1

Developmental Evaluation Defined and Positioned

Science has explored the microcosms and the macrocosms; we have a good sense of the lay of the land. The great unexplored frontier is complexity.
—HEINZ R. PAGELS, *The Dreams of Reason* (1988)

The evaluation version of this observation goes like this:

> Evaluation has explored merit and worth, processes and outcomes, formative and summative evaluation; we have a good sense of the lay of the land. The great unexplored frontier is evaluation under conditions of complexity. *Developmental evaluation* explores that frontier.

Developmental evaluation supports innovation *development* to guide adaptation to emergent and dynamic realities in complex environments. Innovations can take the form of new projects, programs, products, organizational changes, policy reforms, and system interventions. A complex system is characterized by a large number of interacting and interdependent elements in which there is no central control; self-organizing and emergent behaviors based on sophisticated information processing generate learning, evolution, and development

(Mitchell, 2009, p. 13). Complex environments for social interventions and innovations are those in which what to do to solve problems is uncertain and key stakeholders are in conflict about how to proceed. Informed by systems thinking and sensitive to complex nonlinear dynamics, developmental evaluation supports social innovation and adaptive management. Evaluation processes include asking evaluative questions, applying evaluation logic, and gathering real-time data to inform ongoing decision making and adaptations. The evaluator is often part of a development team whose members collaborate to conceptualize, design, and test new approaches in a long-term, ongoing process of continuous development, adaptation, and experimentation, keenly sensitive to unintended results and side effects. The evaluator's primary function in the team is to infuse team discussions with evaluative questions, thinking, and data, and to facilitate systematic data-based reflection

and decision making in the developmental process.

The Distinction Emerges

Developmental evaluation as a distinct niche emerged in response to one of my client's questions and needs. It happened like this.

I had a standard 5-year contract with a community leadership program that specified 2½ years of formative evaluation for program improvement to be followed by 2½ years of summative evaluation that would lead to an overall decision about whether the program was effective, a common design and sequence. The leadership program served small, rural communities throughout Minnesota. During the formative evaluation, the program made major changes in many aspects of how it operated. Recruitment processes were expanded. Program activities were adjusted based on feedback from participants. New curriculum elements and small-group exercises were added and fine-tuned. Follow-up interviews with graduates led to new support initiatives after program completion. Formative evaluation focuses on improving a model. This program team was hungry for feedback and eager to make improvements, which they had done willingly and enthusiastically. Then it came time to close this highly creative phase of formative evaluation and move on to summative evaluation.

On a subzero February morning in northern Minnesota, I opened a program team meeting by announcing:

"We've had a great couple of years changing and adapting the program. I've been impressed by your openness and commitment to use evaluation feedback to make improvements. But now, in the next phase of the evaluation, called *summative evaluation*, the purpose is to make an overall judgment about the merit and worth of the program. Does it work? Should it be continued, perhaps even expanded? Have you come up with a model that others might want to adopt? This means that from now on you can't make any more improvements or changes because we need the program—the model—to stay stable in order to conduct the summative evaluation. Only with a fixed intervention, carefully implemented the same way for each new group of leaders in training, can we attribute the measured outcomes to your program intervention in a valid and credible way."

Mouths fell open. Staff was aghast. They protested:

"We don't want to implement a fixed model. In fact, what we've learned is that we need to keep adapting what we do to the particular needs of new groups. Communities vary. The backgrounds of our participants vary. The economic and political context keeps changing. New technologies like the Internet are coming into rural Minnesota and creating new leadership challenges. Small communities are becoming parts of regional networks. We need to get more young people into the program. Immigrants are moving into rural Minnesota in droves, creating more diverse communities. We need to reach out and adapt what we do to Native Americans. No! No! No! We can't fix the model. We can't stand still for 2 years. We don't want to do summative evaluation."

"But that's what my contract specifies," I replied, disconcerted by their resistance. "This is the way things work," I hastened to explain. "You do a couple of years of formative evaluation to stabilize the program model, then you do summative evaluation to determine if it works, if the targeted outcomes are achieved. That's how things work. That's standard practice."

"But that doesn't make sense for us. We'll just have to change the contract," the direc-

tor offered. "Let's just keep doing formative evaluation. We want to keep improving the program."

"Then when do you want to do the summative evaluation?" I asked.

"Never," he responded without hesitation, "not if it means standardizing the program. We want to keep developing and changing."

"But the purpose of formative evaluation is to get ready for summative evaluation. At some point, you'll need to determine if the model works. At some point the board will need to decide whether to keep funding this program. People outside the program are interested and asking if it works. That means stabilizing the model to do a summative evaluation."

He looked at me sternly, challengingly. "Formative evaluation! Summative evaluation! *Is that all you evaluators have to offer*?"

Frustration, even hostility, was palpable in his tone. I found myself feeling defensive. In truth, those were the field's primary distinctions. That was, in fact, all we had to offer. "Well," I said, seeking inspiration in my coffee cup, "I suppose we could do, umm, we could, umm, well, we might do, you know . . . we could try *developmental evaluation*!"

"What's that?" asked the director.

"It's where you, ummm, keep developing."

"That's what we want to do," he said, obviously relieved. "We can make periodic reports on our developments to the board and to others interested in what we're doing and learning, but we want to keep developing. *Developmental evaluation*. I like it. Let's do that. So, how do we do it?"

"Well, it's kind of a new approach," I said, thinking to myself, like 1-minute new. "But it does seem to fit what you want to do, so I'm sure we can figure it out together." And thus began my foray into and education about developmental evaluation, a learn-by-doing process that has been, and continues to be, "developmental." This books reports what I've learned. But first let me finish the story.

My two evaluation colleagues and I became part of the leadership program's de-sign team, which included a sociologist, a couple of psychologists, a communications specialist, some adult educators, a philanthropic funder, and program training and professional development staff. Our evaluation role was to bring evaluative thinking and data to bear as the team conceptualized, developed, and tried out new approaches for new groups, including immigrants, Native Americans, people from distressed rural communities, elected officials, and young people. The program developed new approaches in light of new federal and state policies affecting rural communities. The ongoing decline in many rural communities led to a more regional focus. As more than one cohort from a community went through the program, the issue of how to connect different cohorts arose. New funding opportunities opened up to support follow-up projects by program graduates. New staffing needs arose. The developmental relationship lasted over 6 years and involved different evaluation designs each year including participant observation, several different surveys, field observations, telephone interviews, face-to-face interviews, focus groups, case studies of individuals and communities, cost analyses, theory-of-change conceptualizations, futuring exercises, and training participants to do their own community-based evaluations. Each year the program changed in significant ways and new evaluation questions emerged. Program goals and strategies evolved. The evaluation evolved. No summative evaluation was ever conducted, no final report was ever written. The program continues to evolve—and continues to rely on developmental evaluation.

Periodic summative-type decisions were made along the way in that the foundation board had to budget to continue funding, sometimes approving major changes in strategic direction and augmenting funding accordingly. Developmental evaluation supported these summative decisions by the board by documenting the nature and results of program developments. What was

judged to be working, however, was not a standardized and routinized model, but rather the ongoing development of leadership programming in response to changing conditions, lessons learned, and the emergent needs of different kinds of participants as the program expanded its outreach.

So, is the distinction between formative and developmental evaluation meaningful? Is it worth distinguishing *improvements* from *developments*? It has certainly proved meaningful and useful to those with whom I work. I think it's valuable to respect and maintain the original connection between formative and summative evaluation, that formative evaluation gets a program model ready for summative testing. I also think, as

my experience with the community leadership program illustrates, that developmental evaluation has a distinct purpose and niche beyond formative and summative evaluation. This book is about that niche.

I hasten to add that I am in no way denigrating of or hostile to formative and summative evaluation, nor am I suggesting that these approaches lack value. Quite the contrary. The point is that each approach, including developmental evaluation, fulfills a specific purpose and adds a particular kind of value. Indeed, in Chapter 7 we'll examine the niche of *preformative* use of developmental evaluation: development of an innovative idea or visionary intervention during a period of exploration to get the emerging model to the point where it is ready for traditional formative and summative evaluation with particular focus on determining if the innovation is a potential model that is scalable for broad impact. Let me elaborate.

