Evolutionary Perspectives on Evaluation: Theoretical and Practical Implications

William Trochim
Cornell University

Paper Presented at the
Colorado Evaluation Network
June 26, 2007

This paper was partially supported through a grant from the National Science Foundation. Evaluation Systems and Systems Evaluation: Building Capacity and Tools for Enhancing STEM Education Evaluation. HER/REC/EREC. NSF Grant #0535492.

The theory of evolution is one of the most important achievements in the history of science. Darwin's *Origin of Species* (Darwin, 1859) and his articulation of the theory of natural selection forms the foundation of virtually all of the life sciences and continues to have profound effects in the social sciences, arts, humanities and, as we all know, in the political and religious realms.

So how is it that the topic of evolution has received so little attention in the field of evaluation? And why is there so little discussion about the implications of evolutionary thinking for evaluation theory and practice? I hope today to encourage us to think about these questions and perhaps to suggest some promising lines of thought. Perhaps I can convey some of my excitement in this topic and a sense of why I think it is so important to our field.

My central premise is a simple one – evaluation cannot be fully understood outside of an evolutionary perspective. It is both influenced by evolutionary forces and influences the evolution of the phenomena it seeks to understand. Our views of evaluation can be deepened by thinking about how evolution relates to it.

I have to warn you at the outset that my understanding of evolution is essentially that of the layperson; my familiarity with the literature does not extend much beyond that available to the general public. But perhaps my technical deficiencies will be a source of encouragement for you and especially for those of you who already know far more than I on this topic.

I would like in this talk to try to state my case as provocatively as I can. I'll begin by sketching briefly what an evolutionary perspective on evaluation might look like. This should help provide a feel for the argument I am making. Then I'll introduce several key topics in evolutionary thinking and discuss some of the practical implications they might have for how we think about evaluation. Throughout this talk, I will limit myself to the arena of program evaluation, although these remarks almost certainly have applicability to other evaluation realms as well.

Preliminary Sketch of an Evolutionary Perspective on Evaluation

Here is my brief and self-consciously provocative sketch of an evolutionary view of program evaluation:

> Every program that we evaluate is essentially an organism[1] in a population of similar programs that constitutes its species[2]. Program theories, whether stated explicitly or not, make up the essential instructions of the program. Within each program species there is variation. Programs have unique characteristics: the people who implement them, the activities that constitute them, the setting and assumptions that guide them, the participants who take part in them. This program variation is essential for their evolution.

> Program variations are implemented, have consequences, and are selected for in subsequent program generations. Some programs, their characteristics and theories survive over time; most become extinct. Programs and program theories get selected and survive because of the fitness of their characteristics to a specific environmental or ecological niche. While most of us probably hope or believe that programs are selected for using rational criteria to yield specific desirable characteristics or outcomes, in many situations they probably survive because people like them, get used to them, or because there are institutional, political and economic forces that favor their survival.
>
> Over time, programs and their theories evolve. This evolution is based on the same principle of natural selection that underlies all evolution in life. The process of consciously developing and evolving programs is a type of artificial selection, a special subtype of natural selection. Evaluation can play a key role in that artificial selection, both in encouraging and enhancing variability and in providing feedback and influencing selection. As in evolution generally, it's not clear where program evolution is heading or whether any adaptation can be said to constitute 'progress.' Slight variations and adaptations can survive that subsequently make little apparent sense. Program features may exist today that were adaptive in the past but remain largely as residuals, long beyond their original adaptive genesis.
>
> Just as with other organisms in nature, in addition to their participation in a broader species, each program has its own individual life, a unique life course that moves through various phases. Programs are born or initiated. They grow and change as they are implemented and revised. They mature and may reach a relatively stable state sometimes becoming routinized and standardized. And, they regenerate, die, are translated and disseminated, and so on, starting new cycles of program instances.

That's it, my preliminary and deliberately provocative sketch of the idea of an evolutionary perspective on evaluation. It is simply a restatement in terms of programs and program theory of the theory of evolution generally. Like the theory of evolution it is simple in conception and readily communicated. And, like that theory it has behind it a world of complexity and implications that I will only begin to explore here.

The Foundation of Evolutionary Epistemology

The foundation of this evolutionary view of programs and program ecosystems actually intersected the field of evaluation, although it is largely neglected and little discussed today. The basis of this line of thought is the subfield of philosophy known as evolutionary epistemology which applies the concepts of biological evolution to the growth and development of human knowledge. The term evolutionary epistemology was reportedly coined by one of the leading thinkers in evaluation, Donald T. Campbell and the field was initially developed by him and the philosopher of science Sir Karl Popper (Popper, 1973, 1975, 1985) in the 1970s and 1980s. In his essay entitled *Evolutionary Epistemology*, Campbell (Campbell, 1974, 1988) argued that "…evolution – even in its biological aspects – is a knowledge process, and that the natural-selection paradigm for such knowledge increments can be generalized to other epistemic activities, such as learning, thought and science."[3] Campbell is not suggesting evolution as a metaphor for learning or thinking or science. He is asserting that evolution is the fundamental process for all of these. And, he is making the argument that biological evolution itself can perhaps most aptly be viewed as a knowledge process.[4]

In his identically titled paper *Evolutionary Epistemology*, Popper (Popper, 1985) describes three levels of evolution: "genetic adaptation, adaptive behavioural learning, and scientific discovery, which is a special case of adaptive behavioural learning" and argues that for all three "the mechanism of adaptation is fundamentally the same." Of course, that mechanism is the process of natural selection. For those who are accustomed to thinking of evolution as something that applies only to biology or genetics, it may initially be somewhat disorienting to accept that both Popper and Campbell are saying that ideas and knowledge follow the exact same process as biological phenomena[5].

