judgment-related decisions. This reflects that fact that designing and implementing a performance measurement system is both a technical process and organizational change process, involving a wide range of organizational/political culture-related decisions. We have not displayed the list of steps involved in re-balancing a performance measurement system (included in Chapter 10) but the range and kinds of judgments involved would be similar to corresponding steps in Table 12.4.

TABLE 12.3 TYPES OF PROFESSIONAL JUDGMENT THAT ARE RELEVANT TO THE PROGRAM EVALUATION FRAMEWORK IN THIS TEXTBOOK

	Types of Professional Judgment			
	Technical	Procedural	Reflective	Deliberative
Steps in designing and implementing a program evaluation				
1. Who are the clients for the evaluation, and the stakeholders?			Yes	Yes
2. What are the questions and issues driving the evaluation?	Yes	Yes	Yes	Yes
3. What resources are available to do the evaluation?	Yes	Yes	Yes	Yes
4. Given the evaluation questions, what do we already know?	Yes	Yes	Yes	Yes
5. What is the logic and structure of the program?	Yes	Yes	Yes	
6. Which research design alternatives are desirable and feasible?	Yes	Yes	Yes	Yes
7. What kind of environment does the program operate in and how does that affect the comparisons available to an evaluator?			Yes	Yes

	Types of Professional Judgment			
	Technical	Procedural	Reflective	Deliberative
Steps in designing and implementing a program evaluation				
8. What data sources are available and appropriate, given the evaluation issues, the program structure, and the environment in which the program operates?	Yes	Yes	Yes	Yes
9. Given all the issues raised in Points 1 to 8, which evaluation strategy is most feasible, and defensible?			Yes	Yes
10. Should the evaluation be undertaken?			Yes	Yes
Steps in conducting and reporting an evaluation				
1. Develop the data collection instruments and pre-test them.	Yes	Yes	Yes	
2. Collect data/lines of evidence that are appropriate for answering the evaluation questions.	Yes	Yes	Yes	
3. Analyze the data, focusing on answering the evaluation questions.	Yes	Yes	Yes	
4. Write, review, and finalize the report.	Yes	Yes	Yes	Yes
5. Disseminate the report.	Yes	Yes	Yes	Yes

TABLE 12.4 ■ TYPES OF PROFESSIONAL JUDGMENT THAT ARE RELEVANT TO THE PERFORMANCE MEASUREMENT FRAMEWORK IN THIS TEXTBOOK

	Types of Professional Judgment			
	Technical	Procedural	Reflective	Deliberative
Steps in designing and implementing a performance measurement system				
1. Leadership: Identify the organizational champions of this change.			Yes	Yes
2. Understand what a performance measurement system can and cannot do and why it is needed.			Yes	Yes
3. Communication: Establish multichannel ways of communicating that facilitate top-down, bottom-up, and horizontal sharing of information, problem identification, and problem solving.	Yes	Yes	Yes	Yes
4. Clarify the expectations for the uses of the performance information that will be created.			Yes	Yes

(Continued)

TABLE 12.4 (CONTINUED)

	Types of Professional Judgment			
	Technical	Procedural	Reflective	Deliberative
Steps in designing and implementing a performance measurement system				
5. Identify the resources and plan for the design, implementation and maintenance of the performance measurement system.	Yes	Yes	Yes	Yes
6. Take the time to understand the organizational history around similar initiatives.			Yes	Yes
7. Develop logic models for the programs or lines of business for which performance measures are being developed.	Yes	Yes	Yes	
8. Identify constructs that are intended to represent performance for aggregations of programs or the whole organization.	Yes	Yes	Yes	Yes
9. Involve prospective users in reviewing the logic models and constructs in the proposed performance measurement system.	Yes	Yes	Yes	Yes
10. Translate the constructs into observable measures.	Yes	Yes	Yes	
11. Highlight the comparisons that can be part of the performance measurement system.	Yes	Yes	Yes	Yes
12. Reporting results and then regularly review feedback from users and, if needed, make changes to the performance measurement system.	Yes	Yes	Yes	Yes

Education and Training-Related Activities

Developing sound professional judgment depends substantially on being able to develop and *practice* the craft of evaluation. Schön (1987) and Tripp (1993), among others (e.g., Greeff & Rennie, 2016; Mejlgaard et al., 2018; Melé, 2005), have emphasized the importance of experience as a way of cultivating sound professional judgment. Although textbook knowledge is also an essential part of every evaluator's toolkit, a key part of evaluation curricula are opportunities to acquire experience and by implication, tacit knowledge.