Facing Complexity and Facing Reality: Or, Facing the Realities of Complexity
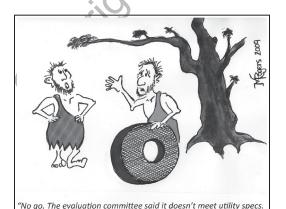
As I've discovered over the last decade, developmental evaluation as a distinct approach to evaluation has proven especially relevant and attractive to social innovators. These people are trying to bring about major social change by fighting poverty, homelessness, community and family violence, and by helping people with AIDS, severe disabilities, chronic diseases, and victims of natural disasters and war. Some of the daunting challenges social innovators face include skepticism, criticism, naysayers, disbelievers, and the ever-present very real possibility of failure, perhaps even the likelihood of failure. Canadian colleagues Frances Westley and Brenda Zimmerman and I studied successful social innovations and visionary social innovators. We reported what we found in a book entitled *Getting to Maybe: How the World Is Changed* (Westley, Zimmerman, & Patton, 2006). We found that fierce con-

---

⌒ **Why Distinctions Matter**

Language matters. Terminology matters. Distinctions matter. That great scholar and observer of all things human, Dr. Seuss (1953), illustrated the consequences of not making distinctions in his children's story "Too Many Daves." Mrs. McCave, it seems, had 23 sons and she named them all Dave. When she wanted one particular Dave and called out his name all 23 Daves came on the run.

Same thing happens if you don't distinguish types of evaluations. An entire volume of *New Directions for Evaluation* was devoted to *How and Why Language Matters in Evaluation* (Hopson, 2000).

This book is about developmental evaluation as a distinct type with its own name.

Edward Sapir (1884–1939), the great linguist and anthropologist, made the same point as Dr. Seuss, but with a bit more of an academic voice:

Human beings are very much at the mercy of the particular language which has become the medium of expression for their society. . . . We see and hear and otherwise experience very largely as we do because the language habits of our community predispose certain choices of interpretation. (quoted in Rheingold, 1988, p. 11)

viction is required to sustain innovation in the face of mounting internal and external obstacles. To be a change agent is to think boldly, to envision grandly. Complexity theory shows that great changes can emerge from small actions. Change involves a belief in the possible, even the "impossible." Moreover, social innovators don't follow a linear pathway of change; there are ups and downs, roller-coaster rides along cascades of dynamic interactions, unexpected and unanticipated divergences, tipping points and critical mass momentum shifts. Indeed, things often get worse before they get better as systems change creates resistance to and pushback against the new.

Traditional evaluation approaches are not well suited for such turbulence. Traditional evaluation aims to control and predict, to bring order to chaos. Developmental evaluation accepts such turbulence as the way the world of social innovation unfolds in the face of complexity. Developmental evaluation adapts to the realities of complex nonlinear dynamics rather than trying to impose order and certainty on a disorderly and uncertain world.

In general I've found that evaluation has a bad reputation among visionaries. This is for a variety of reasons, some fair, some not so fair. Leaders tend to attract and sur-round themselves with believers: true believers, positive thinkers, and hope-springs-eternalists. This adds to the momentum and the flow of social innovation, which is particularly critical in the early stages. Criticism is well known to undermine creativity—which is why it's outlawed in brainstorming exercises. Visionaries, then, often eschew criticism, especially early in the process while creating a vision and recruiting allies and followers. Energy being always in short supply, those aiming to change the world focus their energy on what can be done, on strengths, not weaknesses.

In addition, many of those working in the domain of social innovation, including social entrepreneurs and inventors (Conger, 2009), have experienced evaluation methods that seem entirely unrelated to the nature of their enterprise. Identifying clear, specific, and measurable outcomes at the very start of an innovative project, for example, may be not only difficult but counterproductive. "Outcomes will emerge as we engage," say the social innovators.

"Not in my world," respond the funders and the evaluators. "Clear goals have to be established before you engage. And you need an explicit change model, a logic model to show how you'll attain your goals."

"Not in my world," respond the social innovators. "Time is of the essence and there's no time to lose. Every minute matters. We have to dive in and see what we can do." And thus is the battle between funders, evaluators commissioned by funders, and social innovators enjoined.

Unfortunately, resistance to evaluation can undermine social innovation if and when it becomes a resistance to reality testing. And evaluation is ultimately about reality testing, getting real about what's going on, what's being achieved—examining both what's working and what's not working. Jim Collins (2001), author of the best-selling management book *Good to Great*, studied with his research team how good companies become "great." Not many companies qualified for his study. Few made the transition

"No go. The evaluation committee said it doesn't meet utility specs. They want something linear, stable, controllable, and targeted to reach a pre-set destination. They couldn't see any use for this."

Illustrated by Mark M. Rogers.

⌒ Simple Reactions to Complexity: Context for Developmental Evaluation as an Approach to Complexity

Over the years I've reviewed a large number of evaluation guidebooks, position papers, terms of reference, and scopes of work that explain to evaluators how to deal with complexity. The common themes that have struck me are a two-pronged effort to first deny complexity (redefine the complex as simple) and then, failing that, to control it. Here are sample prescriptions I've collected that give you a sense of a prevailing worldview that gives rise to developmental evaluation as an alternative for dealing with complexity, to wit, actually acknowledging and dealing with it.

Conventional prescriptions for denying complexity

- Yes, the world is complex, but don't let that become an excuse. *Simplify and focus*. Things only appear complex because you haven't yet focused.
- What needs to be done only seems complex when you lack a framework for how to intervene. A clear framework simplifies the complex, makes it manageable, and tells you where to target your resources.
- Cut through the noise and find the essence. Don't be distracted by complexity. Get on with taking action and making a difference on key indicators. Move the needle of those indicators and complexity will take care of itself.

Conventional prescriptions for controlling complexity

- In the face of complexity, the first task is to identify clear, specific, and measurable goals. Clear direction and measurable goals cut right through complexity.
- Everything seems complex until you do a logic model. Sort out the complexities into a sequence of concrete actions that are clear, sequential, and logical.
- At its most effective and useful, evaluation makes the uncertain certain, the ambiguous unambiguous, the unknown known, the unpredictable predictable, and the complex simple.
- Accountability requires that programs manage and control complexity. Evaluation makes that possible.

Conventional quotations found in evaluation documents to justify avoiding complexity

- Nothing is more simple than greatness; indeed, to be simple is to be great.—Poet Ralph Waldo Emerson
- Nothing is particularly hard if you divide it into small jobs.—Industrialist Henry Ford
- Nothing is true, but that which is simple.—German literary luminary Johann Wolfgang von Goethe
- Life is really simple, but men insist on making it complicated.—Chinese philosopher Confucius
- Things should be made as simple as possible, but not any simpler.—Physicist Albert Einstein

And it is Einstein's wisdom that informs developmental evaluation and this book.

from good to great, but those that did all had leaders who lived the paradox between absolute dedication to a great vision and ruthless commitment to staring reality in the face. Collins called this the "Stockdale paradox" in honor of James Stockdale, the fabled U.S. Navy officer who survived years of torture in North Vietnamese prisons. Stockdale had an unwavering belief that he would survive and an equally unrelenting vigilance about his prisoner-of-war reality. He was constantly attuned to what was hap-

pening to him and his fellow prisoners, and adapted his survival strategies and tactics accordingly. When, after a short period of unusual good treatment, he realized that he was about to be used as propaganda to show how well prisoners were cared for, he brutalized his own face so that he could not be so used—or misused. Wondering how Stockdale managed to stay ever hopeful in the face of this day-to-day brutal reality, Collins asked him how he would characterize those who didn't make it, those who died in captivity. That's easy, Stockdale replied immediately: they were the optimists, those who said they'd be out by Christmas, and then by Easter, and then by summer's end, and then again by Christmas, always and only focusing on some future target of hope. They died, he said, of broken hearts.

The "good to great" companies Collins's team studied all shared an unrelenting belief in a future that seemed to those around them a delusion *and* an obsession with data about the reality they faced, monitoring the results of their initiatives and getting real-time feedback about what was working and not working, and how their environment was changing. They did not treat vision and reality testing, hope and data, as opposites. Rather, they immersed themselves paradoxically in vision-directed reality testing: no rose-colored glasses, no blind spots, no positive thinking. *Ruthless attention to reality was the common path to attaining their visions.*

The key to reconciling the tension between optimism and pessimism, dreaming and reality testing, is to tailor the methods of evaluation to the demands of innovation by tracking emergent and changing realities, illuminating perspectives about realities, and feeding back meaningful findings in real time so that reality testing facilitates and supports the dynamics of innovation. This is not simple to do, but it can be critical for adapting and sustaining social innovation. Developmental evaluation is designed to be congruent with and to nurture developmental, emergent, innovative, and transformative processes.