Natural selection involves several concepts, most notably the ideas of variation and selection.[6] What does *variation* mean and how does it get generated? Popper maintains that variation in nature and in science is "at least partly random." Campbell uses the term "blind" in his phrase "blind variation and selective retention" (BVSR) that describes natural selection to convey the notion that there is no certain a priori knowledge that a variation will be adaptive. He states (Campbell, 1988): "In going beyond what is already known, one cannot but go blindly. If one can go wisely, this indicates already achieved wisdom of some general sort." Of course, in the world of programs and program theory the process of creating new programs or program variations does not feel blind or random to us. We think of new programs as rational, sometimes obvious, responses to problems or challenges in our environment, not as blind or random endeavors.

But the notion of "blind variation" doesn't mean that our attempts at program construction are irrational – it just means that when we construct them we don't know whether our programs will work or survive. Evaluation certainly plays a role in the program variation generation process. We conduct needs assessments to identify important problems or issues. We construct program theories using methods like concept mapping and logic modeling. We assist planners and program developers in thinking through program implementation challenges and in constructing approaches for formative and process evaluation. All of these can be considered activities that help generate or describe program variation.[7]

The other major component of natural selection is the *selective retention* of better adapted variants. This is the function that was described by Herbert Spencer with the infamous phrase "survival of the fittest" that conjures up the notion of intense competition often associated with evolution in the popular mind. Today we have a much more nuanced sense of what the term "fitness" means and how characteristics are selected for over time. Margulis and Sagan describe this well when they say: "…the view of evolution as chronic bloody competition among individuals and species, a popular distortion of Darwin's notion of 'survival of the fittest,' dissolves before a new view of continual cooperation, strong interaction, and mutual dependence among life forms. Life did not take over the globe by combat, but by networking. Life forms multiplied and complexified by co-opting others, not just by killing them." (pps. 28-29)(Margulis & Sagan, 1986). In the view I am offering here, programs and their theories are selected for over time because they have characteristics that enhance their fitness to their environment.[8]

So, what does all this have to do with evaluation? My contention is that evaluation is inherently a part of Popper's adaptive behavioral learning. It can be considered one of a class of processes whereby we attempt to learn about the world around us. It is a mechanism for creating the conditions for observing the world and providing feedback about it and about the process of observation itself. Where does evaluation fit into evolutionary thinking? Actually, it has a critically important role to play throughout. Of course, evaluation is a form of feedback, and as such is part of the selection process in the evolution of the evaluand. But evaluation is also instrumental in the creation of variations and mutations for future study. Much of evaluation, especially in the past decade, has been concerned with the generation of program theory, logic models, structured conceptualizations or concept maps, and so on. Each of these can be viewed as a variation generation device that potentially stimulates or describes "blind" variations that are subsequently selected for.

Practical Evolutionary Evaluation

With this brief sketch and overview of evolutionary epistemology and the theory of natural selection in mind, let's now illustrate some practical implications for evaluation of evolutionary thinking.

The Cycle of Program Life

An evolutionary evaluation perspective leads us to think differently about programs. Here, we draw on the idea of ontogeny in evolutionary theory. Ontogeny is one of the central concepts in the field of developmental biology. It "…describes the origin and the development of an organism from the fertilized egg to its mature form" ("Ontogeny", 2007, April 15). Instead of thinking of our programs as static entities, this notion encourages us to think of each program as continuously evolving through different phases in a lifecycle. While this lifecycle will manifest itself differently in each program, much as different people develop at different rates at various times in their lives, we can sketch out a hypothetical sequence that would likely fit many programs.

The first stage in this hypothetical program lifecycle might be termed the *initiation phase* and spans from the initial conceptualization through design and initial piloting. The first time you implement a program – even one that was well established elsewhere – you will likely be dealing with the usual issues of initiation: identifying and training program staff, localizing the program to the immediate context, reacting to the unanticipated problems that arise, and so on. The second lifecycle stage might be labeled the *growth or development phase* and would encompass the process of successive revision of the program as it gets implemented repeatedly over time. With programs in this stage, you are getting used to the program and how it plays out in practice. You have some idea of what to anticipate. You might still be surprised by what occurs, and are still adapting the program as you learn, but you are also increasingly able to anticipate problems before they arise, and you are developing a storehouse of experience in how to deal with them. In the third stage, which we might term the *maturity or stability phase*, the program has been carried out at least several times with some degree of implementation success. By this point you pretty much know what to expect. You have stable processes for implementation, and the program is routinized and often has a written protocol or process guide. The program is no longer dependent on particular individuals for implementation. If something happens to the initial implementers of the program it can still be carried out with high fidelity. The fourth stage might be termed the *translation or dissemination phase*. The primary focus of this phase is on extending the program to other settings or populations of interest, pushing the ecological boundaries of the program as originally conceived into new niches or applications.