There are at least six complementary ways that evaluation education and training can be focused to provide opportunities for students and new practitioners to develop their judgment skills. Some activities are more discrete—that is, are relevant for developing skills that are specific—these are more focused on technical and procedural judgment-related skills. These are generally limited to a single course or even a part of a course. Others are more generic, offering opportunities to acquire experience that spans entire evaluation processes. These are typically activities that integrate coursework into work experiences. Table 12.5 summarizes ways that academic programs can inculcate professional judgment capacities in their students.

The types of learning activities in Table 12.5 are typical of many programs that train evaluators, but what is important is realizing that each of these kinds of activities contributes directly to developing a set of skills that all practitioners need and will use in all their professional work. In an important way, identifying these learning activities amounts to making explicit what has largely been tacit in our profession.

TABLE 12.5 ■ LEARNING ACTIVITIES TO INCREASE PROFESSIONAL JUDGMENT CAPACITY IN NOVICE PRACTITIONERS

Learning Activities	Types of Professional Judgment Involved	Examples
Course-based activities		
Problem/puzzle solving	Technical and procedural judgment	Develop a coding frame and test the coding categories for intercoder reliability for a sample of open-ended responses to an actual client survey that the instructor has provided
Case studies	Technical, procedural, reflective, and deliberative judgment	Make a decision for an evaluator who finds himself or herself caught between the demands of his or her superior (who wants evaluation interpretations changed) and the project team who see no reason to make any changes
Simulations	Technical, procedural, reflective, and deliberative judgment	Using a scenario and role playing, negotiate the terms of reference for an evaluation
Course projects	Technical, procedural, reflective, and deliberative judgment	Students are expected to design a practical, implementable evaluation for an actual client organization
Program-based activities		
Apprenticeships/internships/work terms	Technical, procedural, reflective, and deliberative judgment	Students work as apprentice evaluators in organizations that design and conduct evaluations, for extended periods of time (at least 4 months)
Conduct an actual program evaluation	Technical, procedural, reflective, and deliberative judgment	Working with a client organization, develop the terms of reference for a program evaluation, conduct the evaluation, including preparation of the evaluation report, deliver the report to the client, and follow up with appropriate dissemination activities

Teamwork and Improving Professional Judgment

Evaluators and managers often work in organizational settings where teamwork is expected. Successful teamwork requires establishing norms and expectations that encourage good communication, sharing of information, and a joint commitment to the task at hand. In effect a well-functioning team is able to develop a learning culture for the task at hand. Being able to select team members and foster a work environment wherein people are willing to trust each other, and be open and honest about their own views on issues, is conducive to generating information that reflects a diversity of perspectives. Even though there will still be individual biases,

the views expressed are more likely to be valid than the perceptions of a dominant individual or coalition in the group. Parenthetically, an organizational culture that emulates features of learning organizations (Garvin, 1993; Mayne, 2008) will tend to produce information that is more valid as input for making decisions and evaluating policies and programs.

Managers and evaluators who have the skills and experience to network with others and, in doing so, be reasonably confident that honest views about an issue are being offered, have a powerful tool to complement their own knowledge and experience and their own systematic inquiries.

THE PROSPECTS FOR AN EVALUATION PROFESSION

What does it mean to be a professional? What distinguishes a profession from other occupations? Eraut (1994) suggests that professions are characterized by the following: a core body of knowledge that is shared through the training and education of those in the profession; some kind of government-sanctioned license to practice; a code of ethics and standards of practice; and self-regulation (and sanctions for wrongdoings) through some kind of professional association to which members of the practice community must belong.

The idea that evaluation is a profession, or aspires to be a profession, is an important part of discussions of the scope and direction of the enterprise (Altschuld, 1999; Altschuld & Engle, 2015; Stockmann & Meyer, 2016). Modarresi, Newman, and Abolafia (2001) quote Leonard Bickman (1997), who was president of the American Evaluation Association (AEA) in 1997, in asserting that “we need to move ahead with professionalizing evaluation or else we will just drift into oblivion” (p. 1). Bickman and others in the evaluation field were aware that other related professions continue to carve out territory, sometimes at the expense of evaluators. Picciotto (2011) points out, however, that “heated doctrinal disputes within the membership of the AEA have blocked progress [toward professionalization] in the USA” (p. 165). More recently, Picciotto (2015) suggests that professionalizing evaluation is now a global issue wherein a significant challenge is working in contexts that do not support the democratic evaluation model that has underpinned the development of the field. He suggests, “The time has come to experiment with a more activist and independent evaluation model grounded in professional autonomy reliant on independent funding sources and tailor made to diverse governance environments.” (p. 164).