Developmental Evaluation and Complexity Theory

We have entered the *Age of Adapting Quickly.*
—MICHIKO KAKUTANI, Pulitzer Prize–
winning critic for the *New York Times*
(2009, p. C1)

Complexity as a construct is a broad tapestry that weaves together several threads relevant to innovation and evaluation. Exhibit 1.1 summarizes some complexity concepts that we'll be using throughout this book: nonlinearity, emergence, dynamical systems, adaptiveness, uncertainty, and coevolutionary processes. Innovation as something new, emergent, and adaptive exhibits characteristics and dynamics associated with complex adaptive systems. Developmental evaluation likewise centers on situational sensitivity, responsiveness, and adaptation, and is an approach to evaluation especially appropriate for situations of high uncertainty where what may and does emerge is relatively unpredictable and uncontrollable. Developmental evaluation tracks and attempts to make sense of what emerges under conditions of complexity, documenting and interpreting the dynamics, interactions, and interdependencies that occur as innovations unfold.

Positioning developmental evaluation as especially appropriate for complex situations requires a brief excursion into systems thinking and complexity theory. Chapters 4 and 5 examine these ideas and their implications in depth. As prologue, it's worth warning that this is treacherous terrain, easy to get lost in. Once, when hiking a rugged wilderness area of the Grand Canyon, I missed one switchback on the descent and started down the wrong drainage. Within 15 minutes I recognized my error, but I was on a steep slope run through with drainages, ravines, and ridges, converging, diverging, and crisscrossing. It took a couple of hours trying one direction and then another to find my way back to my companions. Hiking the Grand Canyon wilderness away from the main tourist trails, I learned, offered many

---

EXHIBIT 1.1 Characteristics of Complex Adaptive Systems

Nonlinearity. Sensitivity to initial conditions; small actions can stimulate large reactions, thus the *butterfly wings* (Gleick, 1987) and *black swans* (Taleb, 2007) metaphors, in which highly improbable, unpredictable, and unexpected events have huge impacts.

Emergence. Patterns emerge from self-organization among interacting agents. What emerges is beyond, outside of, and oblivious to any notion of shared intentionality. Each agent or elements pursues its own path but as paths intersect and the elements interact, patterns of interaction emerge and the whole of the interactions becomes greater than the separate parts.

Dynamical. Interactions within, between, and among subsystems and parts within systems are volatile, turbulent, cascading rapidly and unpredictably.

Adaptive. Interacting elements and agents respond and adapt to each other so that what emerges and evolves is a function of ongoing adaptation among both interacting elements and the responsive relationships interacting agents have with their environment.

Uncertainty. Under conditions of complexity, processes and outcomes are unpredictable, uncontrollable, and unknowable in advance. *Getting to Maybe* (Westley et al., 2006) captures the sense that interventions under conditions of complexity take place in a *Maybe World*.

Coevolutionary. As interacting and adaptive agents self-organize, ongoing connections emerge that become *coevolutionary* as the agents evolve together (coevolve) within and as part of the whole system, over time.

Note. Exhibit 5.6 in Chapter 5 presents the developmental evaluation implications of each of these dimensions of complexity.

---

opportunities to get sidetracked and lost, and we unwittingly and unexpectedly got lured into unplanned sidetrack adventures. Such uncertain and emergent adventures are reported by even the most expert Canyon hikers, like the renowned Harvey Butchart, who spent more than 1,000 days hiking the Canyon, covered some 12,000 miles, recorded 23 first ascents, and was often lost, sometimes with dire consequences, including the death of a young hiking companion (Butler & Myers, 2007; Patton, 1999).

Yes, sidetracks. Unexpected detours. Getting lost. Navigating tough terrain. Negotiating ravines and ridges. Steep ascents and terrifying descents. Diverging, converging, and crisscrossing. Watching for what emerges. Expecting the unexpected. Going with the flow. Riding cascades and waves of turbulence. These are the allusions and metaphors of complexity. And of developmental evaluation. Complexity writings are filled with metaphors that try to make complex phenomena understandable to the human brain's hardwired need for order, meaning, patterns, sense making, and control, ever feeding our illusion that we know what's going on. We often don't. But the pretense that we do is comforting—and sometimes necessary for some effort at action.

So complexity theorists talk of flapping butterfly wings that change weather systems and spawn hurricanes, individual slime molds that remarkably self-organize into organic wholes, ant colonies whose frantic service to the queen mesmerize us with their collective intelligence, avalanches that reconfigure mountain ecologies, bacteria that *know* the systems of which they are a part without any capacity for self-knowledge, and

∽ *Framing Poverty
as a Complex Issue*

Mark Cabaj (2009a, 2009b) is a developmental evaluator working with Vibrant Communities, a comprehensive and innovative antipoverty program working across Canada. He and his colleagues recently synthesized 8 years of learning from the collaborations across Canada involved in Vibrant Communities. At the top of their list of important learnings is the importance of viewing poverty through the lens of complexity.

Unlike simple or complicated issues that can be effectively addressed by employing best practices or extensive research and planning, poverty is a complex issue. This means that it:

- Is difficult to define;
- Has tangled up root causes;
- Involves stakeholders with diverse values, interests, and positions;
- Varies from person to person and community to community;
- Is constantly evolving; and
- Has no obvious answers or measures of success

The developmental evaluation work with Vibrant Communities has meant using an adaptive approach to mobilizing stakeholders, crafting and evaluating strategies, and stewarding a long-term effort characterized by unavoidable tensions, fast-moving environments, and blunt and clumsy practices implemented by traditional organizations.

Alan Perlis, an award-winning computer scientist, once observed:

Fools ignore complexity. Pragmatists suffer it. Some can avoid it. Geniuses remove it. Simplicity does not precede complexity, but follows it. (quoted in Cabaj, 2009a)

black swans that appear suddenly and unpredictably to change the world. Complexity science offers insights into the billions of interactions in the global stock market, the spread of disease throughout the world, volatile weather systems, the evolution of species, large-scale ecological changes, and the flocking of migrating birds. Complexity theorists explain the rise and fall of civilizations, and the rise and fall of romantic infatuation. That's a lot of territory. I aim merely to add attention to the rise and fall of evaluations.

Utility

What brings me to complexity is its utility for understanding certain evaluation challenges. Complexity concepts can be used to identify and frame a set of intervention circumstances that are amenable to a particular situationally appropriate evaluation response, what I am calling here developmental evaluation. This makes dealing with complexity a defining characteristic of developmental evaluation's niche. Principles for operating in complex adaptive systems inform the practice of developmental evaluation. The controversies and challenges that come with complexity ideas will also and inevitably afflict developmental evaluation. The insights and understandings of complexity thinking that have attracted the attention of and garnered enthusiasm from social innovators will also envelope developmental evaluation—and be the source of its utility. Forewarned is forearmed. You are entering here the world of uncertain beginnings, muddled middles, and unpredictable endings that ripple on and on without end. This is the paradoxical comfort zone of people like photographer-provocateur Robert Frank who cursed "those god-damned stories with a beginning and an end" (quoted in Lane, 2009, p. 88). For those with a high tolerance for ambiguity and a grand sense of adventure, this is an exciting world. For those with big control needs who prize predictability and strive for certainty, not so much.

Dealing with the Unexpected

There is no such thing as a failed experiment,
only experiments with unexpected outcomes.
—R. BUCKMINSTER FULLER (1895–1983),
visionary and inventor

Developmental evaluation requires what distinguished and experienced evaluator Jon Morell has called "agile evaluators," those who learn to expect the unexpected and adapt with agility and flexibility, including changing the evaluation design, reconfiguring program theory, and responding to emergent stakeholder needs (Morell, 2010). There is a lot of lip service in evaluation about looking for unanticipated consequences and assessing side effects; in reality, these are typically token elements of evaluation designs, inadequately budgeted, and rarely given serious time and attention because of the overwhelming focus on measuring attainment of intended outcomes and tracking preconceived performance indicators. You have to go out into the real world, do fieldwork, engage in open inquiry, talk to participants in programs, and observe what

⌁ *Global Complexity, Local Complexity*

As this book was being written in 2008–2009, the news was saturated with evidence of global complexity: the global economic meltdown and financial crisis that began in October 2008 revealed the complex nonlinear dynamics of the interconnected and interdependent global economy, replete with uncertainties and tsunami-like ripple effects. The election of Barack Obama to the presidency of the United States changed the global political landscape, a fact highlighted when he was awarded the 2009 Nobel Peace Prize, not for anything in particular, but for everything in general. The long-term effects of technological innovation and the Internet are still unfolding, the implications just beginning to become evident. The threat of a worldwide flu pandemic has cascading effects on tourism, commerce, travel, and community life.

These global complexities spiral downward and become manifest in local uncertainties and unexpected developments. When the endowments of philanthropic foundations were hard hit, some programs and agencies that thought they had funding were suddenly without financing and had to close their doors. Local and state governments face huge and growing deficits, with uncertain consequences. The U.S. federal deficit is increasing at an unprecedented rate. Demographic trends, especially the unparalleled aging of the population, are creating new demands for services at every level of society. The effects of health care reform will not be known for years and are far from settled. Climate change looms. No one these days doubts the uncertain but very real dynamics of global change.

