Journal of MultiDisciplinary Evaluation
Volume 15, Issue 33, 2019

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Donald T. Campbell's Evolutionary Perspective and its Implications for Evaluation

Robert Picciotto
University of Auckland

Background: Donald T. Campbell, a scientist, humanist, and generalist, left an indelible mark on the evaluation discipline through his methodological work. He is less well known within the evaluation community for his landmark contributions to biology and philosophy. Yet, the evolutionary epistemology that he pioneered has significant implications for evaluation.

Purpose: This article examines the relevance of Donald T. Campbell's blind variation and natural selection approach to evaluation theory, including an elucidation of its basic logic, its social remit as a discipline and trans-discipline, and its summative and formative functions. It also sketches in broad strokes the implications of evolutionary thinking for evaluation practice, including natural and artificial selection, ontogeny, phylogeny, co-evolution, and feedback. Finally, it comments on Campbell's Experimenting Society vision and the ongoing craze for randomised evaluations in development through an evolutionary lens.

Setting: Not applicable.

Intervention: Not applicable.

Research Design: Not applicable.

Data Collection and Analysis: Not applicable.

Findings: Not applicable.

Keywords: *blind variation; evolution; experiment; epistemology; selective retention; society.*

Introduction

Donald T. Campbell's prominent place in the pantheon of the evaluation profession is secure. Within the evaluation community, he is best known for his exploration of evaluation biases, validity concepts, and experimental methods (Shadish and Cook, 1998). His utopian vision of an Experimenting Society (Campbell, 1971) adds to his notoriety and further underlines his life-long commitment to rigorous evaluation.

Thus, perceptions are widespread that Campbell's intellectual fame rests mostly on the design and use improved social research tools. This explains why Alkin and Christie (2004) positioned Campbell squarely on the methods branch of their evaluation theory tree. Yet, as stressed by Shadish and Luellen (2004), Campbell's evaluation legacy cannot be reduced to his (admittedly pioneering) methodological achievements. To the contrary, his polymathic life work cannot be pigeon holed since his writings consistently link methods, values, and use.

It would be foolhardy to attempt a comprehensive catalogue or a critical appraisal of Campbell's wide-ranging contributions to social research and evaluation within the scope of a journal article. In what follows, I only mean to show that Campbell's epistemological thinking provides a coherent framework for appreciating evaluation's role in society and for identifying new avenues of evaluative inquiry. I also hope to stimulate interest in evolution theory among evaluation practitioners.

The article is structured as follows: First, I introduce Campbell as an inveterate trespasser across social science disciplines; Second, I outline the strands of his evolutionary epistemology; Third, I apply his evolutionary concepts to evaluation theory ("who we are"); Fourth, I do the same for evaluation practice; Fifth and finally, I look at the recent emergence of experimentalism in international development through an evolutionary lens.

The Accidental Evaluator

Campbell joined the program evaluation community fortuitously following the imprimatur of his experimentalist stance by Edward Suchman (1967), a notorious advocate of social research conducted in 'scientific' ways. Suchman's influential sponsorship helps explain why Campbell has long been perceived as an unequivocal champion of quantitative methods. Yet, his advocacy of social research experimentalism was always restrained and nuanced.

Campbell never forsook the ambition to get evaluation practice aligned with scientific research strictures. But he fully recognized that the unruly vagaries of human interaction limit the utility of randomized social experiments. Accordingly, he helped to stock the evaluator's toolkit with a variety of quasi-experimental instruments (Trochim, 1998). He also took pains to advocate the pursuit of convergent validity through the triangulation of multiple methodologies.

In his own words, "the polarity of quantitative vs. qualitative approaches to research on social action remains unresolved, if resolution were to mean a predominant justification of one over the other. Each pole is at its best in its criticisms of the other, not in invulnerability of its own claims to descriptive knowledge...If we are to be truly scientific we must re-establish the qualitative grounding of the quantitative." (Campbell and Stanley, 1974, pp. 29-30).

No wonder then that, according to remarks made at an evaluation conference in 1992 –as reported by Chelimsky and Shadish (1997) – Campbell "consistently tried to integrate opposite poles in evaluation thinking, to bridge the gaps between them. This effort of his was crucial to evaluation survival … when it was under attack". From this perspective, Campbell's distinctive contribution to evaluation has as much to do with evaluation theory ('who we are') as with instrumental evaluation practices. Arguably, the immense debt evaluators owe Campbell is emblematic of a bold, comprehensive, and multi-disciplinary approach.

As a research psychologist (he secured his PhD in psychology at Berkeley in 1947), Campbell's early work anticipated recent

developments in behavioural psychology. He probed the conflicting misconceptions that underlie group biases (ethnic, racial, national, etc.) and he proposed ways to help overcome interviewers' and questionnaire biases (Campbell, 1950 and Campbell and Mohr, 1950). Next, he put his undergraduate exposure to biology, genetics, anthropology, and sociology to work. His first faculty positions were at Ohio State University (1947-1950) and the University of Chicago (1950-1953) but it is at Northwestern University which had long promoted 'no holds barred' intellectual debate and fulsome interaction across disciplines that he found his intellectual home.

This is where he spent 26 years, trained several generations of doctoral students, and made his mark in the field of social psychology through teaching and research. This is also where he initiated his fruitful inquiry about how knowledge is acquired, recognized, used, and transmitted, starting with the study of how false knowledge can be identified and set aside. Towards the discovery of 'a science of science', he undertook a bewildering range of scholarly endeavours so that when he took up his last academic post, at Lehigh University in 1982, he was designated "university professor" with faculty listings in four departments: psychology, sociology, anthropology, and education.