These stages aren't meant to be a straight-jacket or an inflexible taxonomy. For any given program, the progression may not be sequential. A program may be precocious. It may for instance quickly evolve through the development phase and become stabilized or routinized. Or, a program can revert to an earlier stage, much like the young adult that temporarily reverts to juvenile behavior before resuming more mature development. At any phase, we may decide whether to continue the program or not. Sometimes it is apparent even early in a program's development that it is not able to be implemented well or that it has a fundamental flaw in its conception or structure. In the best of all worlds probably only a small minority of programs will or should survive to the translation or dissemination phase.

Envision for a moment a program that you are familiar with, preferably one that you work closely with or are a participant in. Think about the program and its history. For how long has the program been around? Is it a new program or one that is well established? Is it a routinized one or one that is still developing? Does the program have a well-developed written protocol or process guide? Could a new program implementer take the program materials and do a competent job of implementing it? Is the program ready to be distributed widely and tried in new contexts or with new types of participants? Now, try to classify the program into one of the four general phases of initiation, development, maturity or dissemination. Where would you say the program falls? In many cases, it's probably hard to say, at least at first. Most programs simultaneously involve components or facets that are changing and others that are relatively stable. But if you think about the program as a whole, as an organism, where would you classify it?

This notion of a program lifecycle has practical implications for evaluation. How should a program be evaluated at each stage of its lifecycle? In organizations that are simultaneously running multiple programs – and most organizations do this routinely – what are the advantages of thinking about the collection of programs as constituting a type of portfolio and encouraging variation of programs at different stages of development? What role can evaluators play in helping program administrators and organizations assess where their programs are in their development and in encouraging them to think about when and how they will evolve their programs to their next phase?

The idea of a program lifecycle may strike you as obvious, but it is usually not consciously practiced, and is certainly far from normative. In many of our program contexts, we become committed to the program as it currently exists. The program evolves up to a point and then we get a type of “lock-in” where we seemingly get stuck in a phase and are unable to move any further. Program decisions evolve into a struggle between program preservationists who fear change and the potential loss of their familiar context or even their jobs, and program critics who push for ever-extending demonstrable results and emphasize ever-shrinking funding and resources.

An evolutionary perspective on programs and the idea of ontogeny emphasize program change as a normative expectation. Instead of the commitment to preserving the program as it is, they encourage the idea that programs have a finite life-span, that they should not be assumed to have an infinite horizon, that it is normal to see them as part of an ongoing trial-and-error effort, and that the abandonment of an older program and the development of new ones is part of the normal cycle-of-life. From the beginning of the program, and throughout its evolution, the focus is on where the program is in its development and how to move it to the next phase. In effect the idea of a lifecycle creates system pressure to move programs along and not allow them to become static.

Symbiosis and Co-evolution in Evaluation

The second example of evolutionary thinking I would like to consider is based on the ideas of symbiosis and co-evolution that are critically important in evolutionary biology. One of the most familiar examples of this phenomenon is the relationship of the flower and the bee. Each provides something to the other. The flower provides nectar that is produced into honey and the bee acts as the vehicle for sexual reproduction by moving pollen from one flower to another. Both benefit from the exchange. Neither participates in this exchange consciously. Flowers didn’t strategize one day that they needed bees as a vehicle for reproduction. And bees didn’t decide that flowers would be good vessels for honey production. They co-evolved over millennia in a manner that makes them co-dependent.

People are also integrally involved in co-evolution. There’s a wonderful book that I highly recommend called *The Botany of Desire* (Pollan, 2001) that describes four examples of co-evolution of people and plants. Over time each has been “domesticated” to emphasize certain characteristics that people desire. The apple is associated with our desire for sweetness, the tulip with our love of beauty, the marijuana plant with our yearning for intoxication, and the potato with our continual search for control over the source of our sustenance. But from the evolutionary perspective of each plant, one can say equally well that it is humans who have been domesticated to tend to the plant’s interest, providing them with the regular care and safety necessary for their survival.

Here I want to talk about two types of symbiotic relationships that are important for evaluation. The first involves the co-evolution of programs and the methods used for evaluating them. The second has to do with the relationship between the motivations and incentives that different parties to a program evaluation have regarding the measurement of outcomes.

Methods Matching: Linking Evaluation Methods and Program Phase. If all programs can be understood as evolving through various stages of a lifecycle, then we must imagine that the evaluation approaches we use at each stage need to differ throughout the life of the program. That is, the way we would evaluate a program during its initiation stage would not likely be appropriate for evaluating it during its growth stage, and so on.

In fact, evaluators have known this intuitively for some time now. The basic methodological sequence of exploratory, correlational and experimental research almost surely springs from the same intuition that the phenomenon we are studying needs different methods depending on its stage of development. Michael Scriven's distinction between formative and summative methods (Scriven, 1967) also suggests the importance of yoking method to program phase appropriately.

