

Evaluation for an Open Society: Then and Now



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

Robert Picciotto
University of Auckland

Background: Karl Popper's views about science and political economy remain relevant to evaluation theory and practice. His Open Society opus inspired pioneering contributions to experimental evaluation and shaped the evaluation discipline. Yet, his ideas are not widely known without the evaluation community even though populist leaders are once again threatening to undermine democracy.

Purpose: To define the Open Society, probe its epistemological tenets, confirm that they remain valid as the foundation of evaluation practice, identify the ways in which the operating environment for evaluation has changed and, against this background, propose a policy change agenda relevant to the contemporary evaluation discipline.

Setting: The Open Society is once again being undermined. Modern authoritarianism is tightening its grip. The lure of the strong man is once again gaining traction. The dominance of an international order grounded in democracy, human rights, and the rule of law is giving way to a world in which leaders are pursuing narrow nationalist and vested interests. In this troubled context, policy making has become more complex than when evaluation emerged out of the ashes of World War II. Economic and social dysfunctions have led to extraordinary concentration of wealth and privilege. Ominous environmental threats loom. The architecture of international relations designed in the mid-1940's has become obsolete.

Research Design: To design this commentary about the prospects of the evaluation discipline, the author drew on his personal experience as evaluation academic, international development practitioner, manager of the World Bank's Independent Evaluation Group for two consecutive five-year terms and senior independent evaluation adviser to governments and international development agencies.

Intervention: As an intervention, this article adds value to evaluation theory and practice by showing why and how the Popper/Campbell mandate for evaluation needs to be upgraded to protect the public interest in a new operating environment. Specifically, Popper's piecemeal social experimentation concept should be refined to forge links between small scale experiments and the broader fabric of society. In addition, the ambiguity regarding the relationship between the Open Society and evaluation should be lifted through a reconsideration of the democratic evaluation model.

Data Collection and Analysis: The author conducted an extensive review of the literature and consulted with a wide range of evaluation thinkers to examine the extent to which Popper's philosophy remains relevant to the evaluation discipline.

Findings: Popper's Open Society ideas aimed at avoiding the rise and perpetuation of autocracy and remain highly relevant. But the current threats to democracy call for a more ambitious and detailed remit for the evaluation occupation. Beyond the promotion of evaluation in democracy and of democracy in evaluation, evaluation for democracy should be pursued. This implies putting value, ethics, and the public interest at the very center of the evaluation occupation; breaking free of Popper's parsimonious piecemeal social engineering concept to inform systemic social reform; bringing peace to a methodologically divided house; systematic mixing of evaluation methods and models; and the promotion of evaluation independence through professionalization.

Keywords: *democracy; experimentation; falsification; paradigm wars; piecemeal social engineering.*

Introduction

Karl Popper is not widely known within the evaluation community. Yet, he had a major influence on Donald T. Campbell, whose pioneering contributions to applied social research shaped the evaluation discipline from its very start (Rieper, 2004). Popper's perspectives on knowledge creation have been especially influential and his first and magisterial contribution to political philosophy (*The Open Society and Its Enemies*) addresses issues that are still central to the evaluation function today.

This article responds to Campbell's challenge to "contribute to the best possible exploration" of the actions needed for evaluation to contribute to the Experimenting Society, an offshoot of the Open Society (Campbell, 1998, p. 37). First, I define the Open Society, probe its epistemological tenets, and confirm that the Open Society paradigm remains valid as the foundation of contemporary evaluation practice. Second, I contrast the current operating environment for evaluation with the context that prevailed when Popper and Campbell put forward their vision of the good society. Third, I argue that the current evaluation community needs to get its act together to respond to the threats currently faced by the Open Society. Fourth, I lay out a change agenda for evaluation designed to help address these threats. Fifth, I conclude.

What is the Open Society?

The Open Society (Popper, 2013) was an immediate sensation and it is still required reading in political science courses. First published in 1945, the two volumes that make up the book were completed by 1943 when Nazi Germany was still conquering much of Europe. Since the tides of war had not yet turned, Karl Popper viewed the Open Society as his 'war effort' (Schilpp, 1977, p. 91).

By the time the Open Society was published the western democracies had vanquished Hitler's forces, but Popper's classic consolidated the cosmopolitan democratic consensus that triumphed when the Soviet Union imploded in the late 1980s and it remains an inspiration for democratic activists. Grounded in Enlightenment values, it champions human rights, rejects the reactionary excesses of nationalism, and unmask the risks that populist leaderships pose to life and liberty. This explains why its major themes resonate today.

The critical rationalist beliefs that underlie the Open Society influenced evaluation theory and

practice and inspired Donald T. Campbell's contributions to social research. Like Popper, Campbell viewed evidence-based decision making and the resolution of political differences through rational debate as essential to human progress – two major ways in which evaluation contributes to the effective workings of democracy.

The Open Society has Limited Aims

While Popper valued democracy, the Open Society does not embody a specific democratic model. Popper borrowed the concept from Henry Bergson (1935) who contrasted closed societies (static, immune to change, constrained by tribal, collectivist or religious beliefs) with open societies that are dynamic, responsive, pluralistic, and respectful of individuals.

Popper viewed safeguards against dictatorial rule as the acid-test of the Open Society not only because dictators tend to abuse their power but because, even if they are benevolent, they rob citizens of their fundamental rights and civic duties. The Open Society protects freedom of thought, critical debate, and pursuit of knowledge. On the other hand, Popper distrusted all utopian visions and observed that sharply contrasting values and principles are inevitable in a democracy:

...we shall always have to live in an imperfect society. This is not only because even very good people are very imperfect; nor is it because, obviously, we often make mistakes because we do not know enough; even more important than either of these reasons is the fact that there always exist irresolvable clashes of values: there are many moral problems which are insoluble because moral principles conflict (Schilpp, 1977, p.92)

Concentrating on the narrow question of how to avoid the emergence (or perpetuation) of autocracy (Popper, 1988), Popper stood at the centre of the political spectrum. The Open Society is not a neo-liberal laissez faire paradise even though market fundamentalists hold Popper in high regard. Nor is Popper's credo supportive of the libertarians who have long attempted to capture his legacy (Childs, 1976).

Sound Epistemological Foundations

Popper's commitment to the Open Society, shared by Campbell, was epistemological rather than political. Eschewing any blueprint for a democratic society, Popper concentrated on the prerequisites

of human progress through knowledge creation. He was instrumental in debunking the positivist paradigm that previously held sway. His post-positivist stance secured widespread support and credibility within the social scientific community and made possible the advent of evaluation as a distinct discipline.