Professionalizing evaluation now appears to be a global movement, judging by the growing number of Voluntary Organizations of Professional Evaluation (VOPEs), their memberships, and the parallel efforts by some national evaluation organizations to implement first steps in making it possible for evaluation practitioners to distinguish themselves, professionally (Donaldson & Donaldson, 2015). They summarize the global lay of that land this way:

During the 2015 International Year of Evaluation we learned about the profound growth and expansion of VOPEs. While there were relatively few VOPEs prior to 1990, we have witnessed exponential growth over the past 25 years (Donaldson, Christie, & Mark, 2015;

Segone & Rugh, 2013). Rugh (personal communication, 2015, October) reported that there are now approximately 227 VOPEs (170 verified) representing 141 countries (111 verified) consisting of a total of approximately 52000 members. At the same time, there has been a rapid expansion of University courses, certificates and degree programs in evaluation and major growth in the number of VOPEs and other training organizations providing evaluation workshops, online training, and other professional development experiences in evaluation (LaVelle & Donaldson, 2015) (Donaldson & Donaldson, 2015, p. 2).

The growth in evaluation-related voluntary organizations is occurring against a background of the diversity in the field. Donaldson and Donaldson (2015) point out that the core of the evaluation field is its theories, and a contemporary reading of the field suggests that theoretical perspectives continue to emerge and differentiate themselves (Alkin, 2013; Mertens & Wilson, 2012; Stockmann & Meyer, 2016). On the one hand, this richness suggests a dynamic field that is continually enriched by the (now) global contributions of scholars and practitioners.

But if we look at evaluation as a prospective profession, this diversity presents a challenge to efforts to define the core competencies that are typically central to any profession. Imas (2017) summarizes the global evaluation situation this way, “today any person or group can create their own set of competencies. And indeed, that is not only what is happening but also what is being encouraged” (p. 73). She goes on to point out that “most fields recognized as professions, such as health care, teaching, counseling, and so on, have typically developed competencies . . . by asking a group of distinguished practitioners. . . . to first generate [an] initial list of competencies, then to institute an expert review process to edit and refine them. The competencies are then made available to professionals in the field” (p. 71).

Competencies are typically used to structure education/training programs and guide practice. In the evaluation field, bottom-up efforts continue to dominate efforts to define core competencies (King & Stevahn, 2015). Although more likely to be representative of the range of existing theories and practice, they may trade off breadth with depth. Among the recommendations in an evaluation of the Canadian Evaluation Society Professional Designation Program (Fierro, Galport, Hunt, Codd, & Donaldson, 2016), is one to facilitate recognizing specializations for persons who are successful in acquiring the Credentialed Evaluator (CE) designation. In effect, the 49 competencies that are the basis for the CE assessment process (Canadian Evaluation Society, 2018) would be refined to formally acknowledge different theoretical and methodological approaches to evaluation practice.

One way to approach professionalization is to focus on the steps or stages involved. Altschuld and Austin (2005) suggest there are three stages: credentialing, certification, and licensing for practitioners. *Credentialing* involves demonstrating completion of specified requirements (knowledge, skills, experience, and education/training). A profession that credentials its practitioners offers this step on a voluntary basis and cannot exclude practitioners who do not obtain the credential. The Canadian Evaluation Society Credentialed Evaluator designation is such a program (Canadian Evaluation Society, 2018). *Certification* involves testing competencies and other professional attributes via an independent testing process that may involve examinations and practice requirements (practicums or internships, for example). Typically, those who pass the certification process are issued document(s) attesting to their competence to be

practitioners. The profession cannot exclude those who do not seek (voluntary) certification or who fail the process. Finally, *licensing* involves government jurisdictions issuing permits to practice the profession; persons without a license cannot practice. Persons who are licensed to practice are typically certified; for such professions, certification is a step toward obtaining a license to practice.

Aside from practitioner-focused steps, it is also possible for professions to *accredit* formal education/training programs (typically offered by universities) so that students who complete those programs are certified and can (if appropriate) apply to become licensed practitioners. Accreditation typically involves periodic peer reviews of programs, including the qualifications of those teaching, the resources for the programs, the contents of the program, the qualifications of the students (the demand for the program), and other factors that are deemed to predict student competencies (McDavid & Huse, 2015).