In effect, the evaluation of a program has its own lifecycle and one of the major tasks of evaluation is to encourage the symbiotic or co-evolutionary relationship between program and evaluation lifecycles. In the *initiation phase* an evaluation needs to be dynamic and flexible, providing rapid feedback about implementation and process. In many program evaluations this is accomplished with simple monitoring or post-only feedback forms, or with unstructured observation and qualitative methods. While we don't often think about the simple sharing of implementation experiences and informal debriefing sessions among implementers as an integral part of evaluation, we probably should. Evaluators can play a critically important role in the initiation stage in helping to craft rapid informal feedback processes, including communication systems for implementers of new programs. In the *development phase* of an evaluation, the focus tends to shift to the observation and assessment of change. During this phase we are designing observational procedures and measures of key outcomes, assessing the consistency and construct validity of measures, looking at pre-post differences and examining the relationships among different observations, qualitative or quantitative. The *mature phase* of an evaluation tends to focus on the idea of control. At this point the program is routinized and stable enough to compare performance of participants with some standard expectation of performance or with outcomes of people who participate in alternative programs or none at all. This is the realm of experimental and quasi-experimental designs and of more structured and comparative qualitative approaches. The *translation or dissemination phase* in evaluation is typically concerned with generalizability or external validity. It examines the consistency of outcomes across different settings, populations or program variations. This is the realm of secondary and meta-analysis and of program review approaches that seek general inferences about the transferability of the program.

This simple idea of the co-evolution of a program with its evaluation has important implications for one of the most vexing methodological debates of our time – the idea that there is a "gold standard" evaluation methodology that is generally preferable to others. In particular, there has been much debate in the past few years about whether the randomized experiment constitutes such a gold standard for evaluating programs. An evolutionary perspective would argue against such generalized thinking, instead preferring an appropriate symbiosis between program and method phases.

It is intriguing that the current emphasis on randomized experiments, especially in education and international development evaluation, has consciously taken as its model the tradition of randomized clinical trials in medicine and health. In those fields it is indeed the case that randomized experiments are a valued approach and they have a long tradition of successfully implementing such studies. But proponents of grafting experimental methods to other fields often neglect the fact that in medicine the experimental paradigm is inextricably embedded within an evolutionary lifecycle model that is normative and well-established. We can see this most clearly in the idea of phased clinical trials in medicine, such as the trials that are required of all new medical and pharmacological treatments before their widespread adoption in practice. Like the four-phase evaluation lifecycle model I sketched earlier, the system of clinical trials does not commence with a randomized experiment. Typically, Phase I clinical trials are small sample exploratory studies that look at dosage levels, implementation difficulties and potential side-effects of the treatment. Phase II trials concentrate on whether the treatment yields measurable changes and examine relationships among observed outcomes. It is in the late Phase II or Phase III trials that more formal controlled designs like the randomized experiment are employed. Phase IV trials are concerned with fidelity of transfer and implementation of treatments in uncontrolled real-world contexts. In true evolutionary fashion, most treatments do not survive through this lifecycle. In fact, nearly three-fourths of all treatments are abandoned before a Phase III randomized experiment is ever mounted (Mayo Clinic, 2007). Advocates of randomized experimentation in other areas like educational or environmental evaluation should be cautious about transplanting that methodology without also bringing along this essential notion of phased studies over the life course of a program.

We should also recognize that for all its strengths, the system of phased clinical trials in medicine is, like all approaches in science, fallible and imperfect. Ernie House at the University of Colorado has made the case that "methodological fundamentalism" that privileges randomized experimental design is fraught with difficulties (House, October, 2004). He and others have argued that drug evaluations in pharmacology have been corrupted by commercial pressure leading to a number of well-publicized errors, and probably many unpublicized ones, that have had significant costs in terms of both human pain and loss of life (Cundiff, 2007; House, November, 2006). While transplanting the notion of randomized experiments within the context of phased studies over the life cycle of a program is almost certainly preferable to just transplanting the emphasis on randomized experiments alone, even both ideas together will be susceptible to significant errors in inference. Our greatest hope is not in any particular methodological approach. As sobering and humbling as it may be, our primary hope for improved inferences over time is in the gradual corrective mechanism of evolutionary trial-and-error.

The idea of matching methods to program phase has great practical utility in evaluation. Just as we imagined how we might place a program in terms of its phase, for each program we might also classify its evaluation by phase. This would enable us to look at the degree to which the evaluation is out of sync with the program. We are likely to find that many existing programs are more developed programmatically than they are in terms of evaluation. Too much of our typical evaluation work seems to involve catching up; after-the-fact cobbling together of evaluation approaches at the eleventh-hour to address looming reporting deadlines of funders and sponsoring organizations. But the program-methods symbiosis challenge poses dangers in the other direction as well – the too-early imposition of evaluation methods that are designed for more stable and mature programs. This is especially the case these days with the rush to premature experimentation before we have adequately conceptualized our programs, gained experience in implementing them or developed observational systems that are consistent, reliable and have strong construct validity. Encouraging an evolutionary view of the symbiosis or co-evolution of program and evaluation would help address these problems.