First, Popper rejected the Vienna Circle view that there is a single reality that can be conclusively identified by observation even in the absence of a theory. Recognising that research is influenced by the values of investigators and by the hypotheses that they hold, Popper stressed the fallibility of human knowledge and resisted the exaggerated claims of social scientists who look for problems that fit their solutions instead of acknowledging that their theories are only conjectures subject to refutation.

On the other hand, Popper did not share the constructivist belief that all knowledge claims are entirely dependent on social arrangements so that perceived truth cannot be divorced from ideological preconceptions. He opposed the notion of reality as a pure social construct wholly dependent on individuals' worldviews. Nor did they endorse the logical positivist mental model according to which the existence of the external world can be verified conclusively or that valid knowledge claims are independent of the social and cultural context.

In sum, the Popper/Campbell epistemology falls between the extremes. It is grounded in the presumption that while reality exists, it is only experienced indirectly and imperfectly through observation processes that are invariably mediated. It holds that the only inferences that can be legitimately drawn from an experiment depend on the articulation of refutable causation theories. It asserts that rational decision making in the public sphere can only be guided by plausible though fallible, context dependent knowledge derived from rigorous reality testing, scrupulous self-criticism, and principled debate. These remain foundational tenets for the evaluation discipline.

The Dangers of Utopian Thought

By and large, evaluators, a sceptical lot, embrace Popper's rejection of essentialist, historicist, and holistic ideas that are the bedrocks of autocracy. Essentialism aspires to unearth the hidden nature of things, their eternal essence. Historicism seeks to discover basic laws and principles that drive historical change towards an ineluctable end-point. Holism rejects partial explanations solely focused on individual constituents of society.

Together, these beliefs negate diversity of opinion, favour collectivist undertakings and privilege sectional interests over individual rights. They also contribute to a predisposition to brutal revolutionary change irrespective of the human costs involved. Yet, violent revolutions driven by rigid utopian thinking fail to create adaptable social structures. By contrast, the Open Society stands for gradual change, informed by critical discussion, and trial and error practices.

Thus, for Popper and Campbell, sustainable social progress just as scientific advance can only be incremental since knowledge of the social world can only be partial, conjectural, and dependent on free-wheeling debate, checks and balances, and self-correcting mechanisms. These are also the attributes of the Experimenting Society (1998) in which Campbell laid out a vision of evaluation as a feedback mechanism that provides rational guidance for political decision making towards a better world.

The Imperative of Experimentation

From this perspective, social learning, just as the search for scientific truth, is a never-ending voyage. It proceeds systematically through discrete steps with ample room for course corrections. The resulting incremental adjustments can always be improved upon. Thus, Popper and Campbell viewed all scientific knowledge as conjectural and provisional. They held the view that no theory is safe from refutation, both in the physical sciences and in the social sciences.

Theory based evaluation that now dominates the evaluation field is solidly grounded in the scientific method as conceived by Popper/Campbell: the assumptions that make up a program theory can only be refuted through experimentation, the province of summative evaluation. Reality testing of intervention design and implementation is the core objective of the evaluation discipline – an integral part of accountable and responsive management in all sectors of society.

Falsification, a core premise of the scientific endeavour, also describes how evaluation works – or should work. It relies on independent testing of specific conjectures through systematic processes validated by the collective wisdom of peers. This intellectual stance is fully consistent with Popper's post-positivist premise: while conclusive verification of an overarching theory cannot be achieved, it can be ruled out as false through evidence.

A Modest Social Agenda

These epistemological ideas informed Popper's worldview and they were echoed by Campbell who also favoured discrete problem solving through piecemeal social engineering. Rather than seeking comprehensive solutions to social problems, the rational agent of the Open Society makes his way "step by step, carefully comparing the results expected with the results achieved, and always on the lookout for the unwarranted consequences of any reform; and he will avoid undertaking reforms of a complexity and scope which make it impossible for him to disentangle causes and effects" (Popper, 1944, p. 309).

Again, this is the province of evaluation. Unlike social researchers, evaluators do not normally seek generalized answers to social problems. They recognize the uncertainty and risks associated with translating intellectual knowledge into practical action across contexts. They consider their main task to be a search for valid evidence about social interventions in ways that facilitate gradual, iterative, crablike social advances adapted to distinct operational circumstances. This perspective underlies a conception of evaluation as an applied knowledge endeavour: a discipline dedicated to reality testing through case by case assessment.

From this perspective, sound politics just as legitimate science (and principled evaluation) require freedom of thought and the institutionalization of fact-based, knowledge generation processes. Guided by rational scrutiny of competing theories and initiatives—rather than rigid compliance with pre-conceived dogma—both science and evaluation are characterized by a contest of ideas, systematic examination of available evidence, a willingness to reconsider premises, and, when the facts warrant it, admit error and shift course. These are also the societal factors that generate an effective demand for high quality, ethical evaluation services.

The New Operating Environment for Evaluation

Today's evaluation activities take place in an operating context where the Open Society is threatened. The fundamental concerns that led Popper to focus on the risks to humanity posed by the enemies of open societies remain valid, but the contemporary context differs markedly from that which prevailed in the mid-twentieth century when

the Open Society burst onto the intellectual scene. Hence, adjustments to the Popper/Campbell vision are required. This section of the article sketches in broad strokes the contemporary threats to the Open Society and moves on to draw the implications for evaluation of the major challenges faced by the Open Society today.

The Open Society is Under Siege

The global diagnostic regarding democratic trends offered by Freedom House (2017) is stark. The dominance of an international order grounded in democracy, human rights, and the rule of law is giving way to a world in which individual leaders and nations are pursuing their own narrow interests. Modern authoritarianism is tightening its grip around the world through sophisticated strategies of repression that muzzle the press, undermine the civil society, and exploit divisions among open societies. The lure of the strong man is once again gaining traction.

After two decades of major gains (from 34% in 1986 to 47% in 2006), the share of countries classified as free has peaked over the past decade dipping down to 45% in 2016. The number of electoral democracies stood at 123 in 2016, two fewer than in 2015. For the 11th consecutive year 2016 witnessed declines in freedom indicators that outnumbered improvements. While in past years freedom declines were concentrated among autocracies, setbacks in countries classified as free have begun to matter. Almost one-fourth of nations displaying declines in 2016 were in Europe.