Evolutionary Epistemology

Just as Quine (1969), Campbell stressed the futility of seeking an understanding of knowledge 'in itself', i.e., to philosophize about science without resort to science. Instead, he favoured the study of how knowledge is processed using the natural sciences, including biology as well as the social sciences starting with psychology. For Campbell (1988) all knowledge is fallible and partial, but the external world exists independently of the mind, does not need to be proven, and provides the ultimate benchmark against which theories are falsified or temporarily retained.

Thus, inspired by Karl Popper, Campbell (1974) coined the 'evolutionary epistemology' term to relate the processes of knowledge acquisition and growth to theories of biological evolution (especially of Darwinian evolution through natural selection). As he embarked on his intellectual journey at the frontiers of social research, Campbell (1990) applied "selection theory to trial and error learning… to visual perception… and to creative thought". It is primarily through this lens that I propose to examine Campbell's enduring impact on the evaluation discipline.

First, Campbell posited that cognitive mechanisms in animals and humans are products of progressive biological adaptation. Thus, sensory mechanisms, the nervous system, the brain, etc. had gradually evolved to facilitate a proper fit with the external environment and ensure species survival. The acquisition of hearing and/or touch helped to detect obstacles and to improve trial and error movements. Vision eliminated the need for blind trial and error locomotion. Memory improved on the strict randomness and inefficiency of food search. Through the combination of sight and memory, beneficial survival habits emerged and eventually turned into instincts.

Next, the advent of visually supported thought facilitated problem solving towards improved probabilities of survival. Through socially vicarious mechanisms (i.e. observational learning and imitation), trial and error exploration by one member of the group benefited the entire group. Language (created and acquired through trial and error just as the other attributes) facilitated vicarious learning thus enhancing the chances of group survival. Finally, human knowledge was acquired through testing of expectations, hypotheses, or theories using the same variation and selective retention mechanisms that governed the prior phases of biological evolution.

Peter Munz (2001), another epistemologist, went further in articulating Campbell's selective adaptation vision by equating all organisms to embodied theories about their environment and by viewing all theories as disembodied organisms. Conversely, he conceived of the environment as a selection process operating through elimination of the "unfit". Defining any organism as a provisionally true hypothesis about its operating environment that survives until it is falsified, he envisaged environmental studies

focused on the examination of surviving organisms.

Thus, just as Munz, Campbell viewed biological evolution as a knowledge process. Hence, he applied the natural selection paradigm to learning, thought, and science. From this vantage point, he observed that as cultures and technologies evolve, weak theories tend to be discarded through falsification. Thus, scientific progress occurs as new and more promising conjectures are put forward, i.e., a theory prevails until it is replaced: the analogy with biological evolution holds. Furthermore, just as selection explains evolution, the selection process itself complies with a selection process: it is a multiple loop learning mechanism.

In the same vein, Stephen Toulmin (1972), related knowledge creation to natural selection: ideas and concepts are introduced into the gene pool of scientific knowledge through journals, conferences, books etc. Assuming a reasonably free market for ideas, competition among scholars winnows out the field and only the more promising theories survive. Thus, in the knowledge world, trading and recombination of ideas mimic the biological evolution process.

This assumes of course that the relative values of competing theories can be established and, especially in the social sphere, that political forces or cultural biases do not intervene to distort the outcome. They often do, of course, so that progress is not foreordained and when it takes place it is by fits and starts. This obstacle to social learning was one of Campbell's persistent preoccupations as he aspired to turn open societies into experimenting societies through greater resort to eliminative induction, shared inquiry and the systematic use of trial and error mechanisms.

Implications for Evaluation Theory

Popper (1974) used the following metaphor to illustrate Campbell's blind variation/selective retention model of knowledge creation: "… although the blind man who searches for a black hat may bring some order into his trials, the order is not given to him; he may choose or invent one order (method) first, and a different order later; and these choices will be trials too – even though on a higher level…the trials are forays into the unknown… they are blind; while to the degree that past knowledge enters, their blindness is only relative… We may at the beginning of an exploration be blinder than we are after even a short time, though after even a short time we may still be blind: we may still not know where the black hat is, but we may know (or think that we know) where it is not"

Evidently, the blind man could do with some help – just as the hard-pressed policy maker in pursuit of effective solutions to a social problem does. In the social world, even more than in the natural world, uncertainty prevails, and prediction is foolhardy. Policy choices are often shots in the dark. Program design faces a wide range of options. Implementation brings forth new challenges and calls for frequent re-appraisals. Typically, responsible decision makers draw on a store of knowledge and/or an inventory of successes and failures to help chart their course, but this does not eliminate the possibility of error.

Against this background, this section of the paper explains how Campbell's variation and selection concepts help explain "who we are"; how evaluation builds bridges between Popper's three Worlds; the rationale of evaluation as a discipline (and a trans-discipline); the imperative of objectivity in evaluation; and the relevance of its overarching criteria to effective variation and selection.

Blind Variation and Selective Retention

In the natural world, biological traits that are maladaptive are shed while those that are adaptive are retained for an organism to survive and have healthy offspring. In the scientific and social world too, elimination of error facilitates judicious choices, and lays the groundwork for further knowledge advances. This is how social learning takes place and where cultural traditions find their justification. Thus, trial-and-error dominates progress in science and should do so as well in the policy world.

This said, blind variation differs from random variation. In biology, variation is random given genetic mutations and re-combinations. In science and the social

domain, progress is made through systematic inquiry. If an experiment supports an intervention design, it weakens alternative designs. Conversely, a scientific or a program theory is apt to be chosen among many options if it best describes a phenomenon or appears promising. Over time, any theory acquires or loses credibility depending on the outcome of evaluative tests and on other inevitable changes in the body of knowledge surrounding it.