From a big-picture global systems perspective, these complex phenomena are interconnected: economic, political, demographic, environmental, social, cultural, technological, and health systems interlocked, interacting, and interdependent—with unknown and unpredictable consequences. All of this is sometimes labeled CONTEXT by evaluators creating fixed and static logic models that pretend and assume control and predictability: implement these activities and produce these outcomes. But global complexities and dynamics are not just context. They manifest themselves in local realities: changed conditions under which programs operate, new problems that participants bring to programs, and new challenges in meeting emergent needs. These well-documented and pervasive complexities have become part of public consciousness replete with local evidence of what theorists call *nonlinear dynamics* and common folk capture with the bumper sticker slogan "Shit Happens."

I don't find that it takes a lot of effort to convince people that the world is complex. The evidence is all about them. The question is *how to respond and adapt to that complexity*. That's no longer just a question for those trying to bring about change and those trying to survive change. It's a question for those evaluating change. *How do evaluators respond and adapt to the realities of complexity?* Developmental evaluation is one response.

is going on as innovations unfold to detect unanticipated consequences. In contrast to the casual and if-we-get-to-it-and-have-time-and-resources-after-everything-else-is-done way that evaluators typically approach the unexpected and unanticipated, the possibilities of unexpected impacts become likelihoods under conditions of complexity and developmental evaluators make expecting the unexpected fundamental to the work at hand. Organizational development researchers Weick and Sutcliffe (2001) found that high-performance organizations are always on the lookout for the unexpected. So are high-performing developmental evaluators.

Developmental Evaluation and Single-Loop versus Double-Loop Learning

Developmental evaluation supports learning to inform action that makes a difference. This often means changing systems, which involves getting beyond surface learning to deeper understandings of what's happening in a system. Social innovators and social entrepreneurs are typically trying to bring about fundamental changes in systems, to change the world (Bornstein, 2007). To do so, they have to understand how the system they want to change is operating and make changes that get beyond temporary and surface solutions to change the system itself. This involves *double-loop learning*.

For decades three stories have been endlessly repeated: one about the stream of ambulances at the bottom of the cliff instead of building fences at the top; one about the numerous dead bodies floating down the river while all we do is build more impressive services for fishing them out; and one about giving someone a fish versus the value of teaching that person how to fish. In reviewing these stories, distinguished Australian action research scholar and practitioner Yolande Wadsworth (2010) has commented that they are reminders about our

repeated tendency to go for the short-term quick fix rather than to examine, come to understand, and take action to change how a system is functioning that creates the very problems being addressed. Double-loop learning involves systemic solutions and is supported by evaluation attuned to looking for system explanations and offering systemic insights. Chapter 5 explores in depth how systems thinking informs developmental evaluation.

Argyris and Schön (1978) distinguished single-loop from double-loop learning. In single-loop learning, people modify their actions as they evaluate the difference between desired and actual outcomes and make changes to increase attainment of desired outcomes. In essence, a problem-detection-and-correction process is *single-loop* learning. Single-loop learning is like a thermostat that knows when it is too hot or too cold and turns the heat off or on. The thermostat can perform this task because it can receive information (the temperature of the room) and take immediate corrective action.

In double-loop learning, those involved go beyond the single loop of identifying the problem and finding a solution to a second loop that involves questioning the assumptions, policies, practices, values, and system dynamics that led to the problem in the first place and intervening in ways that involve the modification of underlying system relationships and functioning. Making changes to improve immediate outcomes is single-loop learning; making changes to the system either to prevent the problem or to embed the solution in a changed system involves double-loop learning.

An Example of Double-Loop Learning

Harvard Medical School surgeon Atul Gawande (2007a) tells of visiting the Walter Reed military hospital early in the Iraq War. He participated in a session interpreting eye-injury statistics. The doctors were having considerable success saving some soldiers

from blindness, a positive outcome. But digging deeper, the doctors asked why so many severe eye injuries were occurring. Interviewing their patients, they learned that the young soldiers weren't wearing their protective goggles because they considered them too ugly and uncool. They recommended that the military switch to "cooler-looking Wiley X ballistic eyewear. The soldiers wore their eyegear more consistently and the eye-injury rate dropped immediately" (p. A23). By asking these kinds of deeper questions about what's really going on and questioning basic assumptions about why things are happening, developmental evaluators help get at fundamental systems change implications and understandings. That's double-loop learning.

The Importance of Interpretive Frameworks

Management scholars Kathleen Sutcliffe and Klaus Weber (2003) examined the performance of business organizations in relation to the amount and accuracy of information used by senior executives as well as the "interpretive frameworks" they used to make sense of information. In a *Harvard Business Review* article they concluded that *the way senior executives interpret their business environment is more important for performance than the accuracy of data they have about their environment.* That is, they concluded that there was less value in spending a lot of money increasing the marginal accuracy of data available to senior executives compared to the value of enhancing *their capacity to interpret* whatever data they had. Executives were more limited by a lack of capacity to make sense of data than by inadequate or inaccurate data. In essence, they found that interpretive capacity, or "mindsets," distinguish high performance more than data quality and accuracy. After all, they concluded, the role of senior managers isn't just to make decisions; it's to set direction and motivate

⌒ **Real-Time versus Developmental Evaluation**

"Real time" refers generally to rapid feedback and response, linking data and action as close together in time as possible. The ultimate in real-time data analysis is reporting on stock market transitions in microseconds. In hospitals, real time means getting blood analyses or other diagnostic tests back to a doctor within a short time line that can range from minutes to an hour. In evaluation situations, real time typically means getting results to intended users in a day or two, or at most a couple of weeks, rather than in months or on a routine schedule of standard quarterly reports (a common information system reporting time frame).

Developmental evaluation aims for real-time feedback, but *not all real-time data use and evaluation is developmental*. Police departments use real-time data on increasing crime in a neighborhood to reallocate personnel from lower crime to higher crime areas. That is real-time evaluation and data use, but it is not developmental. This real-time use of data by police involves implementing a rapid response management approach, but the police are not developing that approach. In contrast, if crime data in a community indicated a national gang was moving into the community, the police could develop a task force to fight gang recruitment, infiltration, and crime and monitor emergent effects as the gang adapted to police attention so that police could adapt accordingly. That would be developmental evaluation because the intervention is emerging in real time and using evaluation data to adapt the intervention to what emerges in real time.

others in the face of ambiguities and conflicting demands. In the end, top executives must manage meaning as much as they must manage information.

Enhancing the quality and accuracy of our evaluation data through better methods and measures will add little value unless those using the data have the capacity to think evaluatively and critically, and be able to appropriately interpret findings to reach reasonable and supportable conclusions. Systems thinking, complexity theory, and developmental evaluation together offer an interpretive framework for engaging in sense making. As a complexity-sensitive, developmental evaluation unfolds, social innovators observe where they are at a moment in time and make adjustments based on dialogue about what's possible and what's desirable, though the criteria for what's "desirable" may be quite situational and always subject to change.

Developmental Evaluation and Accountability

Complexity-based developmental evaluation shifts the locus and focus of accountability. Traditionally accountability has focused on and been directed to external authorities and funders. Accountability-focused evaluators report independently to decision makers charged with making sure that resources are spent on what they're supposed to be spent on.

In contrast, for vision-and-values-driven social innovators the highest form of accountability is internal. Are we walking the talk? Are we being true to our vision? Are we dealing with reality? Are we connecting the dots between here-and-now reality and our vision? And how do we know? What are we observing that's different, that's emerging? These become internalized questions, asked ferociously, continuously, because they want to know. Those funding innovations join in the questioning and need to understand that the seriousness of inquiry

and resulting learning constitutes accountability.

That doesn't mean that asking such questions and engaging the answers, as uncertain as they may be, is easy. It takes courage to face the possibility that one is deluding oneself. Here the individual's sense of internal and personal accountability connects with a group's sense of collective responsibility and ultimately connects back to the macro, to engage the question of institutional and societal accountability. Throughout such discussions about accountability, the focus remains: What is getting developed? With what implications?

Developmental Evaluation as Utilization-Focused

Developmental evaluation is meant to communicate that there is an option in and approach to conducting evaluations that specifically supports *development*. This book will elucidate the niche, methods, and challenges of conducting developmental evaluations. In so doing, I place this approach within the larger context of *utilization-focused evaluation* (Patton, 2008c). Since utilization-focused evaluation is what I am best known for and most closely associated with, let me take a moment to make explicit how developmental evaluation flows from and can be positioned within the larger context and framework of utilization-focused evaluation.