In this troubled context, policy making has become far more complex than when evaluation emerged out of the ashes of World War II as a transmission belt between the academy and public policy making. Simplistic views, divorced from reality, distort the political discourse (Nielsen and Graves, 2017). With growing polarization, the rational dispositions required to seek pragmatic solutions to human problems, and the tolerance needed to reconcile rival interests through principled debate, are in short supply (Abramovitz and Saunders, 2008).

Doctrinal rigidity and growing religious and ethnic intolerance are undermining human security. Public frustration regarding the broken promises of liberal democracy has induced xenophobia. Referendums that displace principled compromises among elected representatives have become more frequent, a possible sign of indirect democracy in distress. The public sphere is degenerating into a spectacle of polarized, warring

policy stances no longer constrained by facts. Social media dominate public opinion and work as echo chambers that merely confirm biases and prejudices.

Lack of principled debate compounds the problem. Politically motivated disinformation exploits public anger and frustration. It mimics professional journalism and crowds out objective, balanced reporting. Journalists are discredited, intimidated, or prosecuted for doing their job. Absent reliable information, fruitful, open dialogue is stunted. Thus, the fragility of democratic regimes that Popper had diagnosed when he turned his attention from the philosophy of science to the philosophy of politics is again in evidence.

The New Operating Context

The post-World War II era was one of optimism and public faith in the capacity of the industrial democracies to improve livelihoods through government intervention and public investment. By contrast, in the current operational environment, the public sector has lost much of its influence. Earlier preoccupations regarding the excessive state power that propelled neo-liberalism to the commanding heights of policy making world-wide have been superseded by concerns about the threats that vested interests pose to policy making ...and to evaluation (House, 2016).

Being debt burdened, most western governments have become subservient to cash rich, tax evading, multinational corporations that have perfected the art of political influence. Through the new information technologies, trade, aid and migration, economic globalization has taken hold and marginalized the state (Collier and Dollar, 2002).

Multinational corporations have acquired enormous power. The information revolution combined with trade liberalization has rendered distance irrelevant so that borders no longer constrain private investment decisions. Equally, aid has induced developing countries to connect to the mighty engine of the global economy and to encourage foreign direct investment. They have become the engines of the global economy. A great convergence is underway. Global poverty was cut by more than half between 1990 and 2015 while de-industrialization hollowed the middle class in western societies.

Rising Public Discontent

The causes of public disillusionment with the performance of liberal democracies are many (Stiglitz, 2002). Economic and social dysfunctions have led to extraordinary concentration of wealth and privilege, with the deleterious consequences envisaged by Campbell (1998, p. 44) who feared that the Western democratic capitalist model was vulnerable given “the role of wealth in providing grossly uneven weightings of some preferences over those of others” [and in creating] “great obstacles that may effectively sabotage program decisions genuinely based on the public good.”

Ominous environmental threats that transcend political boundaries have also made the fate of nations inextricably linked (Kanie and Haas, 2004). The architecture of international relations designed in the mid-1940s has become obsolete (Raustiala, 2002). Inequality has surged. Eight men own the same wealth as the 3.6 billion people who make up the poorest half of humanity, according to Oxfam (2017).

According to the International Monetary Fund (Dabla-Norris, Kochkar, Suphaphit, Ricka, & Tsolunta, 2015, p. 4): “Widening income inequality is the defining challenge of our time. In advanced economies, the gap between the rich and poor is at its highest level in decades. Inequality trends have been more mixed in emerging markets and developing countries, with some countries experiencing declining inequality, but pervasive inequities in access to education, health care, and finance remain.”

Implications for Contemporary Evaluation

These then are not propitious times for the Open Society. More than ever, evaluation is needed to protect the public interest: it blends legitimacy with expertise and contributes to accountability and social learning in a bewilderingly complicated world characterized by impersonal governance structures, intricate supply chains, opaque financial instruments, and pervasive sub-contracting that diffuse responsibility and undermine the public interest.

The Imperative of Adaptability

There is no doubt that Popper favored democracy. But he did not hold definitive views about the form that democracy should take (e.g. representative vs.

direct, technocratic vs. participatory, etc.) and he would have stoutly resisted the hegemonic notion that Western society constitutes the endpoint of human social evolution (Fukuyama, 1992).

Thus, for Popper, liberal democracy was not the best of all possible worlds – only the best of all existing worlds (Jarvie and Pralong, 1999, p. 37). Still, he constructed the Open Society based on broad, tolerant, and adaptable political governance assumptions focused on “setting free the critical powers of man.” Specifically, he defined the Open Society as “a society based on the idea of not merely tolerating dissenting opinions but respecting them” and democracy as “a form of government devoted to the protection of an open society” (Popper, 1994, p. 110).

On the one hand, the broad, nearly agnostic conception of governance of the Open Society has facilitated the practice of evaluation across all sectors of society and all parts of the world. This conception is aligned with the varied requirements of contemporary evaluation practice: evaluation is now operating in the private and voluntary sectors of society as well as in the public sector. Furthermore, evaluation is becoming more “international in the sense of being at the same time more indigenous, more global, and more transnational” (Chelimsky and Shadish, 1997, p. xi).

At the turn of the century, there were only 20 evaluation associations in existence. Since then, the number has exploded: EvalPartners, under the auspices of the International Organization for Cooperation in Evaluation (IOCE), has identified a total of 158 associations or networks, of which 135 are at the national level. This means that evaluation today must adapt to a wide variety of governance contexts. Its success as a knowledge occupation hinges on its adaptability.

Towards a More Ambitious Mission

Without jettisoning the pragmatic stance that has contributed to the adaptability of the evaluation discipline, Popper/Campbell’s modest evaluation mandate needs upgrading to address the unprecedented challenges faced by the Open Society today. To be sure, evaluation is an “art of the possible” that requires nimble adaptation to a variety of governance contexts. However, the current ambiguity regarding the relationship between the Open Society, democracy, and evaluation should be lifted since it is hindering clarity of purpose for the evaluation occupation.

However, prior to exploring the intersection between evaluation and democracy, it is

appropriate to recognize that a more ambitious remit is required for contemporary evaluation. Updating the Open Society means breaking free of the parsimonious “piecemeal social engineering” compact to rise to contemporary social challenges. Incremental changes applied haphazardly will not be enough to achieve the social transformation required to make meaningful progress towards an inclusive democratic community (Kelly and Gregware, 1998).

While the threats that led Popper to expose the intellectual enemies of the Open Society are once again looming, the negative utilitarian stance that might be inferred from a wholesale rejection of holism runs the risk of discouraging the identification of high leverage interventions capable of triggering major shifts in economic and social policy. Especially relevant to today’s social predicament is Albert O. Hirschman’s “bias for hope” (1971) according to which small ideas and discrete interventions can be selected, sequenced, and aggregated to be transformative.