Variation matters since only with diversity is knowledge likely to make competitive selection effective in setting aside unpromising designs and enhancing the survival prospects of selected interventions. Choice among policy options relies on prior experience and contextual (often vicarious) knowledge and social interventions are only undertaken if they promise to deliver results. This, as I will now try to demonstrate, is where evaluation (whether formal or informal, independent or self-administered) comes in. It distinguishes selective retention in the social knowledge world from natural selection in biology.

Bridging the Natural and Human Worlds

Karl Popper (1978) paints the following picture of our world:

> There is the physical universe, World 1, with its most important sub-universe, that of the living organisms. World 2, the world of conscious experience, emerges as an evolutionary product from the world of organisms. World 3, the world of the products of the human mind, emerges as an evolutionary product from World 2. In each of these cases, the emerging product has a tremendous feedback effect (pp. 166-167).

From this vantage point, and as part of his fulsome endorsement of Campbell's evolutionary epistemology, Popper (1974, p. 1060) stressed a singular characteristic of human knowledge: while animals as well as human beings focus on the physical world (World 1) and use their senses and their brains (World 2), only humans have evolved the capacity to create the knowledge products of World 3.

According to Michael Scriven (2016), evaluation, a cognitive skill has an ancient pedigree that "pre-dates the emergence of true language" and that endowed humans with an enormous evolutionary advantage when "interwoven with demonstration, trial and differential reinforcement". The Scriven conjecture asserts that the human species resorted to evaluation since its origins.

Irrespective of this proposition, it is self-evident, as noted by Carol Weiss (1998, p. 3), that "people evaluate all the time". But human bias is ubiquitous and mistakes in human judgment are the rule rather than the exception. This is amply confirmed by recent advances in behavioural psychology, the "science of error" pioneered by Nobel laureates Daniel Kahneman (2012) and Richard Thaler (2015). Pursuit of objectivity (Campbell's lifelong crusade) is the fundamental rationale of evaluation as a discipline.

Evaluation as a Discipline...and a Trans-Discipline

Popper's World 2 is the domain of subjective knowledge and experience whereas World 3 is the realm of objective and collective (albeit transitional) knowledge created by the human mind. The evaluation discipline is committed to fill gaps between World 2 and World 3. It is rooted in the proposition that intuition is fallible and needs control through systematic review, argumentation, and criticism.

Specifically, evaluation as a discipline is mandated to improve the quality of information used for human decisions: evaluation reduces the scope of error and bias and adds a value dimension that other disciplines often lack. Evaluators are deeply invested in World 1 whereas social researchers often cut themselves from it within the ivory towers of World 3. The legitimation of claims implicit in any social intervention is at the centre of the evaluation discipline – as well as action research, its close cousin. It follows that meta-evaluation – the evaluation of evaluation – is a professional obligation of evaluators, as stressed by Stufflebeam (2001).

To be sure, all knowledge occupations play a role in linking World 1 and World 3. But the evaluation discipline is uniquely distinctive because: (i) it uses all the disciplines relevant

to the problem at hand; and (ii) it fulfils a unique trans-disciplinary function by providing other disciplines with tools that make them more effective, e.g. by "playing the role of arbitrator, scapegoat, trouble shooter, inventor, conscience, jury, or attorney" (Scriven, 1991, pp. 363-364).

As a result, evaluation generates a wide range of feedback loops among all three worlds. While it is governed by mental states and processes (World 2), it draws on the products of the human mind (World 3) to sharpen its assessment of the material environment (World 1). It also serves World 3 by testing the interface between World 1 and 3. Finally, at its cumulative best, it generates products for World 3.

The Imperative of Objectivity

Natural selection reflects different survival rates of organisms bearing adaptive genes whereas theory selection in science and choice of intervention in the social world are or should be performed with the help of specialists using legitimate, time tested protocols that identify designs that are close "fits" within a given context. Thus, only a reliable mapping of the lay of the land and an accurate description of the missing headgear will help the blind man towards the right hat.

It is the validity imperative in all its dimensions that underlies Campbell's passion for research design on which he arguably "had a more profound effect in the second half of the twentieth century than any other single person" (Trochim, 1998, p. 407). For Campbell, objectivity does not imply analytic certainty, but it does justify the claim that a corrigible and fallible knowledge proposition is plausible.

Specifically, validity implies that the theory of action selected for an intervention will survive construct validity tests that ascertain inter alia whether the samples of persons, settings, manipulations, and measures are adequate. Validity also means that evaluation findings are consistent with a theory of change that transparently links outputs and outcomes to impacts (conclusion validity).

Furthermore, observed outcomes should be attributable to the intervention (internal validity). Finally, external validity is the degree to which a causal relationship found in a study generalizes across various persons, settings, treatments, measures as well as cultures (Kirkhart, 2010).

The Evolutionary Logic of Evaluation

It should be clear by now that the evaluation function complies with the basic tenets of the evolutionary process evoked by Campbell. But is its basic logic consistent with evolutionary principles? According to Michael Scriven (1991, p. 139), evaluation is "the process of determining the merit, worth or value of something or the product of that process". Putting this definition to work readily demonstrates that the artificial selection process facilitated by evaluation is precisely designed to ascertain whether a social intervention organism fits its operating and authorizing environment.

Just as every relationship between an organism and its environment can be conceived a knowledge relation, every project, program, or policy (i.e. any social intervention) can be perceived as a theory that illuminates the intersection between the intervention and society. Thus, evaluation may either refute the theory, attest to its provisional fitness, or evoke the results of prior experiments to propose ways of tightening the fit between the intervention and its social environment.

