

An overview of evaluation theories

Michael Scriven



Evaluators play many roles, and, like actors, sometimes fall into the trap of thinking that those roles define reality. There are half a dozen cases in the history of evaluation where it seems to me that this has happened.

- 1 In the early days of program evaluation theory, managers or funders were commissioning evaluations, to assist them in their role as decision-makers, and that led some evaluators to think that it was the essential nature of evaluation to be a *decision support* process. And the decisions to be supported were the typical manager/funder decisions about whether the program should be refunded/cut/abandoned. These were controlled by the managerial definition of success – i.e., was the program on time, on target, and on budget – so those were the data to be gathered. Indeed, these data could be used for both formative and summative purposes. Discrepancy evaluation (Provus) was the most outspoken advocate of this approach, but several others, e.g., those of CIPP and Suppes/Cronbach, supported the general approach. In truth, however, this kind of decision support is only one type of decision support (see 2 below), and decision support in general is only one of evaluation's roles (see 3 to 6 below).
- 2 Later, as the importance of evaluation as a *consumer service*, typically summative, was stressed, some theorists (myself amongst them) began to talk as if this was the essential duty of evaluation. Consumers, by and large, have no interest in whether the program or product designer's goals have been met, and only a secondary interest in improving the program (i.e., formative evaluation), being mainly interested in whether their own needs are met (a different definition of 'success'). However, it was incorrect to act as if the information needs of program funders and managers were illegitimate, though it was important to stress that they were often not the same as those of consumers.
- 3 Then the pendulum swung full arc, and we had the Cronbach group trying to convince us that there was no such thing in the real world as summative evaluation; in reality, they said, evaluation was essentially always formative. This may well have been true of their experiences, but not of a reality containing *Consumer Reports* and the results of jury trials and appeals.
- 4 More recently, the idea of allowing those who are being evaluated to participate in the evaluation has become increasingly popular, for ethical or political reasons; this has been called *collaborative* or *participatory* or *empowerment evaluation*; transactional evaluation was a precursor. It has been suggested by advocates of these positions that the old models are

outdated, and that evaluation, properly done, should always be seen as a collaborative or even an empowerment exercise. Sometimes yes, often no; there are often requirements of confidentiality or credibility or cost or time that require distancing of the evaluator from the evaluated.

- 5 Even more recently, the approach of *theory-driven evaluation* has taken the process of generating explanations of the success and failure of an enterprise to be the core function of evaluation, and concluded that this should be the core of the conception itself. But black-box evaluation, or something close to it, is not a failure of evaluation, as the history of the evaluation of many processes in medicine and education bears witness. It is just one kind of evaluation, often the only kind we can do, and very useful at that.

A. Running alongside these later conceptions of evaluation has been a rowdy gang of ontological subversives – the doctrines associated with epistemological scepticism, constructivism, post-modernism, and some fellow travellers. In these views, the entire enterprise of evaluation, like most others of what they describe as ‘so-called objective’ science, is a fraud, being in reality only a projection of some group’s values onto the subject matter. Since a general doctrine that denies the existence of reality can hardly at the same time claim to be informing us of the real truth, this position is inescapably self-refuting as a universal truth; if, on the other hand, it is making some less general claim, we can proceed until our specific errors are demonstrated. But those impressed by or pressing sub-theme A rarely do this (Mertens is an exception), and hence do not pose the threat they imagine to be their due. Nevertheless, they have influenced or included not only the supporters of positions 4 and 5, but also of our final example of aspect-exaggeration, Model 6. Bob Stake is perhaps the most eminent evaluator attracted by this view, especially in the form in which it simply denies the possibility of objectivity. Bias is always present, he thinks, so don’t try to hide it. But evaluation bias can be prevented; look at the scoring guide to a well-developed mathematics test as an obvious example, or the rating of Einstein as a brilliant scientist, or the evaluation of certain reading programs as unjustifiable expenses (when others are better and cheaper in the relevant respects). The denial of objectivity is either self-refuting (if a universal claim) or something that has to be documented in specific cases and applied only to them.

- 6 Our last example concerns the power of evaluators. On this view, often fuelled by what are seen (I think incorrectly) as the implications of A, it is inevitable that evaluation exerts social power, and that any serious approach to it must begin with this fact, and proceed to maximise the social value of its impact. Here we have Ernie House’s *democratic deliberative evaluation*, and the *transformative evaluation* of many theorists, well summarized by Mertens – who introduced the term – in several recent works. Here, too, we must place, albeit at one remove from activism, those who define evaluation as a process aimed at the solution of social problems; Rossi, Freeman, and Lipsey’s dominant text is now joined in this view by the latest one, by Mark, Henry, and Julnes. But evaluation, even program evaluation, should not be defined as something committed to social engineering. One might as well define a tape measure as something committed to house-building. That’s just one use for it, often not present, as in bridge-building. Evaluation is often concerned only with determining the extent of something, e.g. in computer literacy or epidemiological surveys, and hence a tie to social problem solving can never be defined as part of its nature.