According to Hirschman, fundamental institutional change is more likely to be triggered in periods of latent crisis (such as the world is now going through) since they relax the constraints that normally limit policy makers’ freedom of action: where public discontent turns acute, new opportunities for progressive reforms arise. In such circumstances, a window of opportunity opens and “there is a special justification for the direct search for novelty, creativity, and uniqueness: without these attributes change, at least large-scale social change, may not be possible at all” (Hirschman, 2013, p. 22).

This conception of projects as “privileged particles of the development process” (Hirschman, 1967, p. 1) goes beyond Popper’s piecemeal social experimentation concept since it aims at large-scale replication. Achieving social progress on a significant scale requires forging links between small scale experiments and the broader fabric of society. From this perspective, the social scientist’s search for regularities, stable causal relationships, and predictable sequences should not be summarily dismissed since it may help identify high leverage social interventions suitable for upscaling or replication following systematic experimentation.

Evaluation at a Higher Plane

At a time of unprecedented and interlocked threats to democracy, Popper’s social tinkering remains relevant but only as a first step towards transformative change. If so, evaluations should

serve as policy experiments aimed at the adoption of far-reaching progressive reforms. This was an incipient premise of the Experimenting Society. Thus, Campbell (1998, p. 47) observed that evaluative “skills should be reserved for the evaluation of policies and programs that can be applied in more than one setting.”

In the new operating context, this aspiration has become an imperative and meta-evaluation should come center stage in evaluation theory and practice. But clarity in the relationship between evaluation and democracy is also essential since in the words of Seneca (2016, p. 182): “The archer must know what he is seeking to hit; then, he must aim and control the weapon by his skill. Our plans miscarry because they have no aim. When a man does not know what harbour he is making for, no wind is the right wind.”

Evaluation of Democracy

The term democracy (‘rule of the people’) is malleable. Views among scholars as well as citizens about the meaning of democracy and how it should be assessed vary widely. But the basic concept is so universally popular that all governments, including dictatorships, claim to be democratic. In the west, many variants of democracy exist depending on the degree to which eligible citizens participate in political decision making, either directly, indirectly (through elected representatives) or through a head of state.

By way of illustration, a broad-based consensus of western opinion captured by Freedom House (2017), a US think tank, provides a template for assessing the quality of democracy in different countries. It defines a liberal democracy as one that:

- (i) carries out free and fair elections;
- (ii) tolerates political pluralism and encourages participation;
- (iii) is free of corruption and operates transparently;
- (iv) protects freedom of belief and expression;
- (v) guarantees freedom of assembly and association;
- (vi) respects the rule of law and individual rights.

But this template is not a universally accepted way of defining let alone evaluating democracy. The metrics used for each of the above categories while plausible are not self-evident; and there is no

uncontroversial way of selecting the weights used to aggregate the various measures. Most of all, the model fails to take outcomes into consideration so that leaders of authoritarian regimes argue that they are better equipped to deliver national strength, prosperity, and equity. This may or may not be the case: a vast literature seeking to relate democracy and economic growth has failed to uncover unambiguous associations (Sen et.al, 2017).

From an international perspective, history as well as culture shape the government models adopted by various countries. Thus, socialist regimes rely on a mix of centralized party control and direct democratic rule at the local level. They reject the basic tenets of western liberal democracy which they consider fictitious because of alleged wealthy elites’ control over elections and politicians.

Campbell (1998) recognized that democratic capitalism had major advantages and displayed favourable features for the establishment of an Experimenting Society (rule of law; frequent elections; pluralistic decision making) but he deplored the resulting ‘accumulation of great inequalities in individual and corporate wealth’ that it spawned, and he detected promising perspectives in Marxist theory that offered some hope for the advent of a truly experimental socialism (p.44). Thus, he did not rule out that evaluation may have a role in authoritarian societies since they might gradually evolve and eventually adopt Open Society principles. This remains a legitimate goal for the evaluation discipline.

Evaluation in Democracy

Fortunately, the Open Society is unconstrained by rigid definitions. An ideal construct in the Weberian sense, it reflects the liberal, market-oriented, democratic, cosmopolitan values that have also inspired the spread of evaluation across national borders. According to Anders Hanberger (2006), democratic governance models fall into three categories:

- (i) elitist democracies where decision making is largely delegated to elected officials and technocrats;
- (ii) participatory democracies where citizens are active participants in decision making; and
- (iii) deliberative democracies where collective decisions are taken following

principled debate among free and equal citizens.

Barry MacDonald's (1976) useful distinction between bureaucratic, autocratic, and democratic evaluations helps to pinpoint how evaluation serves each of these three democracy models. Bureaucratic evaluation is the instrument of choice for decision makers in elitist democracies – as well as in authoritarian regimes and the private sector. This model dominates evaluation practice today. It provides unconditional support to decision makers who own the evaluation process and control its design as well as the disclosure of its findings.

By contrast, autocratic evaluation is autonomous. Mostly protected by the ivory towers of academia, it seeks to advance social learning as well as provide expert advice to decision makers. While dictators, wealthy philanthropists, and corporate leaders are loath to use it, enlightened decision makers in elitist democracies do rely on it to secure reliable information about the public interventions that they manage.

Finally, inclusive evaluation serves deliberative or discursive democracies. It takes many forms depending on the extent of citizens' participation in the evaluative process. Barry MacDonald's version merely informs debate and "promotes knowledge of the programme on the part of those who have a right to know, or the duty to advise, or the obligation to provide, or the power to stipulate." It "stops short of advocating courses of action... modesty is called for" (Stake, 2000, p.100).

By contrast, the deliberative democratic evaluation model promoted by House and Howe (1999) is highly participatory: it includes all relevant stakeholders; promotes dialogue, involves extended deliberation processes. It is specifically designed to serve deliberative democracies. Finally, transformative, emancipatory, and empowerment approaches aim to level the playing field of society. Well adapted to civil society organizations, they are used by evaluators (whether bureaucratic, autocratic, or inclusive) whose values give privileged status to social justice.

It should be clear by now that for evaluation to be effective and fulfil its remit in diverse governance contexts an admixture of bureaucratic, autocratic, and inclusive evaluation models is necessary. Each of the three models has its strengths and weaknesses so that synergies may be tapped by combining them judiciously. Thus, bureaucratic evaluation enjoys low transaction costs and suffers from few information asymmetries. Improving its quality makes autocratic and inclusive evaluation more effective.