Evaluation bridges World 1 and World 3 by addressing the merit, worth and value trilogy. But fitness measured with one criterion does not guarantee fitness in terms of the other two. Merit is intrinsic to the intervention. Assessing merit means testing the theory embedded in the intervention conceived as an organism. Is the intervention designed to satisfy its own survival needs? Even if it is, i.e. even if the intervention meets its intended objectives, worth is not guaranteed.

Worth is extrinsic to the intervention. It addresses another question: is the intervention equipped to survive external threats and to meet group expectations and needs? It is a highly pertinent question in evolutionary terms: an intervention that survives a merit oriented summative evaluation is still vulnerable. Just as a natural organism is vulnerable to predatory threats and ostracism from other group members, it is

not enough to satisfy the owners, managers and sponsors of an intervention and meet their short-term objectives.

Specifically, inability to meet the demands of intervention stakeholders, especially those (funders, owners, voters, opinion makers, etc.) that constitute its authorizing environment can doom a social intervention (Moore and Khagram, 2004). Equally, biological survival relies on mutual aid and group support: inadequate group cohesion and cooperation can lead to species extinction. Mutual aid as well as competition propels evolution. Wolves hunt in packs. Birds collectively search for food, Bees and ants work together (Kropotkin, 1989).

Finally, even if an intervention is meritorious and worthy, it may not be significant or valuable, i.e. it may not serve the public interest, the ultimate survival test in a democracy. This may occur when overall societal priorities no longer justify the intervention design. Equally, in the natural world, major changes in the biosphere yield intense episodes of species extinction (Raup, 1994).

In sum, in the social world just as in the natural world, fitness is a relative and contingent concept. No single organism or species can be expected to acquire all the attributes necessary for survival across all possible environmental conditions. Specialization and adaptation to an environmental niche are imperative. Equally, social interventions must be designed to fit within a specific operational and political context and more than one intervention will normally be needed to achieve a pertinent societal goal. Inevitably, trade-offs must be struck since social needs are diverse, resources are limited, and stakeholders' interests vary. The next section seeks to demonstrate that the necessary trade-offs are better struck with evolutionary evaluation than without.

Implications for Evaluation Practice

This section of the paper shifts the focus to the evaluand. After examining summative and formative evaluation through an evolutionary lens, it describes the evaluand's life cycle. Next, it shows how co-evolution and feedback ideas as well as natural/artificial selection processes help explain the emergence of social programs as distinct from social interventions. Finally, it explores the implications of evolutionary thinking for evaluation at this higher plane.

Summative and Formative Evaluation

Evolutionary thinking implies a recognition that the social environment is made up of complex systems that are continually changing in mostly unpredictable ways; that social interventions are always experiments, and that their adaptation to evolving conditions is a condition of their sustainability. Hence, social learning depends on frequent evaluative feedback informed by systems thinking.

As part of the blind variation/selective retention model of decision making, summative evaluation plays a central role. To use Popper's terminology, it aims to refute the conjecture embedded in the evaluand. It may do so at any stage of the intervention life cycle. This conforms to the logic of natural selection: eliminating programs that do not work opens the way for programs that do.

On the other hand, even if the theory or the intervention has survived rigorous tests, there is no guarantee that the problem which the intervention has addressed is significant; that the intervention is the most relevant, that it provides an efficient solution to the problem at hand, or that its design will remain serviceable in other contexts or in the future. Similarly, in the natural world not all species survive the multiplicity of threats of diverse and volatile environments.

In other words, empirical testing (does the intervention work?) is necessary for sound decision making …but it is far from sufficient. Even if the conjecture embedded in an intervention is not refuted based on valid tests, summative evaluation cannot vouch for the optimality of the solution. One may help the blind man find a hat and confirm that it is indeed black…but the hat may not fit the blind man's head, or it may look funny.

Helping to find a better hat (how to improve the intervention) is the province of formative evaluation. It is more speculative and open ended than summative evaluation. It helps

decision makers grope towards better targeted, more effective social interventions. It draws on dominant social science theories and validated meta-evaluation findings, the equivalent of inherited structures in biology. Its recommendations are only conjectures: the past is not necessarily prologue.

Thus, evaluation only makes the selective retention process somewhat more efficient and effective in helping decision makers decide whether the intervention is fit for purpose, and/or whether the prospects for an even closer fit are promising enough to be tried out. While evaluation on its own cannot be used to justify a social intervention, it can examine it, interpret it, confront it, and criticize it to help shape judicious decisions regarding its design, adaptation, and termination.

Evaluation and the Life Cycle

All social interventions have a finite life – just as biological organisms. They evolve through different phases of a life cycle. Ontogeny in evolution theory describes how an organism is born, grows and dies. To thrive within a given environment for the duration of its existence, the organism draws on innate skills embedded in its genetic material as well as on behavioural adaptation secured through trial and error learning.

Any social intervention develops in stages. The evolutionary perspective encourages adaptation of evaluation methods to various stages of the intervention life cycle. (Urban, Hargraves and Trochim, 2014). During the first seminal phase (identification and preparation), evaluation offers relevant social science findings to shape program design and the design of a theory of action that governs behaviour through protocols that are akin to the content of a genetic mix. Due diligence practices follow to help ensure that the intervention design is fit for purpose. Here again evaluation comes into play through small-scale experimental trials followed by upscaling tests.

Next, ex-ante evaluation and evaluability assessments help decide whether the intervention is worthy of full-scale implementation. This is when success indicators are identified, and monitoring systems are constructed. They mimic the feedback mechanisms that help living organisms avert danger, locate food, find shelter, and survive. They also help the intervention adapt through a fine-tuning phase when new capabilities are added, skills are acquired, routines are firmed up, experience is gained, etc.

