

A Critical-Historical Review of Program Evaluation and the Emerging Motif 'Evaluation Science'

William James Fear
Owner, William James Consulting

Background: It is important to distinguish between evaluation as an inherent, automatic, affective process and Program Evaluation (Evaluation, with capitalised 'E') as an institution, and equally important to consider what a good understanding of evaluation tells us about Evaluation. Evaluation is an established social institution whose modern roots can be traced back to 16th century France. Since the early 1900s the institution has developed within and across a range of scientific disciplines with interests in perceived social problems and efforts to resolve the said problems. This can be demonstrated objectively by the number and scale of relevant publications within relevant disciplines. This, in turn, helps us understand more about Evaluation as an institution. Set in this context is the question of Evaluation Science: is this simply a fashionable institutional motif or is it a potential new era for Evaluation?

Purpose: Commentary on the history and development of Program Evaluation.

Setting: Not applicable.

Intervention: Not applicable.

Research design: Not applicable.

Data collection & analysis: Not applicable.

Findings: Not applicable.

Keywords: *program evaluation; evaluation; history; evaluation science.*

Introduction

Program Evaluation (Evaluation, hereafter) is an institution. That is, in the simplest possible terms, it is an established pattern of behaviour, including thinking, feeling, and acting, with an accompanying set of beliefs and values (DiMaggio & Powell, 1991; Gerth & Mills, 1953; Judd, 1926, 1936; Scott, 1995, 2004). By Evaluation I mean a value-based judgement of the merit and/or worth and/or significance of an intervention or program stemming from a policy of some sort or another. The intervention or program is intended to bring about a beneficial change of some sort for a group of people, the evaluand. Evaluation is not bound to any one particular discipline, profession, or field of practice and in this sense it is both somewhat unusual and permeable. Like all institutions it is highly susceptible to fashions and trends. One of the current fashions is the attempt to rebrand Evaluation as Evaluation Science. A number of recent discussions about the notion of Evaluation Science led me to review my long-term interest in the historical development, or evolution, of Evaluation into its current form.

Attempting to address the historical development of an institution is problematic from within the confines of the institution, not least due to the typical short-term focus on recent history and the requirement to adhere to the current beliefs and values of the institution. The Evaluation institution, for example, typically assumes Evaluation emerged in the United States between 1933-1936 and that Evaluation is transdisciplinary. Neither, I argue here, is the case. For clarification, evaluation per se is clearly transdisciplinary. That is, the act or process of evaluating is part and parcel of any discipline and, indeed, is an integral process that has been studied in some depth. In this sense evaluation is not a discipline in and of itself and neither is it an institution as it occurs regardless of either of these constructs. To be sure, it can be exercised within and between disciplines and institutions as an artefact, but it does not depend on either. Program Evaluation, on the other hand, is the use of evaluation as an artefact to make a judgement about an object where the object is a program. As such, Evaluation can be an institution, and

arguably is, given its long history and prevalence. Whether or not Evaluation can be a discipline in its own right is a separate matter.

I suggest, here, that the assumption that an already existing institution can become a distinct self-affirmed discipline that owns a process common to all is fundamentally flawed at best. That is not to say we cannot have a self-affirmed discipline whose object of inquiry is the process of Evaluation. But I suggest there are some difficulties with wider support for that notion given that existing disciplines are already heavily engaged with this activity and have been for a long period of time. Education, for example, has a long history of Evaluating their own Evaluation methods and processes, as does Economics as does Psychology (this is referred to below).

Regarding the question of Evaluation Science, this, again, is neither a current development nor one either limited to, or owned by, Evaluation. It is an emerging fashionable institutional motif already established, at the time of writing, outside of Evaluation. That does not mean it has no use. It does mean that, as with all fashionable motifs, caution should be exercised with regard to adjusting institutional beliefs and values to accommodate such an object.

This paper starts by addressing the question of the very nature of evaluation, and follows this with an exploration of the role of evaluation in science. It then addresses the question, 'What is [Program] Evaluation?', and considers a plausible long-term history of the institution. Next it provides a disciplinary perspective of the recent history of the institution followed by a critique of the concept of Evaluation Science. It concludes with some consideration of how Evaluation might continue to develop.