Globally, the prospects for the field of evaluation evolving to be more professionalized are promising, judging by the interest in evaluation and the growth in evaluation-related associations. Some countries (Canada, Britain and Japan) are taking the next step—credentialing evaluators who are interested in differentiating themselves professionally (UK Evaluation Society, 2018; Wilcox & King, 2013). But there is also evidence of limited movement, particularly among those countries that have taken the lead in professionalizing evaluation so far (United States, Canada, Australia and New Zealand) where “the development can be described as stagnation, with even a certain decline in the number of programs (primarily in Psychology)” (Stockmann & Meyer, 2016, p. 337). In the evaluation of the Professional Designation Program (Fierro, Galport, Hunt, Codd, & Donaldson, 2016), the evaluators asked Canadian Evaluation Society Board members “if they believed that recognition of evaluation as a profession in Canada was increasing, decreasing or remaining the same. While no one reported a decrease in recognition, the board members were split on whether it was increasing or remaining the same” (p.17).

Although it is challenging to offer an overall assessment of the future of evaluation, it seems clear that the recognition of evaluation as a separate discipline/profession/body of practice is growing globally. But taking the next steps toward professionalization is far more challenging. The experience of the Canadian Evaluation Society in embarking on a program to credential evaluators suggests that building and sustaining an interest and involvement in evaluation at this next level is promising but not yet assured.

Stockmann and Meyer (2016) sum up their volume on the global prospects for evaluation this way:

To sum up: the global trends for the future of evaluation are still positive, even if many pitfalls can be identified. While evaluation is steadily on the increase, this continuously produces new challenges for the integration of evaluation as a scientific, practical and politically useful endeavor. Today, the shared perspective of being one global evaluation community dominates and many different ways of doing evaluations are accepted. The tasks for the future will be more scientific research on evaluation and improved utilization in public policy. This will be a dance on the volcano—as it ever has been. (p. 357)

Summary

Program evaluation is partly about understanding and applying methodologies and partly about exercising sound professional judgment in a wide range of practice settings. But, because most evaluation settings offer only roughly appropriate opportunities to apply the tools that are often designed for social science research settings, it is essential that evaluators learn the *craft* of working with square pegs for round holes.

This chapter emphasizes the central role played by professional judgment in the practice of professions, including evaluation, and the importance of cultivating sound professional judgment. Michael Patton, through his alter ego Halcolm, puts it this way (Patton, 2008, p. 501):

Forget “judge not and ye shall not be judged.”

The evaluator’s mantra: Judge often and well so that you get better at it.

—Halcolm

Professional judgment is substantially based on experience and our experiences are founded on what we know, what we learn, what we value, and what we believe. Professional judgment has an important ethical component to it. Professional practice consists in part on relying on our knowledge and skills, but it is also grounded in what we believe is right and wrong. Even evaluators who are making “pure methodological decisions” are doing so based on their beliefs about what is right and wrong in each circumstance. Rights and wrongs are *based in part on values*—there is no such thing as a value-free stance in our field—and are based in part on ethics, what is morally right and wrong.

Professional programs, courses in universities, textbooks, and learning experiences are opportunities to learn and practice professional judgment skills. Some of that is tacit—can only learned by experience. Participating in practica, internships, apprenticeships, are all good ways of tying what we can learn from books, teachers, mentors and our peers (working in teams is an asset that way) to what we can “know” experientially.

Although professional guidelines are an asset as we navigate practice settings, they are not enforceable and because they are mostly based on (desired) values, are both general and can even conflict in a given situation. How we navigate those conflicts—how we choose among moral values when we work—is an important part of what defines us as practitioners. In our field there is a growing concern that evaluators should do more to play a role in addressing inequalities and injustices, globally. As our field globalizes, we encounter practice situations where our clients do not want evaluators to address social justice issues. How we respond to these challenges will, in part, define our efforts to become a profession.