This maturation phase is followed by a full-scale implementation phase which unfolds until the intervention objectives are met or allocated resources run out. Before they do, retrospective evaluation is carried out to examine overall performance, draw lessons of experience, and where appropriate lay the groundwork for a successor intervention, thus completing the life cycle. Passing the evolutionary torch to the next generation is key to long-term sustainability, in the social world just as in the natural world.

Feedback and Co-Evolution

Most programs are the offspring of prior programs. Without the right kind of feedback, programs fail, as suggested by evolutionary theory. Different kinds of feedback can have felicitous or disastrous consequences. This is because all organisms are part of the environment to which they must adapt and as they do they may change the environment in ways that may threaten their own survival, a distinctive feature of Darwinian selection.

Biological feedback is a two-way street (Robertson, 1991). Similarly, social programs are intentionally designed to modify the social world, but they often do so without adequate consideration of their unintended consequences, e.g. depletion of environmental resources. In such situations, sustainability requires the inclusion of projects in existing or new programs to mitigate or reverse these deleterious effects. Climate change comes to mind.

Co-evolution theory examines how two or more organisms evolve together and interact symbiotically. Mutualistic relationships generate collaborative benefits while parasitic and predatory relationships are one-sided and fail the test of Pareto optimality. It follows that projects included in program portfolios should be subjected to prior co-evolutionary selective retention tests.

Equally, the evaluation of social programs that rely on partnerships to achieve collective impact call for performance assessment of individual partners to avoid free riding and provide incentives for partners to fulfil their distinctive accountabilities and respect their reciprocal obligations.

Natural and Artificial Selection

Natural selection explains the wide diversity of species currently in existence. All living things are connected to a common ancestor through a huge tree of life. All species are embedded in nested hierarchies. As natural events disperse organisms of the same species and erect barriers between them, organisms must struggle to fit different environments and cannot interbreed. As a result, over many generations, new evolutionary branches sprout as the organisms adapt to their distinctive ecological niches and accumulate different heritable capabilities through natural selection.

Equally, in the social world, different program designs evolve to address different but interconnected social issues and/or respond to diverse stakeholders' aspirations. Several projects are usually needed to address a significant social problem. Just as breeders select individual plants or animals with desirable characteristics to procreate and generate the desired breed, policy makers resort to programs by selecting and combining several interventions to achieve an overarching social goal.

Qualitative comparative analysis using Bayesian logic, long applied to the study of biological structures, is increasingly used to evaluate public policies and programs in complex environments (Matthews, 2017). Similarly, different types of diagrams display logic models in program theory. They show how the various project mechanisms fit together to make the program work. They evoke the "phylogenetic trees" drawn by evolutionary biologists to show the inferred relationships among various biological species based on similarities and differences in their physical or genetic characteristics.

Evaluation of Project Portfolios (Programs)

It should be clear by now why variation of organisms within a species is essential to survival given the diversity of threats that it faces and the different environmental configurations to which it must adapt. Similarly, in the social policy world, resilience to risk is enhanced by diversity in program content. Evaluation contributes to the variation and contributes to the fitness of program design through situation analyses and evaluability assessments.

Phylogeny is the scientific study of how species (collections of organisms) evolve. It also throws light on the genesis of social programs made up of a portfolio of project interventions. For example, evaluation helps to construct programs through a process akin to selective breeding: a program is typically made up of several projects designed and chosen to tap synergies and/or adapt program characteristics to local conditions (Trochim, 2007). Programs come into being through blind variation and selective retention, just as individual interventions.

Evaluation of project portfolios starts with an aggregation of separate project level evaluations used as building blocks for the overall program evaluation edifice. But programs, just as individual projects, are knowledge organisms so that their evaluation also calls for a generic theory of action and change that takes explicit account of project interlinkages and interactions, i.e., their alignment and coherence towards desired program goals. How do evolutionary processes at the intervention level affect changes at the program level? This is best explained by a hierarchical cycle of positive and negative feedback, resource allocations imposed by program management and artificial selection/retention of projects to achieve synergies (Van de Ven and Grazman, 1999).

Similarly, evaluation networks and practices can be conceived as a complex knowledge organism made up of diverse evaluation approaches and methods that co-exist and interact. Arguably, their co-evolution reflects the same blind variation and selective retention mechanisms that govern other knowledge occupations. Not all monitoring

and evaluation models survive, e.g. adversary evaluation and management by objectives are no longer around. What then lies ahead for evaluation? Evidently, it is now time to turn to Campbell's utopian perspective on the future of evaluation.

Whither Experimentalism?

In this section, I outline Campbell's Experimenting Society; acknowledge Campbell's deep ambivalence about it; identify the antecedents of an on-going international development experimentalist initiative; show how its advent fits within the evolutionary path of the discipline and offer an admittedly tentative, fallible, and corrigible assessment of its record so far.

What is the Experimenting Society?

According to Carol Weiss (2000), the Experimenting Society reflects Campbell's interest in strengthening the role of reason and logic in politics. But as a firm adherent of Popper's Open Society ideas, Campbell believed in participatory democracy and did not advocate technocracy nor an elite status for evaluators. Nor did he condone evaluators' pretensions to shape policy, except indirectly through the assessment of current and past policies. He restricted the remit of evaluation to ascertaining whether programs "work" – as only one element of public deliberation and multiparty negotiations.

The ideology of the Experimenting Society is admirable. Who can object to "a popularly responsive society whose goals and means are determined by collective good and popular preference…an honest society committed to reality testing, to self-criticism, to avoiding self-deception", or to "an accountable, challengeable due process society … decentralized in all feasible aspects … (with) means idealism as well as ends idealism" (Kelly & Gregware, 1998, p. 190)?