Equally, autocratic evaluation enhances the value of other evaluation approaches by attesting to the validity of their methods and providing alternative and expert perspectives to process. Finally, inclusive evaluation amplifies the voice of weak, neglected, underprivileged groups in the evaluative process (whether bureaucratic or autocratic) and as a result focuses attention and resources on the most pressing threats to the legitimacy of Open Societies.

Democracy in Evaluation

Accountability is a common requirement for all democratic governments. In autocracies, lower levels of government are also expected to be accountable. A major function of evaluation is to help ensure that promises are compared with actual delivery through fair, impartial, and transparent evaluative processes.

Democracy in evaluation also implies validity of evaluation findings, objectivity of merit assessments, impartiality of evaluative judgments and transparency as well as timeliness of evaluation processes. Ethical guidelines address professional integrity, respect for the security, dignity and self-worth of the respondents, program participants, clients, and other stakeholders and sensitivity to local and cultural concerns.

Finally, democracy in evaluation requires governance structures and contractual protocols that protect the independence of evaluators and the integrity of evaluation processes. The checks and balances embedded in democratic decision making facilitate the design of appropriate evaluation governance structures, e.g. a reporting relationship to the legislature or to civil society organizations that operate at arm's length from government to ensure unbiased evaluations of programs initiated and carried out by the executive branch.

Is evaluation today equipped to protect the Open Society where it is under threat or to promote a transition towards its advent where it remains elusive? Not unless the evaluation community 'puts its own house in order' to deliver on its more demanding mission.

What is to be Done?

This section highlights five overarching challenges for the contemporary evaluation discipline. First, evaluation policy directions should be reconsidered so that beyond democracy in evaluation and democracy in evaluation, the discipline commits to work for democracy. Second, it is high time to

secure a broad-based consensus about what evaluation is about so that it works for democracy. Third, the paradigm wars should be decisively settled to generate a united front against the enemies of the Open Society. Fourth, beyond mixed methods, evaluation models should be combined in ways appropriate to the context to promote the ideals of the Open Society. Fifth, evaluation should become self-managed through professionalization.

Evaluation for Democracy

Through the long twilight struggle that lies ahead, updating Popper's and Campbell's ideas would help sustain an embattled and divided evaluation community. Evaluation for democracy should complement evaluation in democracy and democracy in evaluation. This implies a more activist stance in authorizing environments that do not tolerate dissent and/or for assignments that are closely controlled by vested interests.

In such contexts evaluators should not be content to operate as pliant advisers or brokers. In dysfunctional governance contexts they should assume ownership of their evaluation products and deserve to be backed up by professional protocols that protect their independence. In parallel, social justice models should come centre stage.

It should no longer be enough for evaluators to attend to their contractual obligations, focus on evaluation utilization, and satisfy those who hold the purse strings. Consultants who use evaluation methods without regard to the public interest should not be called evaluators.

Rather than sitting on the fence, it is time for evaluators everywhere to make their democratic value commitments explicit (Greene, 1997). Equally, the ethical principles that govern evaluation should address the social inequities, power differentials, and human rights abuses that undermine the Open Society. In turn, this requires clarity about the nature of the evaluation endeavour.

What is Evaluation?

Definitions matter since they are the building blocks of theory and since theory is "who we are" (Shadish, 1998, p.5). Among the many formulations on offer, Michael Scriven's Thesaurus definition (1991, p. 139) has evinced broad based support: "Evaluation is the process of determining the merit, worth or value of something or the product of that process". Unfortunately, there is currently no

consensus about the precise meaning of the merit-worth-value trilogy.

Merit is defined as the intrinsic value of the 'evaluand' but does this refer to the value that accrues to an individual consumer (Scriven, 1991, p. 382) or to compliance with quality standards (Scriven, 1998, p.65)? Is it instead assessed by how effective it is in meeting the needs of those it is intended to help (Patton, 2008, p. 113)?

Similarly, worth is the extrinsic value of the 'evaluand'. But is it the value to a specific institution or collective as per Scriven's Thesaurus (1991, p. 382) or does it refer to the value to the larger community or society as proposed by Patton (2008, p. 113)? Should a more recent definition – in which 'significance' stands for importance and value is now embedded in worth – prevail? It states that: "the discipline of evaluation undertakes the systematic, objective, determination of the extent to which any of three perspectives are attributable to the entity being evaluated: merit, worth or significance" (Scriven, 1998, p. 65).

These definitions are goal-free, but this is not how they are always interpreted. Goal-based, merit-oriented approaches have dominated evaluation practice. Goal related assessments can enhance accountability but unless evaluation is genuinely independent (see below) they can be captured by power holders through fee dependence and contractual obligations.

Since power holders, especially in the private sector, have usually reached their position because inter alia they are self-confident in their capacity to assess situations and committed to achieving their chosen objectives, they disdain speculation, favour certainty, and rely on quantitative indicators of progress to manage their programs.

Consequently, program managers have welcomed the instrumental quality of goal-based, indicator driven evaluation and its reliance on a tight results chain that links inputs, outputs, outcomes, and impacts. Yet, as noted by Campbell (1998, p. 55) "the more any social indicator is used for social decision making, the greater the corruption pressures upon it" and, as stressed by House (1990) social validity hinges on an objective assessment of value from a public interest perspective.

Similarly, worth assessments should confront the incommensurability of social preferences and sundry dilemmas of collective action before they can be considered valid for the overall society. Only value driven evaluation informed by ethics can be expected to resolve the aggregation dilemma of valuing interventions across individuals and groups. This calls for ensuring that evidence is

collected and interpreted in ways designed to verify or refute that legitimate intrinsic merit criteria have been met while, at the same time, ascertaining the extrinsic worth of the intervention from the perspective of the citizenry.

In other words, aggregating merit, and worth assessments from a public interest perspective reliant on valid ethical judgments to determine the overall value to the Open Society is what evaluation is ultimately about. Such a remit would be in line with Deborah Fournier's view (2005, p. 140) that "it is the value feature that distinguishes evaluation from other types of inquiry, such as basic research, clinical epidemiology, investigative journalism, or public polling".

This leads to the following question: shouldn't the identity of evaluation, a public service, be grounded in the recognition that, teleologically speaking, worth assessments are vital complements to merit assessments since, according to John Rawls (1971, p.24) the good is defined independently from the right and the right maximizes the good? Is this not the premise implied by the final, rubric directed, synthesis phase of the evaluation process recommended by Scriven (1994)?