Discussion Questions

1. Take a position for or against the following proposition and develop a strong one-page argument that supports your position. This is the proposition: "Be it resolved that experiments, where program and control groups are randomly assigned, are the Gold Standard in evaluating the effectiveness of programs."
2. What do evaluators and program managers have in common? What differences can you think of as well?
3. What is tacit knowledge? How does it differ from public/shareable knowledge?
4. In this chapter, we said that learning to ride a bicycle is partly tacit. For those who want to challenge this statement, try to describe learning how to ride a bicycle so that a person who has never before ridden a bicycle could get on one and ride it right away.
5. What other skills can you think of that are tacit?
6. What is mindfulness, and how can it be used to develop sound professional judgment?
7. Why is teamwork an asset for persons who want to develop sound professional judgment?
8. In this chapter we introduced three different ethical frameworks. Which one aligns most closely with your own ethical approach? Why?
9. What is practical wisdom as an ethical approach in professional practice? How is it different from the three ethical frameworks we introduced in this chapter?
10. What do you think would be required to make evaluation more professional—that is, have the characteristics of a profession?

APPENDIX

Appendix A: Fiona's Choice: An Ethical Dilemma for a Program Evaluator

Fiona Barnes did not feel well as the deputy commissioner's office door closed behind her. She walked back to her office wondering why bad news seems to come on Friday afternoons. Sitting at her desk, she went over the events of the past several days and the decision that lay ahead of her. This was clearly the most difficult situation that she had encountered since her promotion to the position of Director of Evaluation in the Department of Human Services.

Fiona's predicament had begun the day before, when the new commissioner, Fran Atkin, had called a meeting with Fiona and the deputy commissioner. The governor was in a difficult position: In his recent election campaign, he had made potentially conflicting campaign promises. He had promised to reduce taxes and had also promised to maintain existing health and social programs, while balancing the state budget.

The week before, a loud and lengthy meeting of the commissioners in the state government had resulted in a course of action intended to resolve the issue of conflicting election promises. Fran Atkin had been persuaded by the governor that she should meet with the senior staff in her department, and after the meeting, a major evaluation of the department's programs would be announced. The evaluation would provide the governor with some post-election breathing space. But the evaluation results were predetermined—they would be used to justify program cuts. In sum, a "compassionate" but substantial reduction in the department's social programs would be made to ensure the department's contribution to a balanced budget.

As the new commissioner, Fran Atkin relied on her deputy commissioner, Elinor Ames. Elinor had been one of several deputies to continue on under the new administration and had been heavily committed to developing and implementing key programs in the department, under the previous administration. Her success in doing that had been a principal reason why she had been promoted to deputy commissioner.

On Wednesday, the day before the meeting with Fiona, Fran Atkin had met with Elinor Ames to explain the decision reached by the governor, downplaying the contentiousness of the discussion. Fran had acknowledged some discomfort with her position, but she believed her department now had a mandate. Proceeding with it was in the public's interest.

Elinor was upset with the governor's decision. She had fought hard over the years to build the programs in question. Now she was being told to dismantle her legacy—programs she believed in that made up a considerable part of her budget and person-year allocations.

In her meeting with Fiona on Friday afternoon, Elinor had filled Fiona in on the political rationale for the decision to cut human service programs. She also made clear what Fiona had suspected when they had met with the commissioner earlier that week—the outcomes of the evaluation were predetermined: They would show that key programs where substantial resources were tied up were not effective and would be used to justify cuts to the department's programs.

Fiona was upset with the commissioner's intended use of her branch. Elinor, watching Fiona's reactions closely, had expressed some regret over the situation. After some hesitation, she suggested that she and Fiona could work on the evaluation together, "to ensure that it meets our needs and is done according to our standards." After pausing once more, Elinor added, "Of course, Fiona, if you do not feel that the branch has the capabilities needed to undertake this project, we can contract it out. I know some good people in this area."

Fiona was shown to the door and asked to think about it over the weekend.

Fiona Barnes took pride in her growing reputation as a competent and serious director of a good evaluation shop. Her people did good work that was viewed as being honest and fair, and they prided themselves on being able to handle any work that came their way. Elinor Ames had appointed Fiona to the job, and now this.

Your Task

Analyze this case and offer a resolution to Fiona's dilemma. Should Fiona undertake the evaluation project? Should she agree to have the work contracted out? Why?

A. In responding to this case, consider the issues on two levels: (1) look at the issues taking into account Fiona's personal situation and the "benefits and costs" of the options available to her and (2) look at the issues from an organizational standpoint, again weighing the "benefits and the costs". Ultimately, you will have to decide how to weigh the benefits and costs from both Fiona's and the department's standpoints.

B. Then look at this case and address this question: Is there an ethical "bottom line" such that, regardless of the costs and benefits involved, it should guide Fiona's decision. If there is, what is the ethical bottom line? Again, what should Fiona do? Why?