In this connection Campbell visualized two sets of problems. The first had to do with "the widespread resistance of institutions and administrators to have their programs evaluated… as we try to implement high quality program evaluations we meet with continual frustration from the political system. It seems at times to set up just to prevent reality testing" (Campbell, 1988, p. 36). To address it, he called on evaluators "to make a career commitment…to contribute to the best possible exploration – in advance of what such (an experimenting) society would be like" (Campbell, 1988, 37) …and "to invent political/organizational alternatives that would avoid the problem" (Campbell, 1988, p. 47)

This manoeuvre allowed him to avoid tackling the problem directly and to side step the dilemmas associated with multi-objectives, ambiguous, evolving, adaptable social programs, let alone the diverse, conflicting, and sometimes self-serving and covert goals of sponsors and/or funders. Nor did Campbell delve into how evaluation should relate to organizations or address how evaluators might protect their independence in a market driven occupation. Instead, he concentrated on a second set of problems having to do with evaluation methods and processes.

He presumably did so for pragmatic reasons: these are matters that the evaluation community has the wherewithal to address. Thus, he put forward a wide range of sensible albeit modest proposals to facilitate evaluation use – emphasizing evaluation of successful programs; using volunteers in randomized experiments; complying with strict transparency rules regarding funding and data collection; eschewing the use of simplistic indicators in program management; respecting the people involved in and benefiting from programs; etc.

Throughout his exploration of evaluation methods and practices, he was explicit regarding the essential criteria of validity, interpretability, and reliability and he never questioned the primacy of experimentalism. In fact, he made clear that his central concern was "to extend the epistemology of the experimental method into nonlaboratory social science… (and) to treat "the ameliorative effects of government as field experiments (Campbell, op, cit., 1998, p.36).

Campbell's Ambivalence

Revealingly, Campbell withheld the publication of "The Experimenting Society"

that he had crafted in 1971 until the compilation of a selection of his papers in Overman (1988). His unease about prescribing a utopian vision for evaluation in society is unsurprising: he considered any idea of clairvoyance to be fatuous (Campbell, 1966) and he disdained social scientists who are trapped by their theories, do not subject them to reality testing and feel free to "say what is to be done… rather (than) to say what has been done" (Campbell, 1988, pp. 44-45).

Unsurprisingly, Campbell faced vigorous opposition when he presented his ideas to various disputatious communities of "truth seekers". They pointed to several practical, political, theoretical, and ethical issues that Campbell had not addressed convincingly (Shaver & Staines, 1971). This helps explain why Campbell conceived of the Experimenting Society as only one among many alternative utopias. He acknowledged that once the full implications of information overload implicit in the proposed Experimenting Society, "this monster of measurement", might well be rejected out of hand.

Experimentalism in Development: The Antecedents

Campbell's Experimenting Society could not have been conceived today. It was created at a time of optimism about public affairs. Published a year before the Berlin wall crumbled, it made assumptions about politics and the society that are no longer valid (Kelly & Gregware, 1998). Specifically, it was shaped by the confluence of two waves of evaluation diffusion (Vedung, 2010). First, the experimental wave of the 1950s and 1960s, a radically rationalist, positivist, and meritocratic phase of evaluation history grounded in quantitative methods that he was instrumental in triggering. Next, the dialogic wave which swelled in the late 1960's and prevailed until the mid-1980s: the heyday of qualitative methods and participatory processes geared to social learning and community empowerment.

But by the time the Experimenting Society paper was published a neo-liberal wave had engulfed the evaluation discipline. Suddenly government was perceived as the problem rather than the solution. This is when the new public management movement, emphasizing devolution, decentralization, and deregulation, introduced market thinking into government. The next wave was evidence-based and it too was powerfully influenced by Campbell's ideas. Its prime movers sought a sharper edge in evaluation methods that would produce incontrovertible evidence of verifiable "results" as justification for public action.

The International Development Experimentalist Initiative

The onslaught on qualitative development evaluation occurred at the turn of the century. These results differed from those of a cottage industry of policy research studies that could not identify robust correlations between aid volumes and economic growth. Their ambiguous findings, due to the methodological limitations of cross-country correlations, amplified development aid pessimism within an environment of budget austerity in donor countries.

This new context created a strategic opportunity for young economists based at the Massachusetts Institute of Technology (MIT). Its Poverty Action Lab (J-PAL) charismatic co-founder famously declared during a World Bank Conference in 2003: "Just as randomised evaluations revolutionized medicine in the 20th century, they have the potential to revolutionize social policy during the 21st" (Duflo and Kremer, 2005).

Is Randomization the Answer?

In the right circumstances experimental methods do establish causality by providing a valid measure of the counterfactual. They successfully address the issue of selection bias and allow evaluators to establish a measure of statistical significance to evaluation findings. These are formidable advantages. On the other hand, randomization is mostly suited to simple, tunnel-type interventions with easily identified participants and non-participants, where spillover effects are not likely to bias the results. Yet, this is where knowledge gaps are the deepest.

Randomized controlled trials (RCTs) are redundant when no other plausible explanation for the results observed is available. They are not always feasible, for example when no untreated target group can be identified or when an intervention or policy is intended to be universal. They may not even be decisive in establishing attribution, for example where latent and unobserved causal factors affect the experiment. They are utterly unsuitable for the evaluation of complicated or complex programs in unstable environments that characterize most development situations.

Nor is external validity, a pre-requisite of development policy relevance, their forte. The notion of a universal data base that would provide valid answers to what works and does not work in international development was termed 'crazy' by a former World Bank Chief Economist (Bedecarrats, Guerin & Roubaud, 2017, p.11) and it would probably have been dismissed out of hand by Campbell as an example of the misplaced assurance of scholars "quite out keeping with the scientific status of their fields" (Campbell, 1998, p. 45).