Social Ecology: Co-Evolution and Incentives in Evolutionary Evaluation. The principle of co-evolution also has important practical implications for measurement in evaluation. In many evaluation contexts, one hears a series of laments related to the motivations and incentives of the different stakeholders. For instance, the evaluator asks "Why don't these program people just cooperate when I ask them for data?" Program implementers ask "Why don't these evaluations address something that would be useful for us?" Program participants want to know "Why do they keep bugging us for data? We don't get anything from this." In the ideal, we would want the situation to be a co-evolutionary one where program participants are providing information naturally as part of their participation, where program administrators are getting what they want from the provided data, and where evaluation happens almost transparently as an integrated aspect of program implementation. That is, the ideal is the flower and the bee. This is a difficult ideal to achieve in practice. It requires that the evaluation be engineered in such a way that each stakeholder group's incentive to participate in the evaluation is well understood.

I can give a simple example of this from an evaluation I am currently working on of an educational biology course. The program is an informal one that is supplementary to the formal classroom education in biology. From an evaluation perspective, we want to know how well the students are learning the content of the program. Giving students a formal test of knowledge would be the most straightforward way to assess knowledge acquisition, but it is an imposition for the classroom teachers and something the students consider a burden.

Our symbiotic approach is to couple the educational program with a separate simultaneous program on test-taking skills for students. We provide instruction in test taking and have students practice their skills on the practice biology exams. Like the flowers and the bees, each party gets something that they want while giving something that the other party needs. Teachers are eager for this assistance in preparing their students, and are willing to give up the time needed for such training. The students get help in getting better test scores that enhance their chances of graduating and are more willing to take the tests we need them to take. And we evaluators get outcome measures of biology knowledge in exchange for providing this extra training. Of course, there are other potential problems that arise from this arrangement including the additional costs and effort and the need to control for test preparation effects in assessing program outcome. But the example illustrates how we might be able to engineer a more useful symbiosis in evaluation measurement, based on principles of dynamic exchange and co-evolution.

Program Ecology: Programs as Species and the Ecology of Programs in Evolutionary Evaluation

An evolutionary evaluation perspective suggests that we think differently about programs in program evaluations. Programs are, by their very nature, conscious interventions. In this sense, the most apt evolutionary parallel is the notion of *selective breeding*. It's interesting to note that Darwin didn't begin the *Origin of Species* by talking about natural selection, the key concept in his theory of evolution. Instead, he started by discussing something that many of his contemporaries were already quite comfortable with, the domestication of animals and plants. He termed the process of selective breeding 'artificial selection' as distinct from the idea of 'natural selection' although he later points out that there is no fundamental difference between the two ideas. It's just that Darwin recognized that the notion of artificial selection would be less threatening and more understandable to his audience.

In many ways, there is a direct parallel between the notion of artificial selection or domestication as described by Darwin and what we evaluators and our colleagues are doing in developing programs. Both are essentially human manipulations. Both are attempts to intervene in the 'natural order' of things. The plant or animal breeder is trying to manipulate a species to a specific set of characteristics or level of performance; a program developer is manipulating the program to achieve certain desired outcomes or levels of performance. Both are interested in refining the object of evolution over time. A plant breeder tries to produce new hybrids that achieve certain desirable characteristics; program developers try to develop better variations of the program to better achieve the desired results that we evaluators attempt to assess. In the ideal world, program developers wish to improve their programs continuously just as animal or plant breeders wish to encourage the evolution of their species along desired lines.

So while program evaluators work locally to improve programs and achieve desired results, they and the people who are designing and implementing the programs, are making a multiplicity of choices based on their local preferences, on what they think makes sense or will work best. They are no more trying consciously to develop a new species at a more global or general level than the gardener who selects the best of the lot of plants and weeds the rest, or the breeder who selects the best animal to produce the next generation of offspring. People were consciously breeding plants and animals long before scientists came up with the theory of evolution and the ideas of artificial and natural selection. But whether evaluators, program designers and program implementers are conscious of this or not, the myriad choices they make and the reflections of those choices in evaluation results are part of a larger evolutionary process. To the extent we are aware of this process, we are engaging in something that is very much like the artificial selection that Darwin described. To the extent that we are not, we are still unwitting participants in the ongoing process of natural selection.

There are important practical implications of the idea that evaluation contributes to the process of artificial selection of programs. At the center of the theory of natural selection is the idea of variation. Without variation there is no basis for comparative advantage, for improved fitness to the environment or ecological niche. The selective breeding of the potato provides an interesting historical example and warning (Pollan, 2001). The potato was originally domesticated in the highland regions of Peru. At the time of the Spanish invasion, the Incas had cultivated over three thousand varieties of potato. In part, this variety was due to the challenging and rugged mountainous climate. Potatoes that could grow successfully in one area would not survive in nearby adjacent areas that differed in altitude, soil, shade, or any number of other factors. In addition "…since the margins and hedgerows of the Andean farm were, and still are, populated by weedy wild potatoes, the farmer's cultivated varieties have regularly crossed with their wild relatives, in the process refreshing the gene pool and producing new hybrids. Whenever one of these new potatoes proves its worth – surviving a drought or storm, say, or winning praise at the dinner table – it is promoted from the margins of the field and, in time, to the neighbors' fields as well" (p. 193)(Pollan, 2001). The conquering Spanish brought only a very small fraction of this great variety back to Europe. In Ireland, essentially one particular variety, the Lumper potato, became a staple food because it was easily cultivated, grew well in otherwise marginal agricultural land and freed the Irish economically and politically from dependence on the British for staple foods. Ireland's population rose from three million to eight million in less than a century. A diet of potatoes and cow's milk was nutritionally sustainable, providing the basic carbohydrates, protein and nutrients needed to survive. But this ended abruptly in the summer of 1845 when a wind-borne fungus arrived, probably on a ship from America. The Lumper, like all potatoes, is a clone, with every one genetically identical, all descended from the same single plant. The fungus quickly ravaged the monoculture of potato crops, even infecting stored potatoes. The resulting famine was the worst catastrophe in terms of human life in Europe since the Black Death in 1348. About a million Irish, roughly one in eight, died and many more went blind or insane because of the loss of vitamins in their diets. Within a decade Ireland lost half its population to starvation or emigration. At the heart of this tragedy was the problem of monoculture. The Incas were never susceptible to a comparable disease because of their potato polyculture.