Utilization-focused evaluation is evaluation done for and with specific primary intended users for specific, intended uses. Utilization-focused evaluation begins with the premise that evaluations should be judged by their utility and actual use; therefore, evaluators should facilitate the evaluation process and design any evaluation with careful consideration for how everything that is done, from beginning to end, will affect use. Use concerns how real people in the real world apply evaluation findings and experience the evaluation process. Therefore,

the focus in utilization-focused evaluation is on achieving *intended use by intended users.* In developmental evaluation, the intended use is development, which I shall argue is a distinct and important evaluation purpose. The primary intended users are social innovators and others working to bring about major change.

In any evaluation there are many potential stakeholders and an array of possible uses. Utilization-focused evaluation requires moving from the general and abstract, that is, possible audiences and potential uses, to the real and specific: actual primary intended users and their explicit commitments to concrete, specific uses. The evaluator facilitates judgment, decision making, and action by intended users. Developmental evaluation, conducted from a utilization-focused perspective, facilitates ongoing innovation by helping those engaged in innovation examine the effects of their actions, shape and formulate hypotheses about what will result from their actions, and test their hypotheses about how to foment change in the face of uncertainty in situations characterized by complexity.

The utilization-focused approach is personal and situational. The evaluation facilitator develops a working relationship with intended users to help them determine what kind of evaluation they need. This requires negotiation in which the evaluator offers a menu of possibilities within the framework of established evaluation standards and principles. Thus, while concern about utility drives a utilization-focused evaluation, the evaluator must also attend to the evaluation's accuracy, feasibility, and propriety (Joint Committee on Standards, 1994). Moreover, as a professional, the evaluator has a responsibility to act in accordance with the profession's adopted principles of conducting systematic, data-based inquiries; performing competently; ensuring the honesty and integrity of the entire evaluation process; respecting the people involved in and affected by the evaluation; and being sensitive to the diversity of interests and values that may be related to the general and public welfare (American Evaluation Association [AEA], 1995).

Utilization-focused evaluation does not advocate any particular evaluation content, model, method, theory, or even use. Rather, it is a process for helping primary intended users select the most appropriate content, model, methods, theory, and uses for their particular situation. Situational responsiveness guides the interactive process between evaluator and primary intended users. This book presents and discusses developmental evaluation as one of the options now available in the feast that has become the field of evaluation. Utilization-focused evaluation can include any evaluative purpose (formative, summative, developmental), any kind of data (quantitative, qualitative, mixed), any kind of design (e.g., naturalistic, experimental), and any kind of focus (processes, outcomes, impacts, costs, and cost–benefit, among many possibilities). Utilization-focused evaluation is a process for making decisions about these issues in collaboration with an identified group of primary users focusing on their intended uses of evaluation.

A psychology of use undergirds and informs utilization-focused evaluation. In essence, research and my own experience indicate that intended users are more likely to use evaluations if they understand and feel ownership of the evaluation process and findings; they are more likely to understand and feel ownership if they've been actively involved; and by actively involving primary intended users, the evaluator is training users in use, preparing the groundwork for use, and reinforcing the intended utility of the evaluation every step along the way. Developmental evaluation carries this user involvement further than usual by creating a dynamic partnership between social innovators and the developmental evaluator. How that partnership gets built, and its potential pluses and minuses, will be one of the subjects I'll elucidate later. It is sufficient to say at this point that the language of "partnership" is not the norm in describing the re-

lationship between an evaluator and those whose work is being evaluated. Thus, developmental evaluation invites both skepticism and controversy. We'll deal with both along the way.

Situation Recognition

Astute situation recognition is at the heart of utilization-focused evaluation. There is no one best way to conduct an evaluation. This insight is critical. The design of a particular evaluation depends on the people involved and their situation. The standards and principles of evaluation provide overall direction, a foundation of ethical guidance, and a commitment to professional competence and integrity, but there are no absolute rules an evaluator can follow to know exactly what to do with specific users in a particular situation. Recognizing this challenge, situation analysis is one of the "essential competencies for program evaluators" (Canadian Evaluation Society, 2010; Ghere, King, Stevahn, & Minnema, 2006; King, Stevahn, Ghere, & Minnema, 2001).

The idea—admittedly an ideal—is to match the type of evaluation to the situation and needs of the intended users to achieve their intended uses. This means—and I want to emphasize this point—*developmental evaluation is not appropriate for every situation.*

$\sim$ **Situational Practice as an Evaluation Competency**

The Canadian Evaluation Society has adopted five domains of competencies for Canadian evaluation practice, one of which is *situational practice*, defined as follows:

> *Situational Practice* competencies focus on the application of evaluative thinking in analyzing and attending to the unique interests, issues, and contextual circumstances in which evaluation skills are being applied. (*www.evaluationcanada.ca/txt/20090531_competencies_companion.pdf*)

Not even close. Indeed, I shall argue that its niche is small and demanding. It will not work if the conditions and relationships are not right. I'll be specifying what those conditions and relationships are as we proceed. The point here is that every evaluation involves the challenge of matching the evaluation process and approach to the circumstances, resources, time lines, data demands, politics, intended users, and purposes of a particular situation. Matching requires astute situation recognition. This is not as easy as it may sound. Indeed, it is quite difficult and worth understanding why, so a brief excursion into breakthrough understandings in cognitive science and philosophy of science, heavy-going stuff, is worth mentioning as a context for understanding and framing developmental evaluation.

Substantial research has focused on human nonrationality, including the influential works of Nobel Prize in Economics recipient Daniel Kahneman, one of many who have established that how we decide what to do is far from rational.[1] Our rationality is "bounded" (Simon, 1957, 1978). This applies no less to well-educated professionals than to common folk. We all rely on deeply embedded heuristics, rules of thumb, standard operating procedures, practiced behaviors, and selective perceptions. We operate within and see the world through paradigms. A *paradigm* is a worldview built on implicit assumptions, accepted definitions, comfortable habits, values defended as truths, and beliefs projected as reality. As such, paradigms are deeply embedded in the socialization of adherents and practitioners. Our paradigms tell us what is important, legitimate, and reasonable. Paradigms are also normative, telling us what to do without the necessity of long existential
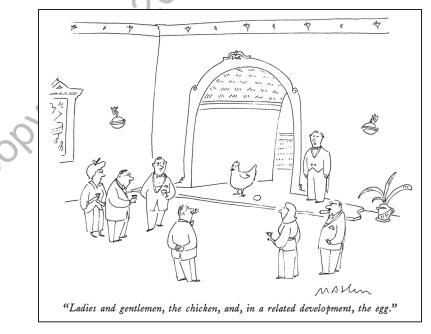
---

[1]For samples of a half-century of research on the nonrational nature of decision making, see Gigerenzer, Todd, and ABC Research Group (1999); Groopman (2007); Inbar (1979); Kahneman and Tversky (2000); Kuhn (1970); Simon (1957, 1978); Thaler and Sunstein (2009); Tversky and Fox (2000); and Tversky and Kahneman (2000).

or epistemological consideration. But it is this aspect of paradigms that constitutes both their strength and their weakness— their strength in that it makes action possible, their weakness in that the very reason for action is hidden in the unquestioned assumptions of the paradigm. This is now widely understood and generally accepted, but it is worth taking a moment to revisit the insights of Thomas Kuhn (1970) regarding how paradigms work. This excerpt is from his influential classic *The Structure of Scientific Revolutions*:

> Scientists work from models acquired through education and subsequent exposure to the literature, often without quite knowing or needing to know what characteristics have given these models the status of community paradigms. . . . That scientists do not usually ask or debate what makes a particular problem or solution legitimate tempts us to suppose that, at least intuitively, they know the answer. But it may only indicate that neither the question nor the answer is felt to be relevant to their research. Paradigms may be prior to, more binding, and more complete than any set of

rules for research that could be unequivocally abstracted from them. (p. 46)

That's what we're up against when we set forth the ideal of matching the evaluation to the nature of the situation. I repeat, then: *developmental evaluation is not appropriate for every situation.* This book will detail when it is appropriate. Chapter 7, for example, looks at when and how a developmental evaluation may generate a promising model that an innovator wants to take to scale, so the appropriate evaluation of that model becomes traditional formative and summative evaluation to assess its scalability and capacity for dissemination. It is also worth noting that developmental evaluation may appear alien to evaluators trained only in the traditional and dominant evaluation research paradigm, so it can evoke their hostility. The stakes can be high. Reactions to paradigm departures can be fierce. Evaluation distinctions matter because evaluation matters in this manic and politicized world of outcomes accountability. More on that later, too. So here's where we're headed.