RCTs raise complex ethical issues, involve high costs, require superior skills, large samples, and specialized quality assurance arrangements. They inhibit resort to cheaper and more effective evaluations. They also hinder fulsome participation of beneficiaries in the evaluation process. Finally, they privilege the selection of simplistic programs and projects that may not be fit for purpose. Most high-level policies, programs and projects that are now privileged by international development agencies are not evaluable through randomized treatment.

These propositions are amply confirmed by the travails of four RCT driven studies carried out in Mexico, Cambodia, and Kenya (Bedecarrats, Guerin & Roubaud, 2017, pp.11-16). There are better alternatives. Mixed methods guided by theories of change and backed up as appropriate by quasi-experimental methods pioneered by Campbell are better equipped to assess goal relevance and determine the reasons for success or failure of achieving intended effects (and the extent and nature of unintended effects). They help to distinguish design issues and implementation problems. They can be shaped by the questions of interest to stakeholders and the assumptions embedded in program and project interventions (Bamberger, Rao & Woolcock, 2010).

Conclusions

Campbell's evolutionary epistemology provides a robust intellectual framework for probing knowledge creation, social learning, and evaluation. Blind variation and selective retention illuminate the unique remit of the evaluation discipline, its criteria, and its mechanisms. The parallels between natural and artificial selection and the metaphors associated with ontogeny, phylogeny, biological co-evolution, positive and negative feedbacks, nested hierarchies, etc. are instructive.

Thus, biological evolution theory helps to adapt evaluation methods to various phases of social intervention life cycles, provokes useful reflections about complexity theory in evaluation, and generates helpful guidance towards the assessment of program theories, the evaluation of project portfolios and the design of social partnerships.

Finally, an evolutionary perspective on the advent of an experimentalist initiative in international development suggests that Campbell's utopian vision of an Experimenting Society came into being on an international stage because it happened to fit the contemporary neo-liberal, evidence-based public policy climate and because it was funded and controlled by aid donors and wealthy private foundations for implementation in the zones of conflict and turbulence of the developing world rather than in the industrial democracies, i.e., without need for deep public involvement and support.

Based on available evidence, the serious reservations that Campbell's tentative proposals for an Experimenting Society had evinced within the evaluation community when it was unveiled have been validated. Wisely, Campbell had suspended his advocacy of the Experimenting Society pending reality testing and he stressed that "we should anticipate its dangers and misuse as well as its promises…and that we should keep open the possibility that we will end up opposing it".

Notwithstanding the recent awarding of the Nobel Memorial Prize in Economic

Sciences to three pioneers of randomized control trials, the disappointing results of the development experimentalist initiative as applied so far suggest that, from an evolutionary perspective, it would be judicious to revise the approach and adopt mixed methods (Stern et. al., 2012) while continuing to celebrate Campbell's extraordinary and inspiring intellectual achievements in the social research and evaluation domain.

References

Alkin, M.C., & Christie, C.A. (2004). Evaluation Theory Tree Revisited, in Alkin, M.C. (Ed.), *'Evaluation Roots: Tracing Theorists' Views and Influences'* (pp. 381-392). Sage Publications.

Bamberger, M., Rao V., & Woolcock, M. (2010), Using Mixed Methods in Monitoring and Evaluation, Experiences from International Development. *World Bank Research Group, Policy Research Working Paper 5245*

Bedecarrats F., Guerin I. & Roubaud F. (2017). All that Glitters is not Gold. The Political Economy of Randomized Evaluations in *Development, Development, and Change*, 0(0) 1-28, International Institute of Social Studies, pp. 1-27

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

Campbell, D. T. (1971), Methods for the Experimenting Society, Paper presented to the Eastern Psychological Association, New York City and to the American Psychological Association, Washington, DC

Campbell, D. T. (1974). Evolutionary Epistemology. in The philosophy of Karl R. Popper (pp. 413-463). P. A. Schilpp (Ed.). LaSalle, IL: Open Court.

Campbell, D. T., & Stanley J. (1974). Qualitative knowing in action research. Kurt Lewin Award Address to the Society for the Psychological Study of Social Issues, 81st Annual Meeting of the American Psychological Association September. 29-30.

Campbell, D.T. (1988). Descriptive epistemology: psychological, sociological, and evolutionary, in Campbell, D.T. Methodology and epistemology for social science: selected papers., E.S. Overman, Ed. (pp. 435-486). Chicago, Ill. : Chicago University Press.

Campbell, D.T. (1990). Epistemological Roles for Selection Theory, in N. Rescher Ed., Evolution, Cognition and Realism (p.7). Lanham, Maryland: University Press of America.

Campbell, D.T. (1998). The Experimenting Society, In Dunn W.N. (Ed.) The Experimenting Society, Essays in Honor of Donald T. Campbell (pp. 35-36). New Brunswick, New Jersey: Transaction Publishers.

Campbell, D.T. & Mohr, P.J. (1950). The effect of ordinal position upon responses to items in a checklist, *Journal of Applied Psychology*, 34, 62-67.

Cook T.D., & Campbell, D.T. (1979). Quasi experimentation: Design and Analysis Issues for Field Settings. Boston, MA.: Houghton Mill

Chelimsky, E., & Shadish, W.R. Eds. (1997). Evaluation for the 21st Century: A Handbook. (foreword). Thousand Oaks, California: Sage Publications

Deaton, A. (2010). Instruments, Randomization and Learning about Development, *Journal of Economic Literature, 48 (2).* 424-455

Duflo, E., & Kremer, M. (2005). Use of Randomization in the Evaluation of Development Effectiveness, in George K. Pitman et.al. (Eds). Evaluating Development Effectiveness, World Bank Series on Evaluation and Development, Volume 7. New Brunswick, N.J., and London, UK: Transaction Publishers

Kahneman, D. (2012). Thinking Fast and Slow, London, UK.: Penguin Books

Kelly R. M., & Gregware, P. (1998). The Experimenting Society: Towards an Inclusive Democratic Community, in Dunn W.N., (Ed.). The Experimenting Society, Essays in Honor of Donald T. Campbell, (pp.189-210). New Brunswick, NJ: Transaction Publishers.