This story and its evolutionary lessons suggest several practical implications for program evaluation. First, we need to be alert to the dangers of program monocultures. Reliance on centrally sanctioned distribution of single program models, however convenient administratively or successful in the short run, potentially limits the variety that is essential for evolution of programs and program theory. With the recent rise of evidence-based practice as a movement and its corresponding dissemination mechanisms such as the Cochrane Collaboration in medicine, the Community Guide in public health and the What Works Clearinghouse in education we need to be concerned about the potential dangers of encouraging universal adoption of specific programs or practices. We should be developing models of evidence-based practice that explicitly encourage continual trial-and-error of program varieties while noting currently favored solutions. Second, we need to develop better methods for linking program implementers with evaluative feedback and for enabling rapid translating and transplanting of promising programs, much like the informal network of Inca farmers and their rapid trial-and-error and distribution of promising hybrids. Finally, although we program evaluators tend to focus on the programs immediately in front of us, we need to recognize the broader program ecologies within which our programs live. New technologies, and especially the World Wide Web, offer special promise for encouraging broader and more effective networking of programs and people, of practitioners and evaluators.

Conclusions

The three broad topics I've discussed here – ontogeny and the program lifecycle; symbiosis and co-evolution and their corresponding methods matching and incentives analysis; and the issues of variety, polyculture and program ecology – are just an initial sampling of how evolution and evolutionary thinking can have important practical implications for evaluation.

There is a lot more potential in this area than I have been able to touch on today. In a recently published book entitled *Evolution for Everyone*, the author reports on a small e-mail survey of social scientists who published articles that used evolution and evolutionary thinking. He wrote them to discover how they had acquired their expertise on evolution and he discovered that most of them "…received their formal training in other areas (such as psychology, anthropology or linguistics) with little or no exposure to evolution in college and graduate school" and that "…they encountered evolutionary theory on their own, often by happenstance, and gradually built up their expertise until it became a guiding force in their research" (p. 7)(Wilson, 2007). This has certainly been my experience as well. I hope that this talk stimulates you to venture into the broad area of evolutionary theory if you haven't already, to wander into those untravelled aisles of your local bookstore that introduce the topic to public audiences, and to examine its potential applicability for program evaluation.

Endnotes

[1] The definition of organism is very broad and there is considerable dispute about its boundaries. While the common sense meaning of the term would seem to exclude an entity like a program be described as an organism, this interpretation falls within at least some technical senses of the term. I deliberately chose this language to be provocative and to jar the listener into thinking differently about the relationship of evolution to evaluation. As to whether the term is stretching the definitional boundaries beyond propriety, we'll need to see how interpretations of evolution evolve.

To get a feel for the semantic challenges, consider the Wikipedia entry:

> The word "organism" may broadly be defined as an assembly of molecules that influence each other in such a way that they function as a more or less stable whole and have properties of life. However, many sources, lexical and scientific, add conditions that are problematic to defining the word.
>
> The Oxford English Dictionary defines an organism as "[an] individual animal, plant, or single-celled life form"[1] This definition problematically excludes non-animal and plant multi-cellular life forms such as some fungi and protista. Less controversially, perhaps, it excludes viruses and theoretically-possible man-made non-organic life forms.
>
> A polypores mushroom has symbiotic relationship with this Birch Tree
> An ericoid mycorrhizal fungusChambers Online Reference provides a much broader definition: "any living structure, such as a plant, animal, fungus or bacterium, capable of growth and reproduction"[2]. The definition emphasises life; it allows for any life form, organic or otherwise, to be considered an organism. This does encompass all cellular life, as well as possible synthetic life. This definition does lack anything approximating to the word "individual" which would exclude viruses.
>
> The word "organism" usually describes an independent collections of systems (for example circulatory, digestive, or reproductive) themselves collections of organs; these are, in turn, collections of tissues, which are themselves made of cells. The concept of an organism can be challenged on grounds that organisms themselves are never truly independent of an ecosystem; groups or populations of organisms function in an ecosystem in a manner not unlike the function of multicellular tissues in an organism; when organisms enter into strict symbiosis, they are not independent in any sense that could not also be conferred upon an organ or a tissue. Symbiotic plant and algae relationships do consist of radically different DNA structures between contrasting groups of tissues, sufficient to recognize their reproductive independence. However, in a similar way, an organ within an "organism" (say, a stomach) can have an independent and complex interdependent relationship to separate whole organisms, or groups of organisms (a population of viruses, or bacteria), without which the organ's stable function would transform or cease. Other organs within that system (say, the ribcage) might be affected only indirectly by such an arrangement, much the same way species' affect one another indirectly in an ecosystem. Thus, the boundaries of the organism are nearly always disputable, and all living matter exists within larger heterarchical systems of life, made of wide varieties of transient living and dead tissues, and functioning in complex and dynamic relationships to one another.
>
> "Organism." *Wikipedia, The Free Encyclopedia*. 20 Apr 2007, 06:57 UTC. Wikimedia Foundation, Inc. 22 Apr 2007 <http://en.wikipedia.org/w/index.php?title=Organism&oldid=124300874>.