"Ladies and gentlemen, the chicken, and, in a related development, the egg."

Charting the Developmental Journey: Overview of the Book

Chapter 2, *Developmental Evaluation as a Distinct Purpose and Niche*, looks more deeply at the role and distinct contributions of developmental evaluation, including the implications of offering an option beyond formative and summative evaluation, the classic distinctions that have dominated evaluation for four decades. There are two distinct niches for developmental evaluation. The first is to support exploration and innovation *before* there is a program model to improve and summatively test. In that sense, developmental evaluation is preformative, but can lead to generation of a model that is subsequently evaluated formatively and summatively. The second niche is for those dynamic situations, like the one involving the leadership program example discussed earlier in this chapter, where program staff and funders expect to keep developing and adapting the program, so they never intend to conduct a final summative evaluation of a standardized and hypothesized best practice model. This niche is nonsummative in that it doesn't render an overall judgment of merit and worth about whether a model is effective and worthy of adoption by others, but rather supports ongoing real-time decisions about what to change, expand, close out, or further develop. The chapter will emphasize the differences between improvement versus development, and the implications of that distinction for evaluation practice. We'll also look at ongoing strategic thinking versus periodic strategic planning, positioning developmental evaluation as a form of thinking and acting strategically as an innovative intervention unfolds.

Chapter 3, *Thinking Outside Evaluation's Boxes*, introduces an extensive case example of program development and evaluation's role in supporting that development. The case illustrates and deepens our understanding of the implications of distinguishing program *improvement* from program *development*, while opening up discussion of developmental evaluation facilitated and conducted by both internal and external evaluators (it can be done by either or by both together). The case example illustrates some of the constraints that arise in complex development situations when traditional evaluation approaches are inappropriately imposed. In looking at situational responsiveness and matching an evaluation to the circumstances in which the program is operating and unfolding, we'll consider the pragmatic questions: *What is sensible evaluation? How do we decide what makes sense? What does it mean to be pragmatic?* This requires considering dominant notions of accountability and common barriers to evaluation utility and actual use, including cautions about misevaluation, misuse, and corruption of evaluation. The chapter closes with 10 key points about developmental evaluation illustrated by the case example. These include the importance of timely engagement and rapid feedback, and how evaluation can become the engine for program development such that ongoing program development and evaluation become mutually reinforcing, a way of doing business—indeed, a way of thinking. Project leadership and support for doing developmental evaluation are crucial, as are competent evaluators attuned to the challenges of developmental evaluation.

Chapter 4, *Situation Recognition and Responsiveness*, provides a framework for distinguishing simple, complicated, and complex situations, and the evaluation implications of these distinctions. *Complexity* is defined as those situations where uncertainty about what to do is high because both knowledge is insufficient and key stakeholders are in substantial conflict. The dynamics and uncertainties of complex adaptive systems make what to do to solve problems and change systems essentially *unknowable* in advance—thus the need for trying things out and quickly assessing what happens and what emerges, both intended and unintended, to inform the next steps in exploration, experimentation, innovation, and development.

The chapter closes with an evaluator's guide to decision making about design priorities matched to different contexts (simple, complicated, and complex).

Chapter 5, *Systems Thinking and Complexity Concepts for Developmental Evaluation*, opens with an excursion into systems thinking and its implications for evaluation, especially in contrast to the linear logic models that so completely dominate current evaluation thinking. Logic models have contributed tremendously to clarifying the IT question in evaluation: When we say IT works, or IT doesn't work, what is the IT? The program logic model describes the IT, which is why it has become the dominant and preferred tool in designing evaluations. But the very notion of an IT connotes a static, fixed, and mechanical cause–effect model where inputs lead to activities, which lead to outputs, which produce outcomes and impacts. That works well in simple situations of high certainty and high agreement about what to do. But such modeling has significant downsides and distorting effects in complex and dynamic situations where the IT is emergent, evolving, and adapting. Systems thinking and mapping offers an alternative to linear logic modeling. Having established that alternative framework, we'll examine the implications for developmental evaluation of the six characteristics of complex adaptive systems presented in Exhibit 1.1: nonlinearity, emergence, dynamic systems, adaptiveness, uncertainty, and coevolutionary processes.

Chapter 6, *How the World Is Changed: A Dialectic*, opens by considering the thesis that the world is changed top-down through widespread dissemination of best practices (a predominant theory of change). The antithesis or opposing proposition is that the world is changed bottom-up through grassroots adaptations of effective principles attuned to local contexts. I proceed to reject both the thesis and the antithesis and propose a complexity-sensitive, developmental evaluation synthesis position: *In the global village, change occurs in the middle where top-down and bottom-up forces collide, intersect, get entangled together, do battle, and otherwise encounter real-world complexities.* In considering this action-in-the-middle synthesis, we'll distinguish best practices from effective principles as a form of evaluation finding, and examine the implications of the distinction for both theories of change and evaluation results. We'll look at how this action-in-the-middle played out in an actual program example. The chapter concludes:

> When the primary source of change is bottom-up, the developmental evaluator helps local innovators take a broader systems perspective, including understanding and attending to larger cross-scale forces that can affect the success of local action, helping them draw on knowledge and principles from elsewhere. When the primary source of change is top-down, the developmental evaluator helps conceptualize and test local adaptations, as appropriate. When the sources for change are simultaneously top-down ("It's blowin' in the wind") and bottom-up ("All politics is local"), the developmental evaluator helps facilitate and navigate the interactive dynamics of the muddled middle.

Chapter 6 also looks at developmental evaluation of networks of change (in contrast to programs and discrete interventions). The chapter concludes with further elaboration of developmental evaluation questions for different situations.

Chapter 7, *The Adaptive Cycle and Developmental Evaluation*, looks at the concept of and research on ecosystem resilience and its implications for both social innovation and developmental evaluation. Ecologists studying the health and resilience of forests have found that these complex ecological systems adapt to fires, disease, and periods of drought through four phases that make up a recurring adaptive cycle: *release* (forest fire or other destruction); *reorganization/exploration* (new growth); *exploitation* (accelerated growth of some varieties over others in the competition for resources); and *conservation* (a mature forest dominated by one species).

This cycling through phases, with major transitions from one stage to another, can be observed not only in healthy ecosystems, but also in resilient social systems. However, if adaptation doesn't occur from one phase to another, the health of the system, or the organization, is threatened. The adaptive cycle has significant implications for evaluation with different approaches to and types of evaluation appropriate for different phases of the adaptive cycle. Developmental evaluation is especially well suited for the reorganization/exploration phase. This chapter aims to deepen our understanding of what it means to match evaluation to particular situations, and offers another framework for doing so. I'll also use the adaptive cycle to discuss and illustrate how developmental evaluation can generate an intervention model that leads to formative and then summative evaluation as an innovation moves through the phases of the cycle.

Chapter 8 examines *Developmental Evaluation Inquiry Frameworks.* Developmental evaluation focuses on developmental questions: What's being developed? How is what's being developed (what's emerging) to be judged? Given what's been developed so far (what has emerged), what's next? The developmental evaluator inquires into *developments*, tracks *developments*, facilitates interpretation of *developments* and their significance, and engages with innovators, change agents, program staff, participants in the process, and funders around making judgments about what is being developed, what has been developed, and the next stages of development. That's the broad panorama. But within that broad panorama, specific questions relevant to specific developmental process and impacts still have to be generated. And there are lots and lots of frameworks for generating and focusing questions. Since a dominant theme throughout the book is situational matching, this chapter offers guidance in how to decide which questions to use to frame a developmental evaluation inquiry. We'll look at 10 distinct inquiry frameworks as examples of alterna-

tive ways of focusing developmental evaluations based on the dynamics of the complex situation in which the evaluation is being undertaken and the predilections and worldviews of those engaged in social innovation.

Chapter 9, *Developmental Evaluation Bricolage*, is about the developmental evaluator as *bricoleur*, a kind of jack-of-all-trades do-it-yourself person who draws on eclectic traditions and integrates diverse approaches to get the job done usefully in a way that fits the situation at hand. The *bricolage* in the chapter will include reflective practice, action research, sensitizing concepts, abductive reasoning, systems change, methodological diversity, and retrospective developmental evaluation. We will look in depth at how reflective practice focused on an innovative sensitizing concept (like the idea of *innovation* itself, or systems change, or social justice) can be a powerful developmental evaluation approach for facilitating ongoing learning, engagement, and adaptation. We'll also take a brief look at pragmatism as one of the epistemological underpinnings for developmental evaluation.