The listed set of six positions is, I think, one more example of the argument between members of a blindfold committee investigating the nature of an elephant. The reality is both simpler and more complex than their individual conclusions: simpler at the meta-level, more complex in detail. At the meta-level, the truth of the matter – if we may be so bold as to suggest that there is such a thing – is surely that these positions all represent an exaggeration of one important role of evaluation into an oversimplified conception of its entire nature. The complex detail, the hard work, in evaluation theory, involves unpacking the way in which evaluation is a pervasive multi-function, multi-role, multi-player enterprise: context-dependent here, context-independent there, biased here, objective there. It is part of the great knowledge-seeking effort that includes substantial parts of science, technology, law, ethics, and other humanistic disciplines. And all of us, throughout all our personal and professional lives, play many roles in evaluation – as evaluators, as evaluatees, and as commissioners or users of evaluation.

Why has this view not been more commonly accepted? Apart from the tendency to over-generalize from a limited range of experiences (large in number, but of the same species), and the perpetual attraction of simplification, there is an affective pressure towards a monolithic conception. Many of our roles in evaluation are involuntary, and many, not just those same ones, are painful. Hence it is very attractive to persuade oneself and others that evaluation has either a more controlled and less painful (or pain-threatening) role – or a disreputable one, as the ‘value-free’ movement averred. But these are bad reasons for poor theories.

A reason to which one can be more sympathetic is that the language of evaluation is often and usefully contrasted with the language of description, so it seems wrong to say that evaluation is a descriptive

enterprise. However, the appearance here is entirely context-dependent: there are indeed many contexts in which evaluation requires a long-chain inference from the observations or other data, so the distinction makes good sense. But there are many other contexts in which we say that a person can be described as good-looking, intelligent, etc.; a program as the best available; an essay as excellent; a dive as graceful or polished, etc. So there is no intrinsic logical difference between description and evaluation; the nearest intrinsic difference is that between the involvement of evaluative terminology and its absence, and such terminology can sometimes be supported by direct observation and sometimes requires complex inference. (Incidentally, it may or may not involve judgement, contrary to the common view that evaluation is of its nature judgemental: sometimes, as in the evaluation of answers to questions on an easy mathematics or translation test, the merit of the answer is just a matter of observation.)

At the most basic level, evaluation is a survival-developed brain function of supreme importance. It is, in its simplest forms, where it is close to the perceptual level, the near-instinctive process that connects knowledge with action. It is probably not justified to say that the totally instinctive connection which prey species make between birds of the hawk shape and flight to hiding is evaluation in the sense in which evaluators do it, which involves at least one dip into the knowledge base or the inference process, but it is the precursor of evaluation or at least a different kind of evaluation, a mere valencing. One step up, and we're at primitive forms of real evaluation, which leads us, on some occasions, to flee from the mammoth, and on others to attack it.

So some kinds of evaluation are in the genes – including some part of our evaluative response to the bodies and behaviours of others – but much of it is learnt. In its most complex form, it elevates the simple process of instant or near-instant appraisal into the complex and systematic investigation of complex systems, in this process becoming, of

necessity, a complex discipline itself. It remains one of our key survival mechanisms, whether in the evaluation of complex weapons systems for our armed forces, or in our civil efforts to improve the rewards and punishments attendant on conviction of crime; but it has also become a key part of the life of the mind, ranging from the satisfaction of curiosity (How good is X, compared to Y, and absolutely?), through science (What's the best variation of string theory?) to our aesthetics (Is this style of painting a revolution – or a failure?).

Evaluation seeks only to determine and describe certain aspects of the world in terms of the vocabulary of value, i.e. in terms of either merit, worth, or significance. That's the dictionary

At the most basic level, evaluation is a survival-developed brain function of supreme importance.

definition, and there's no good reason to abandon it. Of course that value is often dependent on context; but sometimes mainly on more or less acontextual considerations such as deductive logic, or on more or less transcendent ethical standards such as the evils of murder, torture, and child/spousal/worker abuse. Many things are legitimated, precipitated, or barred by such descriptions -both good and bad things one might say. But those things are not part of the evaluation, just consequences, in the social context of the time. Evaluation itself is a strict discipline, fully comparable in its validity with any part of science. That is the core of the matter, and one should try to separate the core issues from the contextual ones. Otherwise, one alters the evaluation in order to improve its consequences, thereby violating the canons of professional conduct. As the early utilitarians learnt long ago, ethics – professional standards being one branch of ethics – transcends first order consequences; and ethics is of course just another branch of evaluation. ♦

EVALUATION OF A DIFFERENT KIND

A jongleur endeavours to entertain the conference dinner at the 2001 AES Conference. Scott Bayley, Judy Pearce and Anton Muller seem less than convinced.

Brennon Dowrick, Olympic gymnast, entertained delegates with word and deed at lunch on the first day of the conference.

