

# ULTIMATE EVALUATION QUESTIONS<sup>1</sup>

Michael Scriven

Claremont Graduate University

**0. Introduction.** The following are some thoughts about expanding the working area of the discipline of general-purpose evaluation. They do not suggest that we should drop our current specialized work in, for example, program evaluation, only that there are some things we should add to our concerns—background duties, one might say.

**1. Ultimate sustainability.** Checklists for evaluating programs, organizations, and educational systems usually and correctly list sustainability as one of the key criteria of merit. There is always an explicit or implied time and space window within which sustainability is to be calculated, but the profession should now be putting on some peer pressure to make sure that framework is the appropriate. In particular, given that we live on a planet with limited resources many of which are rapidly approaching exhaustion, and a rapidly expanding population, many large-scale projects

---

<sup>1</sup> TYPOGRAPHICAL NOTE: A double exclamation point at the beginning of a sentence or paragraph indicates that it will contain a recommendation. Vast areas of evaluation, e.g., all of the scholarly discipline of history, contain and should contain no recommendations. But some areas, e.g., program and policy evaluation, contain fewer than I think they should, and this note is intended to encourage more thinking about that possibility.

are not sustainable for more than one average human lifetime, although planned for several; for example, many large dam projects, and some big plans for school reform including curricular reform in social studies that does not include planning for the future of life on Earth. Such discrepancies should be made clear, and recommendations may be required for improving the consistency between plans and real possibilities. Well, not much new about all this so far, but applying it needs some action, for example,...

There has been a great deal of political pressure to demonize the conclusions of the pessimists about our future survival on this planet (e.g., the forty years of mostly invalid attacks on *The Limits to Growth*),<sup>2</sup> but the case for its conclusions has continued to build as more data accumulates. Evaluations that fail to look at *really long-term* sustainability, when that's the relevant framework, are part of the problem of public avoidance of confrontation with reality when immediate citizen action is urgent. Despite the devastation of recent hurricanes in the north-eastern and south-eastern USA, few coastal communities elsewhere have budgeted (e.g., .25% on the sales tax for non-food items) and changed building codes to prepare for local replications. !!This aspect of evaluation—checking for

---

<sup>2</sup> For example, it is often suggested that we should not be concerned about population growth since the global population looks as if it might stabilize by the end of this century. Even if this is true, it's irrelevant: a planetary population doesn't survive by stabilizing, it just starves slower. !!See Richard Hake (Google) for a review of the expert literature.

the evaluand's contribution or opposition to long-term survival systems for the planet (or parts of it like those near sea level)—should surely be listed in our various codes of professional ethics and competencies. (See also Section 3 below.)

**2. Ultimate crowd-sourced evaluation (CSE).** It is clear that the most common CSEs—the kind you find attached to almost everything Amazon or Zagat covers—are subject to serious abuse, corruption, and superficiality. This is no doubt why CSE has not been given much serious attention by evaluators, especially product evaluators on whose turf it is played out. But I am now convinced that this is like condemning statistics because it only produces approximations. CSEs can be treated as a valuable—indeed often invaluable—source of information if you follow a few guiding principles.<sup>3</sup> So this is another boundary to evaluation that is falling, and it's one whose removal, like the staggering success of Wikipedia,

---

<sup>3</sup> !!Here are some, of which the last is perhaps the most important: (i) ignore cases of favorable ratings when  $n < 5$  always, and  $n < 10$  if possible, and treat  $n > 50$  as worth twice the ratings where  $n \leq 50$ ; (ii) always get CSE ratings from two sources if possible—AMZ and Consumer Reports (CR) if possible; (iii) get a CSE as well as a non-CSE rating whenever possible, even when you can get CR and America's Test Kitchen as the non-CSE (toasters are a good example where both the latter fail); (iv) ignore 'bare' ratings if possible, where no non-trivial reasons are given (that's the kind that most often are faked); (v) focus on the low ratings, not the high ones: always read at least half a dozen of the one-star and two-star (out of five) ratings. For example, for the top of the line current Samsung TVs, the CSE critics identify a critical flaw that none of the professional evaluators had noticed, because it doesn't show up until you've run the set longer than the pros kept each model.

has facilitated the democratization of evaluation, an important part of increasing the feasibility of participatory democracy..

If you run into someone who dismisses attempts to retrieve good evaluation results from the CSE tailings in which they can be found, you might get luckier by stressing the following point: in this country we always live our lives in a way that is massively determined by one particular type of CSE. It is the American Presidential Election. The evaluators who study it, and their predecessors over the life of the country, are called political analysts, and they are not alone in considering that set of CSEs to be evaluands worthy of study since a democracy is supposed to be governed in their interest. And the best of the evaluators *beat all the pollsters* in the 2012 election, beat them by a mile. In other words, expertise in determining the nation's view on controversial issues is not only possible in principle but exists in practice and can be taught, so building policy on it can be done with validity and the help of some evaluators. !!Assignment: evaluate one of the policy proposals of a recent candidate for office, using a checklist designed for evaluating programs.