This phase of the evaluation process, the most contentious but also the most important, implicitly recognizes the need to combine merit and worth assessments (Scriven, 2016). To be sure, the original definition that allowed free standing merit or worth assessments to qualify as evaluation had the beneficial effect of giving wide berth to different conceptions of the evaluation enterprise. But it did not bring about a cohesive and principled stance about the social remit of the evaluation discipline.

Given the unintended consequences associated with Scriven's initial, highly adaptable, and permissive definition, it would be timely to promote a stricter meaning of evaluation. An explicit re-definition of evaluation, legitimized by evaluation societies, would put forward the final integrative phase of the evaluation process at the centre of the discipline and recognize the crucial role of ethics in aggregating merit and worth assessments.

Delivering on this promise will not take place as long as doctrinal differences continue to fragment the evaluation community to the puzzlement of evaluation users and the public at large. To this chronic challenge I now turn.

Bringing Peace to a Divided House

The current troubled state of the evaluation discipline grows out of a sustained confrontation between rival advocates of quantitative and qualitative methods. Vicious paradigm wars have dominated evaluation history. An uneasy truce now prevails but skirmishes still erupt. A genuine rapprochement has proved elusive. This unsettled state-of-affairs undermines the credibility of the evaluation function in the public sphere.

Pragmatic decision makers faced with urgent social problems cannot grasp what the fuss is about and increasingly seek comfort in the bland certainties offered by management consultants, auditors, and econometricians. It is high time for evaluators to put their house in order. Paradoxically, it is the very success of Campbell's quantitative methods at the creation of the evaluation discipline that ignited the paradigm wars.

The spurious claim that randomization is the evaluation gold standard brought forth a furious reaction from three sources: (i) anti-naturalists who argued that scientific methods are invalid in human affairs since social phenomena are unstable and variable across contexts; (ii) constructivists who asserted that the search for objectivity is futile since intentions, goals and values inevitably feed into socially constructed interpretations; and (iii) critical theorists who detected power machinations in the experimentalists' knowledge construction strategies.

Together these qualitative methods proponents elicited the rise of a dialogic, inclusive, and participatory wave of evaluation diffusion (Vedung, 2010) that swelled in the 1970's, a highly creative period in evaluation history. The pendulum swung back in the 1980's when a neo-liberal wave engulfed the evaluation discipline. Soon thereafter, mostly through attrition, the paradigm wars abated as the evidence-based wave swelled in the 1990s. Mixed methods were touted as the solution to the perennial inter-disciplinary conflict (Stern, Stame, Mayne, Forss, Davies, & Befani, 2012), but doctrinal divisions persist.

The "randomistas" continue to pursue incontrovertible evidence of verifiable "results" while interpretative-qualitative studies and critical theory analyses flourish (Gage, 2009). The evaluation scene is now characterized by competitive turmoil under the big tent of a rapidly expanding and increasingly cosmopolitan evaluation movement (Chelimsky and Shadish, 1997).

Table 1
More Evaluation Paradigms and Models (Adapted from Mertens & Wilson (2012, p. 41))

Paradigm	Branch	Models	Proponents
Post-positivist	Quantitative	Experimental/Quasi-experimental	Tyler, Cook, Shadish, Rossi, Henry
Constructivist	Qualitative	Responsive	Stake, Cronbach
		Case Study	Morra, Simons
		Consumer-oriented	Scriven
		Fourth Generation	Guba and Lincoln
Pragmatic	Use	UFE/Developmental	Patton
		Ratings and rubrics	Davidson
		Cost benefit analysis	Harberger, Squire
Transformative	Social Justice	Empowerment	Fetterman
		Democratic	McDonald, House & Howe
		Social justice	Mertens
		Culturally sensitive	Kirkhart

Given chronic inter-disciplinary tensions and rival doctrinal temperaments, Howe's view (1988, p. 16) that "no incompatibility between quantitative and qualitative methods exists at either the level of practice or that of epistemology [so that] there are thus no good reasons for ...researchers to fear forging ahead with 'what works'" should be vigorously promoted. Public confidence in evaluation will remain elusive without mutual recognition within the evaluation community that rejecting the scientific method is self-defeating and that qualitative methods offer invaluable support to experimental methods.

Mixed Methods Require Mixed Models

Combining evaluation models would help implement the pragmatic, mixed methods approach to evaluation. Such a stance would not preclude the use of valid social experiments envisaged by the Popper/Campbell vision of an Open Society. But it would make clear to the wider public that evaluation is a serious knowledge occupation committed to the public interest.

Uneasy co-existence characterizes today's highly fragmented evaluation market. Neither the qualitative nor the quantitative camp has prevailed. The wider public remains uncertain as to what evaluation stands for. By using practical consequences, i.e. impact, as the ultimate arbiter,

mixed models would inspire ingenious methodological syntheses that would be gradually refined in line with the experimental thrust of the Open Society.

A host of evaluation models currently compete for attention and resources (Table 1). To secure broad based public support, collaboration among models rather than competition is called for. This would also enhance evaluation quality. Indeed, a judicious combination of models and instruments adapted to the context would enhance the validity of evaluation experiments.

Equally diverse methodological orientations among evaluation team members would help tap synergies. Different evaluation models (Table 2) address different questions and focus on different assessment criteria. They are complementary rather than antagonistic. Since they rely on different methods, they favour triangulation and facilitate cross-verification of findings.

The Imperative of Independence

Even if current doctrinal divisions are overcome, living up to the tighter definition of the evaluation mission, moving evaluation to a higher plane and achieving greater policy influence to strengthen the Open Society will not materialize without professionalisation. This will require significant shifts in evaluation policy directions and vigorous

implementation of the Global Evaluation Agenda universally endorsed as the culmination of the 2015 Evaluation Year (Eval Partners, 2015).