Kirkhart, K.E. (2010). Eyes on the Prize: Multicultural Validity and Evaluation, *American Journal of Evaluation, Vol 31, Issue 3,* 400-413.

Kropotkin, P. (1989) Mutual Aid. Montreal, Canada: Black Rose Books.

Matthews, M. (2017). Transformative Public Policy: A new strategy for coping with uncertainty and risk. London, UK: Routledge.

Moore, M., & Khagram, S. (2004). On Creating Public Value: What Business Might Learn from Government about Strategic Management In *Corporate Social Responsibility Initiative, Working Paper 3, (March)* Cambridge, Mass.: Kennedy School of Government, Harvard University

Munz, P. (2001). Philosophical Darwinism: on the origin of knowledge by means of natural selection. (pp. 151-160). London, UK: Routledge.

Overman, E. S., (Ed.) (1988). Methodology and Epistemology for Social Science: Selected Papers of Donald T. Campbell, (pp. 290-314). Chicago, Ill.: University of Chicago Press.

Patton, M.Q. (2011). Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use. New York, NY: The Guilford Press

Popper, K. R. (1972). Objective Knowledge: An evolutionary approach. Oxford, UK: Oxford University Press.

Popper, K. R. (1974). Replies to my critics. In The philosophy of Karl R. Popper (pp. 1061-1062)
P. A. Schilpp (Ed.). LaSalle, IL: Open Court.

Popper, K. R., (1978). Three Worlds, The Tanner Lecture on Human Values, Delivered the University of Michigan, April 7. http://www.15926.info/functional-physical-object/popper80.pdf,

Quine, W.V. (1969), Naturalized Epistemology In Ontological Relativity and Other Essays. (pp. 69-90). New York, NY: Columbia University Press.

Raup, D.M. (1994), The role of extinction in evolution. In Colloquium Paper. *Proceedings of the National Academy of Science, Vol. 91,* 6758-6763,

Ravallion, M. (2009). Evaluation in the Practice of Development. In Washington, DC (pp. 29-53): *The World Bank Research Observer, 24 (1).*

Robertson, D.S. (1991). Feedback theory and Darwinian Evolution, In *Journal of Theoretical Biology, 152,* (pp. 469-484) Academic Press Ltd.

Scriven, M. (1991). Evaluation Thesaurus, (pp. 363-364), London, UK: Sage Publications.

Scriven, M. (2016). The Last Frontier of Evaluation: Ethics, in Donaldson, S.I. and Picciotto, R. (Eds.), Evaluation for an Equitable Society', (pp. 38-39) Charlotte, NC: Information Age Publishing.

Shadish W. R. & Cook, T.D. (1998). Donald Campbell and Evaluation Theory, In *American Journal of Evaluation, Vol 19, No.3.* (pp. 417-422) Colloquium Paper. Sage Publications.

Shaver, P. & Staines, G. (1971). Problems Facing Campbell's 'Experimenting Society. In *Urban Affairs Quarterly, Vol, 7, Issue 2, December* (pp. 173-186). Sage Publications.

Stern, E., Stame, N., Mayne, J., Forss, K., Davies, & Befani, B. (2012). Broadening the Range of Designs and Methods for Impact Evaluation. Working Paper 38. London, UK: Department for International Development.

Stufflebeam, D. L. *(2001).* The Metaevaluation Imperative. In *American Journal of Evaluation, 22.* (pp. 183-209). Sage Publications.

Suchman, E. A. (1967). Evaluative Research: Principles and Practices in Public Service and Social Action Programs. New York, NY: Russell Sage Foundation.

Thaler, R. H. (2015). Misbehaving: The Making of Behavioural Economics. New York, NY: W.W. Norton Publishers.

Toulmin, S. E. (1972), *Human Understanding*, Princeton, NJ: Princeton University Press.

Trochim, W.M.K. (1998), Donald T. Campbell and Research Design. In *American Journal of Evaluation, Vol. 19, No. 3*, 407-409.

Trochim, W.M.K. (2007). Evolutionary Perspectives in Evaluation: Theoretical and Practical Implications, Paper presented at the Eastern Evaluation Research Society, April 24. pp.9-10. https://www.socialresearchmethods.net/research/EERS2007/Evolutionary%20Perspectives%20in%20Evaluation%20Theoretical%20and%20Practical%20Implications.pdf

Van de Ven, A. H. & Grazman, D. N. (1999). Evolution in a Nested Hierarchy: A Genealogy of Twin Cities Health Care Organizations, 1853-1995. In Baum, J.A.C. & McKelvey, B., Variations in

Organization Science, In Honor of Donald T. Campbell.

Vedung, E. (2010). Four Waves of Evaluation Diffusion. In *Evaluation, Sage Publications, July, Vol 16, Issue 3,* 263-277

Weiss, C.H. (1998). Evaluation: Methods for Studying Programs and Policies. Second Edition. , Upper Saddle River, NJ: Prentice Hall Publishers.

Weiss, C.H. (2000). The Experimenting Society in a Political World. In Chapter 11, Bickman, L. (Ed.). Validity and Social Experimentation, Donald Campbell's Legacy. Thousands Oaks California: Sage Publications.