**3. The ultimate place of evaluation in the cosmos.** Astrobiologists argue that because there are so many planets in the universe—billions by the current best estimate (Wikipedia)—it is very likely that there are other intelligent beings 'out there.' But there is another possibility. The human species time on Earth has so far been around for only an eye-blink

in cosmological time, or even in the life of our Earth. Suppose that many other species on other planets have evolved a comparable degree of intelligence; suppose that this happened on every tenth habitable planet. Suppose that all these species did develop atomic and hydrogen bombs, and did develop a planetary population that ate and gassed itself into a non-sustainable global situation. At that point, they all have qualified for, and now must show they are smart enough to handle the Ultimate Survival Problem. It's the ultimate graduation test for intelligent species, and it's a very tough test because almost all the traits that led them to develop buildings, tools, science, logic, ethics, and art—aggressiveness, perseverance, imagination—also drive them towards war and the disregard of other obvious barriers to progress.

So it's quite plausible to think they have all failed the ultimate test—thus far. That would explain why the search for ET signals has not produced any results: there *were* ETs but there aren't any *now*. Of course, you can't *reliably* reverse that reasoning—you can't argue that because there are no ET signals, there are no ETs at the moment. There might still be many, perhaps millions, perhaps mostly smarter than us; and their winning strategy may include resolute isolationism, argued for by their best evaluators (in this case, policy analysts), because their study of their predecessors revealed that those who solved their own 'ultimate problem' were then wiped out by savages they contacted via an ET search. Of course, you can't be sure that's the case, either; but you *can* be sure it's a

possibility, and smart planning covers as many of the possibilities as possible.

So who's doing this planning for First Contact? It's a job for global policy analysts, if there are any buried in the corridors of the United Nations. And they are a group of evaluators who probably could use some help in planning First Contact because there aren't very many of them. !!But even if they don't want our help, we'd better start a little more explicit thinking about this problem because our lives or our descendants' lives are at stake, and we have a Disaster TIG (although more kindly named) in AEA, so we're not starting from zero. There are some great predecessors on this path; famously, Enrico Fermi, who brought up what has become known as the 'Fermi paradox' in 1950, with the question "Where is everybody?" !!But not every schooled child knows about it, not even every evaluator, and evaluators working on public education, particularly but not only curriculum evaluators, should change that. (See Richard Hake's extensive writings on the required changes in the school curriculum: references in Google.)

These changes are far-reaching. They must surely include some study of how to reshape our own habits and values that are moving us nearer to planetary collapse through war, famine, flood, or pestilence (e.g., Aristotle's insight that we can change our values by changing our friends, since they strongly influence our values). But much other knowledge and

skills that help with doing that that are still not in our basic curriculum e.g., game and decision theory, evaluation, global problems and proposed solutions, and critical thinking throughout. !!Bottom line; it is time for us to expand our horizons so as to include, whenever relevant, the evaluation of global survival strategies, and programs that are precursors or infrastructure for them, both for this century and those that follow. That's a pretty good New Year's Resolution.

**4. The ultimate value in evaluation.** Almost every evaluation involves combining some value or values with some non-evaluative data. When those values conflict (e.g., cost vs. deluxe features), some must be ranked over others in order to get the required evaluative conclusion. What is the top-ranked value of the 16 or so most important ones<sup>4</sup>—the ultimate consideration in the 'weighting game'? Of course, it's ethics, or at least we pay lip service to that view. But whose ethics? Some multi-culturalists say "everyone's" but that would mean granting cannibals and wife-stoners immunity from condemnation, a view that few clients and evaluators—and governments—are willing to accept. Now, competent professionals don't use values they can't validate—so how do you validate this ultimate value in the ranking of values? You can't appeal to higher values to justify it, because there aren't any. Your usual route in basic validation of values is via a needs assessment; but can you prove that we need ethics? That's

---

<sup>4</sup> See the Values checkpoint in the Key Evaluation Checklist on my website [michaelscriven.info](http://michaelscriven.info)

what the utilitarians tried to do, by arguing that ethics maximizes happiness—and we all need, or at least strongly prefer, happiness—but the usual forms of utilitarianism fail.

In an attempt to solve this problem I use game theory to show that a novel form of utilitarianism *can* be supported as the optimal survival strategy by a related proof.<sup>5</sup> That makes ethics a branch of evaluation, which is a discipline more far-reaching than, but including, the essential foundations of science and scientific method. And so it raises the possibility that those arguing for another basis for ethics, one whose premises cannot be validated via scientific and logically sound arguments, are subversive, i.e., undermining the foundations of our way of life and our best chances of survival. This group currently includes adherents to all major Western religions. !!Assignment: assess the arguments for banning (i) theists, and (ii) agnostics from government employment. Not exactly a good Christmas present, so... Happy New Year!

---

<sup>5</sup> I replace happiness with 'all consistent values'; I avoid claiming that altruism is the best strategy for extremely powerful and evil dictators (but I do claim it's a duty to help bring them down); and I expand the timeline for the calculation of benefits from immediate to within one generation. I retain the definition of ethics as the system of rules and attitudes based on the proposition that all people have prima facie equal rights (a.k.a., golden rule, love thy neighbors as thyself, etc.) A fuller but earlier form of this argument is in my *Primary Philosophy* (McGraw-Hill, 1966)); it's a time-spread solution to the Prisoner's Dilemma.