To perform their social duty as professionals committed to the Open Society, it will not be

Table 2
A Taxonomy of Evaluation Models

Model	Major audiences	Main evaluation criteria	Typical questions
Experimental/ Quasi-experimental	Executive Directors, Managers, Economists	Merit	Were the intended outcomes achieved and were they attributable to the intervention?
Responsive	Program staff, beneficiaries.	Worth, Significance	Does the intervention reflect the diversity of stakeholders' views?
Case Study	Program designers, sociologists, anthropologists	Worth	How do various intervention elements fit together to generate impact?
Consumer	Consumers, market researchers	Worth, value	Does the product or service meet expectations at the right price?
Fourth Generation	Stakeholders, Critical theorists	Significance	Does the intervention reflect the full range of political, social, cultural factors?
UFE/Developmental	Decision makers, intended beneficiaries	Merit	Does the evaluation help improve the intervention for its sponsors?
Ratings and rubrics	Funders, citizens, managers	Merit, worth, value, significance	Does the intervention achieve its relevant goals efficiently in a sustainable fashion and with positive impacts?
Cost benefit analysis	Funders, citizens	Value	Does the intervention make effective use of scarce resources?
Empowerment	Neglected groups	Worth	Do beneficiaries have the tools and knowledge to evaluate their own performance?
Democratic	Citizens	Worth, value, significance	Does the evaluation promote inclusion, dialogue, and deliberation?
Social justice	Citizens	Worth, value, significance	Does the intervention address social validity, diversity, and human rights?
Culturally sensitive	Evaluators	Worth, value	Does the evaluator recognize the influence of culture on human behaviour?

enough for evaluators to avoid technical blunders. Nor should it be enough for evaluators to observe current ethical guidelines issued by evaluation societies since they are largely limited to narrow professional ethics (validity, data access, privacy, respect for subjects, etc.).

To be sure, evaluation generates useful, timely and context dependent answers to a wide range of questions asked by decision makers. But incentives

drive behavior and without structural independence, fee dependence acts as an instrument of control: it allows commissioners to shape terms of reference, influence data interpretation, define the meaning of collected evidence and censor evaluation findings.

This leads us back to Campbell's recognition (1998, op. cit., p. 36) that the existing political system seems "at times set up just so as to prevent

reality-testing” and his valid concern that evaluators are “apt to become unwittingly co-opted into a pervasive bias in favour of the already-established governmental and extra-governmental powers who, after all, will usually be the source of our past and future salaries” (1998, p. 57).

This is not to deny the important role that evaluation consultants (who use evaluation tools but are not independent evaluators) play and should continue to play in helping managers improve project and programme design and implementation: as was made clear earlier, bureaucratic evaluation has value, but only autocratic evaluation can stand up to the autocracy of vested interests.

Just as independent auditors attest to the validity of financial accounts, independent evaluators should be routinely tasked with verification of self-evaluation claims by engaging stakeholders and focusing on the public interest.

Evaluation societies have yet to press for reforms in the enabling environment within which evaluation currently takes place. Yet the legitimacy of evaluation as a public service hinges on the extent to which the commissioner’s values and interests are aligned with those of the citizenry. Only independent evaluation keeps the commissioner, the evaluand, and the evaluator at arm’s length.

Promoting such a governance model as good practice would help adopt evaluation protocols that protect evaluation independence without incurring evaluator isolation. It would make use of the checks and balances that distribute authority between different branches of government (or between the owners and managers of a private or philanthropic organization). Only through structural or professional autonomy can the democratic deficit of evaluation as currently practiced be filled.

Rising to the top of the occupational ladder through professionalization would also enhance evaluators’ influence. This requires not only nurturing ethical dispositions, orientation towards the public interest, responsibility for the quality of one’s work, expertise grounded in high quality education and exposure to practice but also autonomous control over the evaluator designation. Beyond professional development initiatives, oversight and accreditation mechanisms characteristic of well-established knowledge professions will have to be considered (Picciotto, 2011).

Professional self-management would help to minimize the threats to evaluation integrity and quality caused by weak standards, capture by vested interests and state interference. It would

require agreement on administrative rules, peer reviews of work quality; disciplining of errant members and in extreme cases stripping them of their designation. These measures would involve risks associated with elitism and meritocracy. But such risks would be manageable by an invigorated and united evaluation community dedicated to Open Society ideals.

Conclusions

The Open Society had a limited aim: to avoid the rise and perpetuation of autocracy. Popper’s conception of democracy was epistemological: a government dedicated to knowledge creation, respectful of dissent and open to change. This provided the right paradigm for an evaluation occupation on the march across sectoral and national borders.

Various approaches have been used to cater to the needs of policy makers in the diverse governance environments within which evaluators are called upon to operate. A flexible ethical stance, a carefully delineated professional remit, a predilection for assessing merit, a reluctance to engage in advocacy have characterized evaluative practice under the big tent of the evaluation occupation.

However, the Open Society is now under siege and evaluation should play its part in protecting it and promoting it. Beyond evaluation in democracy and democracy in evaluation, it is time to take reasoned steps towards evaluation for democracy. A pragmatic approach will be needed to tailor ambition to what is feasible in individual contexts, but the ultimate evaluation goal should not be in doubt: contributing to the Open Society to the maximum feasible extent.

This implies putting value, ethics, and the public interest at the very centre of the evaluation occupation; breaking free of the parsimonious piecemeal social engineering concept to embrace and inform systemic social reform; bringing peace to a methodologically divided house; systematic mixing of evaluation models; and the promotion of evaluation independence through professionalization. This is an ambitious change agenda but as Carol Weiss once remarked: “evaluation is not a stroll on the beach” (1998, p. 325).