Or, look at the second definition provided in the American Heritage Dictionary (as listed in answers.com):

> A system regarded as analogous in its structure or functions to a living body: the social organism.

The American Heritage® Dictionary of the English Language, Fourth Edition Copyright © 2004, 2000 by Houghton Mifflin Company. Published by Houghton Mifflin Company.

[2] As with "organism", the term "species" is extremely difficult to define (the following segments are taken from the Wikipedia entry).

> It is surprisingly difficult to define the word "species" in a way that applies to all naturally occurring organisms, and the debate among biologists about how to define "species" and how to identify actual species is called the 'species problem'.
>
> Most textbooks define a species as all the individual organisms of a natural population that generally interbreed at maturity in the wild and whose interbreeding produces fertile offspring. Various parts of this definition are there to exclude some unusual or artificial matings:
>
> - Those which occur only in captivity (when the animal's normal mating partners may not be available) or as a result of deliberate human action.
> - Animals which may be physically and physiologically capable of mating but do not normally do so because only their normal mating partners perform the courtship rituals or some other behavior "correctly".
> - Animals whose offspring are normally sterile.
>
> … The typical textbook definition (above) works well for most multi-celled organisms, but there are several types of situations where it breaks down… By definition it applies only to organisms which reproduce sexually… Horizontal gene transfer makes it even more difficult to define the word "species". There is strong evidence of horizontal gene transfer between very dissimilar groups of prokaryotes… All definitions of the word "species" assume that an organism gets all its genes from one or two parents which are very like that organism, but horizontal gene transfer makes that assumption false.
>
> Species. (2007, April 21). In Wikipedia, The Free Encyclopedia. Retrieved 03:07, April 22, 2007, from http://en.wikipedia.org/w/index.php?title=Species&oldid=124569593

But horizontal gene transfer is precisely the issue here. I am arguing that the model of such transfer among bacteria (prokaryotes) is a more apt one for ideas (and for programs) than that of sexual reproduction, which clearly cannot be true. If this supposition is correct, then it may be as accurate to refer to species of programs as it is of bacteria.

[3] I seek to avoid entirely the problem of the similarities and distinctions between evaluation and scientific method for purposes of this argument. I will assume that for Popper and Campbell, evaluation would be considered a type of learning process and, as such, would fall within the same evolutionary framework of any such process.

[4] Toulmin makes the same point: "In talking about the development of natural science as "evolutionary," I have not been employing a mere *façon de parler*, or analogy, or metaphor. The idea that the historical changes by which scientific thought develops frequently follow an "evolutionary" pattern needs to be taken quite seriously; and the implications of such a pattern of change can be, not merely suggestive, but explanatory." (p. 470) (Toulmin, 1967)

[5] For example, Popper points out that all three levels share an inherited structure. At the genetic level it is obvious of course that the inherited structure is the genome. But it may be less obvious that at the behavioral level there is also an inherited structure – "the innate repertoire of the types of behavior which are available to the organism." And, perhaps most intriguingly, the corresponding 'inherited' structure in science consists of the "dominant scientific conjectures or theories" that get passed down through academia and distributed throughout communities of researchers. It is this last sense that comes closest to evaluation.

[6] Campbell describes it concisely with the phrase "Blind Variation and Selective Retention" (Campbell, 1988) (or BVSR) which he says encompasses three essentials: "(a) mechanisms for introducing variation;

(b) consistent selection processes; and (c) mechanisms for preserving and/or propagating the selected variations." (p. 402)

[7] In the biological world, genes are the central mechanism for the transmission of evolutionary knowledge. Variation in genetic instructions, brought about by sexual reproduction and random mutations form the foundation for the diversity and variation that subsequently gets selected for. But this cannot also be the case for programs and their theories, for the obvious reason that there exists no mechanism for sexual reproduction of ideas. The lack of a cognitive equivalent of a gene is not a barrier to an evolutionary theory of evaluation, anymore than it was to the biological one. After all, Darwin advanced his theory of natural selection ignorant of Mendel's work or the idea of genetics. We are still waiting for the equivalent of a Mendel to articulate the evolutionary carrying mechanism for cognitive structures. One promising alternative to the mechanism of mutation and sexual reproduction comes from recent advances in microbiology. We now know that the fundamental division in forms of live on earth is not the one that we formerly assumed between plants and animals. Instead, the fundamental distinction is between prokaryotes or bacteria and eukaryotes, including all plants and animals. Bacteria do not reproduce sexually. Instead, they trade and recombine genes. As Margulis and Sagan point out: "…all the world's bacteria essentially have access to a single gene pool and hence to the adaptive mechanisms of the entire bacterial kingdom. The speed of recombination over that of mutation is superior: it could take eukaryotic organisms a million years to adjust to a change on a worldwide scale that bacteria can accommodate in a few years...Human beings are just learning these techniques in the science of genetic engineering, whereby biochemicals are produced by introducing foreign genes into reproducing cells. But prokaryotes have been using these 'new' techniques for billions of years. The result is a planet made fertile and inhabitable for larger forms of life by a communicating and cooperating world-wide superorganism of bacteria" (pps. 30-31)(Margulis & Sagan, 1986). This seems like a more apt description of what happens with the evolution of thought and of theories – rapid recombination and adaptation of ideas.