Finally, Chapter 10 examines *Utilization-Focused Developmental Evaluation*, with a look at the implications of focusing on intended use by intended users for engagement practices, diverse designs, and adaptive methods. Developmental evaluation does not rely on any particular evaluation method, design, or tool. A developmental evaluation can include any kind of data (quantitative, qualitative, mixed), any kind of design (e.g., naturalistic, experimental), and any kind of focus (processes, outcomes, impacts, costs, and cost–benefit, among many possibilities), depending on the nature and stage of an innovation and the priority questions that will support development of and decision making about the innovation. This can include randomized controlled trials, surveys, focus groups, interviews, observations, performance data, community indicators, network analysis—whatever sheds light on key questions. Given the infinite possibilities, Chap-

⌒ **Developmental Evaluation versus Development Evaluation**

Developmental evaluation is easily confused with development evaluation. They are not the same, though developmental evaluation can be used in development evaluations. Confused? You are not alone. Read on.

Development evaluation is a generic term for evaluations conducted in developing countries, usually focused on the effectiveness of international aid programs and agencies (e.g., Carlsson, Eriksson-Baaz, Fallenius, & Lövgren, 1999; De Coninck, Chaturvedi, Haagsma, Griffioen, & van der Glas, 2008; Hanna & Picciotto, 2002; Independent Evaluation Group, 2009; Picciotto, 2002). *The Road to Results: Designing and Conducting Development Evaluations* (Imas & Rist, 2009) is an exemplar of this genre, a book based on the World Bank's highly successful International Program for Development Evaluation Training (IPDET), which the book's authors founded and direct, and on which their book is based. Full disclosure: I have been on the IPDET faculty since the program began.

Developmental evaluation, as defined and described in the *Encyclopedia of Evaluation* (Mathison, 2005, p. 116), has the purpose of helping develop an innovation, intervention, or program. In developmental evaluation the evaluator typically becomes part of the program or innovation design team, fully participating in decisions and facilitating discussion about how to evaluate whatever happens. All team members together interpret evaluation findings, analyze implications, and apply results to the next stage of development. The evaluator becomes involved in improving the intervention and uses evaluative approaches to facilitate ongoing program, project, product, staff, and/or organizational *development*. The evaluator's primary function in the team is to facilitate and elucidate team discussions by infusing evaluative questions, data, and logic, and to support data-based decision making in the developmental process. In this regard, developmental evaluation is analogous to research and development (R & D) units in which the evaluative perspective is internalized in and integrated into the operating unit. In playing the role of developmental evaluator, the evaluator helps make an intervention's development an R & D activity.
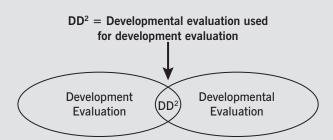
Part of the value of an experienced developmental evaluator to an innovation team is bringing a reservoir of knowledge (based on many years of practice and having read a great many evaluation reports) about what kinds of things tend to work and where to anticipate problems. Experienced evaluators have typically accumulated a great deal of knowledge and wisdom about what works and what doesn't work. More generally, as a profession, the field of evaluation has generated a great deal of knowledge about patterns of effectiveness. That knowledge makes evaluators valuable partners in designing as well as evaluating social innovations.

An evaluation focused on development assistance in developing countries could use a developmental evaluation approach, especially if such developmental assistance is viewed as occurring under conditions of complexity with a focus on adaptation to local context. Many of the examples in this book are of development evaluations, especially Chapter 3. But developmental evaluations are by no means limited to projects in developing countries. Developmental evaluation can be used anywhere that social innovators are engaged in bringing about systems change under conditions of complexity.

The *al* in *developmental* is easily missed, but it is critical in distinguishing development evaluation from developmental evaluation.

DD² = Developmental evaluation used for development evaluation

When I first labeled and wrote about *developmental evaluation* 15 years ago (Patton, 1994), development evaluation was not a distinct and visible category of evaluation practice and scholarship. Evaluations in developing countries were certainly being conducted, but an identifiable body of literature focused on evaluating development assistance had not attracted general professional attention. One of the most important trends of the last decade has been the rapid diffusion of evaluation throughout the world (Patton, 2008c, chap. 1), including especially the developing world, highlighted by formation of the International Development Evaluation Association (IDEAS, 2009), which launched in Beijing, China, in 2002.

Confusion about the distinct and sometimes overlapping niches of development evaluation and developmental evaluation is now, I'm afraid, part of the complex landscape of international evaluation. I hope this book helps sort out both the distinctions and the areas of overlap.

ter 10 presents 10 developmental evaluation engagement approaches and evaluation designs that are particularly appropriate for different complex systems challenges. The examples are meant to be generative and suggestive of the great variety of methods that can be used, not by any means prescriptive or exhaustive of design and methods possibilities. The specific developmental evaluation examples will be presented in a utilization-focused evaluation template so that for each example the following are specified: nature of the complex systems challenge; primary developmental purpose; primary intended users and developmental evaluation partners; key developmental evaluation questions; time line for feedback; and appropriate matching developmental evaluation engagement approach, design, and methods options. Examples presented will include rapid feedback interviews with program participants, bellwether surveys of influential policymakers, participatory action research, social network analysis, and randomized comparison trials of advocacy campaign messages.

Five Developmental Evaluation Purposes and Uses

As the book unfolds, I'll be making the case that developmental evaluation is particularly appropriate for but needs to be matched to five different complex situations and developmental purposes.

1. *Ongoing development* in adapting a project, program, strategy, policy, or other innovative initiative to new conditions in complex dynamic systems (the focus of Chapters 2, 3, 4, and 5).

2. *Adapting effective general principles to a new context* as ideas and innovations are taken from elsewhere and developed within a new setting, the work of developmental evaluation in the dynamic middle between top-down and bottom-up forces of change (the focus of Chapter 6).

3. *Developing a rapid response* in the face of a sudden major change or a crisis, like a natural disaster or financial meltdown, exploring real-time solutions and gen-

erating innovative and helpful interventions for those in need.

4. *Preformative development of a potentially scalable innovation* to the point where it is ready for traditional formative and summative evaluation; preformative developmental evaluation works with emerging ideas and visionary hopes in a period of exploration to shape them into a potential model that is a more fully conceptualized, potentially scalable intervention. (As models emerge out of exploratory and innovative initiatives, some may move into more traditional formative and summative evaluation to determine scalability and generalizability, while others remain in developmental mode, either undergoing further development or continuous experimentation in the search for new models.)

5. *Major systems change and cross-scale developmental evaluation*, providing feedback about how major systems change is unfolding, evidence of emergent tipping points, and/or how an innovation is or may need to be changed and adapted as it is taken to scale, that is, as its principles are shared and disseminated in an effort to have broader impact (discussed in Chapter 7). Horizontal scaling across systems or vertical scaling to broader systems may involve more than adaptation; these dissemination and scaling processes can evolve an essentially new development, the emergence of which can be documented and analyzed as part of a developmental evaluation.

Exhibit 10.1 at the beginning of Chapter 10 summarizes these five purposes including identifying particular complex systems challenges that give rise to each, primary specific developmental evaluation uses appropriate for each type, real-world examples of each with specific primary intended users for each type, and the implications of the different types for evaluation and social innovation.

Throughout the book I'll be positioning developmental evaluation as serving these five particular purposes and uses that, taken together, are different approaches for and windows into developing and evaluating social innovations. These five different uses of developmental evaluation match different situations. They provide different lenses through which to understand and engage in evaluating social innovations under conditions of complexity. Taken together they constitute a specific niche in the large and diverse field of evaluation.

Exhibit 1.2 provides an overview of the niche of developmental evaluation. I've contrasted developmental evaluation generally with some broad-brush traditional approaches to evaluation to help position developmental evaluation in the many-starred evaluation universe. These comparisons and contrasts are meant to be suggestive and illuminative, not definitive. Any one contrast is arguable, possibly overgeneralized, and oversimplified. Viewed as a whole, however, I hope the integration of these many elements provides a sense of what developmental evaluation offers.