References

- Abramovitz, A.I., & Saunders, K.L. (2008). Is polarization a myth? *The Journal of Politics*, 70(2), 542-555.
- Bergson, H. (1935). *The two sources of morality and religion*. New York: Henry Holt and Company.
- 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.
- Campbell, D.T. (1998), The Experimenting society. In W. N. Dunn (Ed.). *The experimenting society: Essays in honour of Donald T. Campbell* (pp. 35 – 68). *Policy Studies Review Annual*, 11. New Brunswick: Transaction Publishers.
- Collier P., & Dollar, D. (2002). *Globalization, growth, and poverty: Building an inclusive world economy*. Washington DC: World Bank.
- Chelimsky, E. & Shadish, W. R., Eds. (1997), *Evaluation for the 21st century: A handbook*. Thousand Oaks, California: Sage Publications.
- Childs, R. A. (1976). The open society and its enemies, book review, *The Libertarian Review*. Retrieved from <https://www.libertarianism.org/publications/essays/popper-open-society-its-enemies>.
- Eval Partners (2016). Global evaluation agenda 2016-2020: Developed by a global multi-stakeholder consultative process. Retrieved from <https://www.evalpartners.org/sites/default/files/documents/EvalAgenda2020.pdf>.
- Fournier, D. (2005). Evaluation. In S. Mathison (Ed.). *Encyclopedia of Evaluation* (p. 140). Thousand Oaks, California: Sage Publications.
- Freedom House (2017). *Freedom in the world, 2017: Populists and autocrats: The dual threat to global democracy*. Retrieved from <https://freedomhouse.org/report/freedom-world/freedom-world-2017>.
- Fukuyama, F. (1992). *The end of history and the last man*. New York: Avon Books.
- Gage, N.L. (2009). The paradigm wars and their aftermath. *Educational Researcher*, 18(7), 4 – 10.
- Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park: Sage Publications.
- Hanberger, A. (2006). Evaluation of and for democracy. *Evaluation*, 12(1), 17 – 37.
- Hayek, F.A. (1982). *Law, legislation, and liberty: A new statement of the liberal principles of justice and political economy*. London: Routledge and Kegan Paul Ltd.
- Hirschman, A. O. (1967). *Development projects observed*. Washington, DC: Brookings.
- Hirschman, A. O. (1971). *Bias for hope: Essays on development and Latin America*. Newhaven, Connecticut: Yale University Press.
- Hirschman, A. O. (2013). *The essential Hirschman*. J. Edelman (Ed.). Princeton, N.J.: Princeton University Press
- House, E. R., (1990). Research news and comment: Trends in evaluation. *Educational Researcher*, 19(3), 24 – 28.
- House, E. R. (2016). Inequality and conflict of interest in evaluation. In S. I. Donaldson & R. Picciotto. *Evaluation for an Equitable Society* (pp. 169 - 80). Charlotte, N.C.: Information Age Publishing.
- House, E.R. & Howe, K. R. (1999). *Values in evaluation and social research*. London and Thousand Oaks, California: Sage Publications.
- Howe, K. R. (1988). Against the quantitative-qualitative incompatibility thesis or dogmas die hard. *Educational Researcher*, 17(8), 10–16.
- Dabla-Norris, E. Kochkar, K., Suphaphihat, N., Ricka, F., & Tsolunta, E. (2015). *Causes and consequences of income inequality: A global perspective*. Washington, DC: International Monetary Fund.
- Jarvie, I., & Pralong, S. (1999). Popper's open society after fifty years: The continuing relevance of Karl Popper. London & New York: Routledge.
- Kanie, N., & Haas, P. M. (Eds.) (2004). *Emerging forces in environmental governance*. Tokyo: United Nations University Press.
- Kelly, R. M., & Gregware, P. (1997). The experimenting society: Toward an inclusive democratic community. In W. N. Dunn, Editor, *The Experimenting Society: Essays in Honour of Donald T. Campbell*, *Policy Studies Review Annual*, Volume 11, Transaction Publishers, New Brunswick.
- Lester, J.C., (1995). Popper's epistemology versus Popper's politics: A libertarian viewpoint. *Journal of Social and Evolutionary Systems*, 18(1), 87 - 93
- MacDonald, B. (1976). Evaluation and the Control of Education. In D.A. Tawney (Ed.). *Curriculum evaluation today: Trends and implications, schools and council research studies* (pp. 125 – 136). London: Macmillan.
- Mertens, D.M., & Wilson, A.T. (2012). *Program evaluation theory and practice: A*

- comprehensive guide. New York: Guildford Press.
- Nielsen, R.K., & Graves, L. (2017). News you don't believe: Audience perspectives on fake news. Reuters Institute, and University of Oxford Factsheet. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Nielsen%26Graves_factsheet_1710v3_FIN_AL_download.pdf.
- Oxfam (2017). *Just 8 men own same wealth as half the world*. Retrieved from <https://www.oxfam.org/en/pressroom/pressreleases/2017-01-16/just-8-men-own-same-wealth-half-world>.
- Patton, M.Q. (2008). *Utilization focused evaluation* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Patton, M.Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: The Guilford Press.
- Patton, M.Q., (2016), In search of blue marble evaluators. *Better Evaluation*. Retrieved from http://www.betterevaluation.org/en/blog/blue_marble_evaluators.
- Picciotto, R. (2011). The logic of evaluation professionalism. *Evaluation*, 17(2), 165-180.
- Picciotto, R. (2013). The logic of development effectiveness: Is it time for the broader evaluation community to take notice? *Evaluation*, 19(2), 155-170.
- Popper, K. (1944). Piecemeal social engineering. In D. Miller (Ed.) *Popper Selections*. Princeton, N.J.: Princeton University Press.
- Popper, K., Ryan, A., & Gombrich, E. H. (2013). *The open society and its enemies*. Princeton and Oxford: Princeton University Press.
- Popper, K. (1988). From the archives: The open society and its enemies revisited. *The Economist*. Retrieved from <http://www.economist.com/blogs/democracyinamerica/2016/01/karl-popper-democracy>.
- Popper, K., (1994). Science, problems, and responsibilities. In M.A. Notturmo (Ed.). *The myth of the framework*. London: Routledge.
- Raustiala, K. (2002), *The architecture of international cooperation*. Los Angeles: UCLA Law School. Retrieved from https://www.law.berkeley.edu/files/raustiala_ArchitectureofInternational.doc.
- Rawls, J. (1971). *A theory of justice*. Cambridge, Massachusetts: Harvard University Press.
- Rieper, O. (2004). Exploring Popper's relevance for the evaluation community. *Evaluation*, 10(1), 92 - 100.
- Schilpp, P.A. (ed.) (1974). *The philosophy of Karl Popper, Part 1*. The Library of Living Philosophers (Vol XIV). Lassalle, Illinois: Open Court.
- Scriven, M. (1991). *Evaluation thesaurus*. London: Sage Publications.
- Scriven, M. (1998). Minimalist theory: The least theory that evaluation practice requires. *American Journal of Evaluation*, 19(1); 57 - 70.
- Scriven, M. (2016). The last frontier of evaluation: Ethics. In S. I. Donaldson & R. Picciotto. *Evaluation for an equitable society*. Charlotte, N.C.: Information Age Publishing
- Sen, K., Pritchett, L., Kar, S., & Raihan, S. (2017). *Democracy vs. dictatorships? The political determinants of growth episodes*. Faculty research working paper, RWP 17-009, Harvard Kennedy School.
- Seneca, L.A. (2016). *Seneca's letters from a stoic*. Mineola, New York: Dover Publications.
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1 - 19.
- Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R., & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluation: Report of a study commissioned by the department of international development*. Working Paper 38. London: Department for International Development.
- Stiglitz, J. E. (2002). *Globalization and its discontents*. New York: W.W. Norton
- Vedung, E. (2010) Four waves of evaluation diffusion. *Evaluation*, 16(3), 263 - 277.
- Weiss, C. H. (1998). *Evaluation: Methods for studying programs and policies* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Williams, B., & Imam, I. (Eds.) (2007). *Systems concepts in evaluation: An expert anthology*. Point Reyes, CA: EdgePress of Inverness.