[8] Popper argues that all three levels of evolution – genetic adaptation, adaptive behavioural learning, and scientific discovery – are subject to selection of the various mutations and variations. He doesn't so much support the stereotypical notion of the "survival of the fittest" as he does the idea of elimination of error – those which are badly adapted are eliminated – or what he dubs the "method of trial and error" or better, the "method of trial and the elimination of error." The selection process is what we mean by natural selection (although I will say more about the distinction of this from 'artificial selection' later). Essentially selection for Popper is a negative feedback function, a feedback function that provides negative information about the viability of that variation. Popper was famous for advancing this falsificationist view of selection that argued that hypotheses or conjectures survive until falsified by empirical evidence.

The idea then is that all three levels – genetic, behavioral and scientific – are subject to the exact same process of evolution and that this process encompasses three central components: an inherited structure, variations in the instruction of the inherited structure that are then subjected to pressure, and selection of some of these variations by elimination of those that are nonviable. Campbell summarizes the entire process as Blind Variation and Selective Retention or BVSR for short.

Campbell translated this impulse into the magnificent edifice of validity theory in social research. In his view, any research project consists of a series of subpropositions – that a program or measures reflect their purported constructs (construct validity); that we identify relationships between programs and outcomes (conclusion validity); that some of these relationships may demonstrate that what we did caused what we observed to happen (internal validity); and, that our conclusions might generalize to other persons, settings or times (external validity). For each such proposition one might construct any number of alternative explanations. The task is to rule out the more plausible of these alternative explanations, leaving the inference of interest as the best surviving explanation.

Toulmin takes a similar approach to the selection process in science in his paper on The Evolutionary Development of Natural Science (Toulmin, 1967) when he remarks: "Science develops (we have said) as the outcome of a double process: at each stage, a pool of competing intellectual variants is in circulation, and in each generation a selection process is going on, by which certain of these variants are accepted and incorporated into the science concerned, to be passed on to the next generation of workers as integral elements of the tradition." (p. 465)

References

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of karl popper*. LaSalle, IL: Open Court Publishing Co.

Campbell, D. T. (1988). Evolutionary epistemology. In E. S. Overman (Ed.), *Methodology and epistemology for social science: Selected papers of donald t. Campbell*. Chicago: University of Chicago Press.

Cundiff, D. K. (2007). Evidence-based medicine and the cochrane collaboration on trial. *Medscape General Medicine, 9*(2), 56.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

House, E. R. (November, 2006). *Blowback: Consequences of evaluation for evaluation.* Paper presented at the Annual Conference of the American Evaluation Association, Portland, OR.

House, E. R. (October, 2004). *Democracy and evaluation.* Paper presented at the Biannual Meeting of the European Evaluation Society, Berlin.

Margulis, L., & Sagan, D. (1986). *Microcosmos: Four billion years of microbial evolution*. Berkeley, CA.: University of California Press.

Mayo Clinic. (2007). Clinical trials: A chance to try evolving therapies. from http://www.mayoclinic.com/health/clinical-trials/DI00033. Accessed June 26, 2007

Ontogeny. (2007, April 15). (2007, April 15). In *Wikipedia, the free encyclopedia. Retrieved 17:01, april 21, 2007, from http://en.Wikipedia.Org/w/index.Php?Title=ontogeny&oldid=122978259*.

Pollan, M. (2001). *The botany of desire: A plant's-eye view of the world*. New York: Random House.

Popper, K. (1973). Evolutionary epistemology, *Sections I-VI of "The Rationality of Scientific Revolutions" given at the Herbert Spencer Lecture at the University of Oxford*. Oxford.

Popper, K. (1975). Chapter 6 (evolutionary epistemology, 1973). In R. Harre (Ed.), *Problems of scientific revolution*. Oxford: Oxford University press.

Popper, K. (1985). Evolutionary epistemology. In D. M. Miller (Ed.), *Popper selections* (pp. 78-86). Princeton, NJ: Princeton University Press.

Scriven, M. (1967). The methodology of evaluation. In R. Tyler, R. Gagne & M. Scriven (Eds.), *Perspectives on curriculum evaluation, aera monograph series on curriculum evaluation* (Vol. 1, pp. 38-83). Skokie, IL: Rand McNally.

Toulmin, S. (1967). The evolutionary development of natural science. *American Scientist, 55*(4), 456-471.

Wilson, D. S. (2007). *Evolution for everyone: How darwin's theory can change the way we think about our lives*. New York: Delacorte.