But Exhibit 1.2 presents a lot of elements to keep track of and put together. So, bottom line: *How can you tell if an evaluation is truly developmental?* I'll offer a more sophisticated answer as the book unfolds, but let's start simply with purpose and outcomes: *Is the purpose and focus of the evaluation helping develop something? Is something getting developed? Did something get developed? If so, what? How? With what implications?* The focus of developmental evaluation is on (drum roll, please) *developing innovations.*

To borrow an old saying, the proof of the pudding is *in the eating.* Since the next chapter distinguishes *developments* from *improvements,* and will position developmental evaluation as different in important ways from formative and summative evaluation, let me offer this segue. Distinguished evaluation theorist and practitioner Bob Stake has explained, "When the cook tastes the soup, that's formative; when the guests taste the soup, that's summa-

EXHIBIT 1.2 **Contrasts between Traditional Evaluations and Complexity-Sensitive Developmental Evaluation**

Introduction and cautionary note: Evaluation is a diverse field with many models, approaches, methods, and purposes. Any generalizations about predominant tendencies in *traditional* evaluation (both formative and summative) are bound to be overgeneralizations. These contrasts are offered as a heuristic device to suggest thematic tendencies and general distinctions. The themes in the right column define developmental evaluation's niche. To focus that niche, I've contrasted developmental evaluation with some broad-brush traditional approaches to help position developmental evaluation in the many-starred evaluation universe. In doing so I emphasize that these comparisons and contrasts are meant to be suggestive and illuminative, not definitive. Any one contrast is arguable and oversimplified, and may not apply to a particular situation. Viewed as a whole, however, I hope the integration of these many elements provides a sense of what developmental evaluation offers in toto—and in tone.

	Traditional program evaluation tendencies	Complexity-sensitive developmental evaluation
1. Purpose and situation		
1.1. Evaluation purposes	Formative–summative distinction dominant: formative improves; summative tests, proves, and validates program models; accountability.	Supports *development* of innovations and adaptation of interventions in dynamic environments.
1.2. Situation where it is appropriate	Manageable and stable situation; root cause of the problem being addressed is known and bounded; intervention reasonably well conceptualized; goals known; the key variables expected to affect outcomes are controllable, measurable, and predictable.	Complex, dynamic environment; no known solution to priority problems; no certain way forward and multiple pathways possible; need for innovation, exploration, and social experimentation.
1.3. Dominant niche and mindset	Finding out if a program model works: focus on effectiveness, efficiency, impact, and scalability.	Exploring possibilities; generating ideas and trying them out; preformal model, so preformative; nonsummative in that ongoing innovation and development is expected, never arriving at a fixed intervention.
2. Focus and target of evaluation		
2.1. Target of change	Identified outcomes for intended program beneficiaries and participants; change in individual behaviors and performance indicators.	Systems change along a continuum from small local systems to disruptive social innovations aimed at major, cross-scale impacts on big problems.

(cont.)

	Traditional program evaluation tendencies	Complexity-sensitive developmental evaluation
2.2. Driving force of the intervention	Outcomes-driven; systems viewed as context.	Systems-change-driven; specific outcomes emergent, dynamic.
2.3. Evaluation results focus	*Formative*: improve and fine-tune the model; prepare for summative evaluation. *Summative*: Render overall judgments of merit and worth, success or failure.	*Development*: provide timely feedback for development; generate learnings and support action in the development process.
2.4. Evaluation locus	Evaluation is top-down (theory-driven) or bottom-up (participatory).	Evaluation helps innovators navigate *the muddled middle* where top-down and bottom-up forces intersect and often collide.
3. Modeling and methods		
3.1. Modeling approach	Design the evaluation based on a linear cause–effect logic model: specify inputs to activities/processes, then outputs to outcomes to impacts. Causality is modeled, hypothesized, and predicted, then tested.	Design the evaluation using systems thinking to capture and map *complex systems dynamics* and interdependencies, and track emergent interconnections. Causality is based on pattern detection (inference to the best explanation), retrospectively constructed from observations.
3.2. Counterfactuals	Counterfactuals a dominant concern to deal with attribution.	Counterfactual formulations meaningless because of complexity: far too many variables and possibilities emerging and interacting dynamically to conceptualize simple counterfactuals.
3.3. Measurement approach	Measure performance and success against predetermined goals and SMART outcomes: specific, measurable, achievable, realistic, and time-bound.	Develops measures and tracking mechanisms quickly as outcomes emerge; measures can change during the evaluation as the process unfolds. Tracking the forks in the road and implications of key decisions as innovation evolves.
3.4. Attention to unexpected consequences	Typically token attention, if any at all, to unanticipated consequences and side effects.	Expect the unexpected. Serious attention to the unanticipated and emergent as a fundamental evaluation function.

(cont.)

	Traditional program evaluation tendencies	Complexity-sensitive developmental evaluation
3.5. Evaluation design responsibility	Evaluator determines the design based on the evaluator's perspective about what is rigorous. The evaluator has responsibility for and controls the evaluation even when and if stakeholder input is solicited.	Evaluator collaborates with those engaged in the change effort to *co-create* an evaluation that is useful and matches the innovation process philosophically and organizationally.
3.6. Methods approach and philosophy	*Rigorously methods-focused*: an evaluation is judged by validity and methodological criteria first and foremost; utility is viewed as methods-dependent. Traditional research and disciplinary standards of quality dominate.	*Utilization-focused*: methods are chosen in service to developmental use; methods derive from utility and pragmatic considerations; judgments about methodological quality are context-and-intended-use-dependent.
3.7. Interpretation and reasoning processes	Deduction first and foremost; some induction some of the time if qualitative methods used. Attribution analysis.	Abduction (inference to the best explanation) and pragmatism (discussed in Chapter 9). Contribution analysis.
4. Roles and relationships		
4.1. Ideal evaluator stance	Evaluator is independent, whether located internally or externally. Credibility depends on independence.	Evaluator is part of the innovation team, a facilitator and learning coach, bringing evaluative thinking to the group, supportive of the innovators' values and vision. Credibility depends on a mutually respectful relationship.
4.2. Locus and focus of accountability	Accountability focused on and directed to external authorities and funders based on explicit preordinate criteria.	Accountability centered on the innovators' deep sense of fundamental values and commitment to make a difference; funders must buy into what gets developed and learned as the focus of accountability.
4.3. Organizational locus of evaluation	Evaluation often a compliance function delegated down in the organization and/or outside to an external evaluator.	Evaluation a leadership function: nurturing *reality-testing, results-focused, learning-oriented leadership*.

	Traditional program evaluation tendencies	Complexity-sensitive developmental evaluation
5. Evaluation results and impacts		
5.1. Desired and ideal evaluation findings	Validated best practices, generalizable across time and space.	Effective principles that can inform practice and minimum specifications (*min specs*) that can be adapted to local context.
5.2. Evaluation approach to a going-to-scale initiative or model dissemination	In evaluating dissemination of models and taking "best practices" to scale, the focus is on high-fidelity replication.	In evaluating dissemination and going to scale, the focus is on applying principles and adaptation to local context.
5.3. Reporting mode	Often ponderous, detailed formal reports; scholarly voice (third person, passive).	Rapid, real-time feedback. Engaged, present voice (first person, active).
5.4. Impact of evaluation on organizational culture	Evaluation often engenders *fear of failure*.	Evaluation aims to nurture *hunger for learning*.
5.5. Evaluation capacity built through the evaluation process	Usually not an objective; the focus is on getting credible evaluation results based on rigorous methods.	Building ongoing and long-term capacity to think and engage evaluatively is built into the process.
6. Approaches to complexity		
6.1. Approach to uncertainty	Aims for as much certainty and predictability as possible.	Expects uncertainty and unpredictability as givens in complex and dynamic situations.
6.2. Approach to control	Evaluator attempts to control design implementation and the evaluation process.	Learning to respond *to lack of control*; staying in touch with what's unfolding and responding accordingly—and agilely.
7. Professional qualities		
7.1. Key evaluator attributes	Methodological competence and commitment to rigor; independence; credibility with external authorities and funders; analytical and critical thinking.	Methodological flexibility, eclecticism, and adaptability; systems thinking: creative and critical thinking balanced; *high tolerance for ambiguity*; open and agile. Teamwork and people skills: able to facilitate rigorous evidence-based reflection to inform action.
7.2. Evaluation standards and ethics	Knowledgeable about and committed to evaluation's professional standards.	Knowledgeable about and committed to evaluation's professional standards.

tive" (quoted in Scriven, 1991, p. 169). More generally, anything done to the soup during preparation in the kitchen is improvement-oriented; when the soup is served, summative judgment is rendered by the guests who consume the soup. And what of developmental evaluation in this metaphor?

Developmental evaluation begins when, before cooking, the chef goes to the market to see what vegetables are freshest, what fish has just arrived, and meanders through the market considering possibilities, thinking about who the guests will be, what they were served last time, what the weather is like, and considers how adventurous and innovative to be with the meal. If the chef decides to follow a standard recipe, the situation remains appropriate for formative and summative evaluations based on fidelity to the prescribed recipe. If the chef decides to attempt a new creation, innovate, and develop a new dish especially well suited for these particular guests in the context of this particular evening, then the situation opens up the possibility for creativity and developmental evaluation. And when a guest and a cook create and concoct a soup together, that co-creation is developmental.

Bon appétit.