REFERENCES

- Abercrombie, M. L. J. (1960). *The anatomy of judgment: An investigation into the processes of perception and reasoning*. New York: Basic Books.
- Alkin, M. C. (Ed.). (2013). *Evaluation roots: A wider perspective of theorists' views and influences (2nd ed.)*. Thousand Oaks: Sage.
- Altschuld, J. (1999). The certification of evaluators: Highlights from a report submitted to the Board of Directors of the American Evaluation Association. *American Journal of Evaluation*, 20(3), 481–493.
- Altschuld, J. W., & Austin, J. T. (2005). Certification. *Encyclopedia of Evaluation*. Thousand Oaks, CA: Sage.
- Altschuld, J. W., & Engle, M. (Eds.). (2015). *Accreditation, Certification, and Credentialing: Relevant Concerns for U.S. Evaluators: New Directions for Evaluation* (No. 145). Hoboken, NJ: John Wiley & Sons.
- American Evaluation Association. (1995). Guiding principles for evaluators. *New Directions for Program Evaluation*, 66, 19–26.
- American Evaluation Association. (2004). *Guiding principles for evaluators*. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- American Evaluation Association. (2011). *American Evaluation Association Statement on Cultural Competence in Evaluation*. Retrieved from: <http://www.eval.org/ccstatement>
- American Evaluation Association. (2018). *Guiding principles for evaluators*. Retrieved from <http://www.eval.org/p/cm/ld/fid=51>
- Arsene, O., Dumitrache, I., & Miha, I. (2015). Expert system for medicine diagnosis using software agents. *Expert Systems with Applications*, 42(4), 1825–1834.
- Arora, V., Johnson, J., Lovinger, D., Humphrey, H. J., & Meltzer, D. O. (2005). Communication failures in patient sign-out and suggestions for improvement: A critical incident analysis. *BMJ Quality & Safety*, 14(6), 401–407.
- Astbury, B. (2016). Reframing how evaluators think and act: New insights from Ernest House. *Evaluation*, 22(1), 58–71.
- Barber, M. (2007). *Instruction to Deliver: Tony Blair, the Public Services and the Challenge of Achieving Targets*. London Politico's.
- Bickman, L. (1997). Evaluating evaluation: Where do we go from here? *Evaluation Practice*, 18(1), 1–16.
- Canadian Evaluation Society. (2012a). *CES guidelines for ethical conduct*. Retrieved from <https://evaluationcanada.ca/ethics>
- Canadian Evaluation Society. (2012b). *Program evaluation standards*. Retrieved from <https://evaluationcanada.ca/program-evaluation-standards>
- Canadian Evaluation Society. (2018). About the CE Designation. Retrieved from <https://evaluationcanada.ca/ce>
- Chen, H. T. (1996). A comprehensive typology for program evaluation. *Evaluation Practice*, 17(2), 121–130.
- Chouinard, J. A., & Cousins, J. B. (2007). Culturally competent evaluation for Aboriginal communities: A review of the empirical literature. *Journal of MultiDisciplinary Evaluation*, 4(8), 40–57.
- Cox, R. & Pyakuryal, S. (2013). Tacit Knowledge. In H. Fredrickson and R. Ghore (Eds.), *Ethics in Public Management* (pp. 216–239) (2nd Ed.). New York: Sharpe.
- Cronbach, L. J. (1980). *Toward reform of program evaluation* (1st ed.). San Francisco: Jossey-Bass.
- Davies, H., & Kinloch, H. (2000). Critical incident analysis: Facilitating reflection and transfer of learning. In V. Cree and C. Maccaulay (Eds.) *Transfer of Learning in Professional and Vocational Education* (p. 137–147), London, UK: Routledge.
- Dobkin, P., & Hutchinson, T. (2013). Teaching mindfulness in medical school: Where are we now and where are we going? *Medical Education*, 47(8), 768–779.
- Donaldson, S., & Donaldson, S. I. (2015). Visions for using evaluation to develop more equitable societies. In S. Donaldson and R. Picciotto (Eds.), *Evaluation for an equitable society* (pp. 1–10). Charlotte, NC: Information Age Publishing.
- Donaldson, S., Christie, C., & Mark, M. (Eds.). (2015). *Credible and Actionable Evidence: The Foundation for Rigorous and Influential Evaluations*. (2nd Ed.). Thousand Oaks, CA: Sage.
- Donaldson, S., & Picciotto, R. (Eds.) (2016). *Evaluation for an equitable society*. Charlotte, NC: Information Age Publishing.