As noted above, the developmental history of an institution is difficult to conduct from within the boundaries of the institution, no matter how permeable those boundaries. This paper makes extensive use of texts from outside of Evaluation as objects. This may be uncomfortable for the institution not least as it may challenge current beliefs and value positions endorsed within the institution's proprietary literature. Nonetheless, the very nature of Evaluation as an institution makes

it important that it rise to just such a challenge.

What is Evaluation?

Without the need to be prosaic, any evaluation has, at its core, a value judgement. That is, evaluation is the process of valuing, judging, or assessing, the inherent and/or pragmatic worth of an object according to set of values ('worth' can of course be segmented further into 'merit' and 'significance' or other sets of elements) and endowing it with meaning, where meaning is dependent upon values and is itself a process of valuing. Any object (and a process can be an object) can be evaluated and it will be evaluated in relation to the set of values employed.

A value judgement, an evaluation, is an affective process that is to some extent automatic, although to what extent varies according to other available mechanisms. The process is deeply rooted in emotions and, by default, in the beliefs the individual and/or group are most invested in. The degree to which this process can be managed, controlled if you prefer, is open to question and seems determined by attitude towards the object, attitude being equally value based. (This is explored in depth in *The Psychology of Evaluation*; Musch & Klauer, 2008).

Automatic evaluations can be spontaneously and immediately constructed without previous experience and conscious appraisal of the object (Ferguson & Bargh, 2008) and yet high fluency with the object is associated with more favourable evaluations of the object. (Winkielman et. al., 2008). Simply articulating, even non-verbally, how one feels about an object influences subsequent decisions (e.g., see White, Barqué-Duran, & Pothos, 2016).

The relationship between emotion, values, and evaluation has been further explored by Fingerhut and Prinz (2018). They argue that evaluative concepts qualify as abstract and are based on emotions linked to bodily perceptions. They extend this understanding of evaluation into morality, including justice, and aesthetics. Once again we see the argument that evaluation is an affective process that is largely automatic and which interacts bi-directionally with sensory

perception—even in relation to morality based constructs such as justice.

These arguments can be readily extrapolated to any deliberate process of valuing, or judgement and assessment, and are important in relation to Evaluation. In simple terms, when a program is Evaluated there will be an impact of fluency with the program, attitude towards the program, and the set of beliefs and values held by evaluator/s and how those values and beliefs relate to those of the various groups of stakeholders. The extent to which methodological tools and processes can constrain these forces is an important question.

We should also take note that there are no noumenal values. That is, there are no absolute and unequivocal values that have an independent existence in and of themselves. Notions of value can be applied to any object, the value of 'good' being a particularly pertinent example. A bad thing can be a 'good' bad thing in that it is a good, or ideal, example of that bad thing. To suggest there is an absolute good is to completely ignore all relevant context and knowledge on the one hand, and to completely misunderstand the value of good on the other hand. (Issues of the various forms of social good—public good, common good, global good, and greater good—are discussed elsewhere. The point here is simply that there is no known form of noumenal good; any interpretation of good is relative in a number of different and varying relations.)

What we do when we assign meaning, in the simplest possible terms, is we categorise something as good (beneficial), bad (detrimental), or indifferent (of no immediate affective consequence). Our categories of good, bad, and indifferent are formed, in large part, based on experience of an object and our affective response to that experience. Our experience of an object is always, to some extent, automatic and based on prior experience/s, but can be moderated by fluency and attitude and, probably, by any artefacts we use to engage with the object. Evaluation then, as a natural process in its base form, is simply the attribution of value and meaning to emotion in relation to physical perception of, and sensory response to, an object.

The Role of Evaluation in Relation to Science

At its root the fundamental purpose of science, and indeed all other disciplines, is to isolate objects/phenomena and describe and explain them and their perceived significance within a clearly stated conceptual framework. In order to describe and explain objects science relies on critical evaluation (Kantor, 1919). Critical evaluation employs both a set of pre-defined and clearly articulated values and a set of processes, methods, designed to constrain the influence of values and value sets that are not directly related to valuing the object.

What science does is to endow objects with meaning in accordance with a clearly stated conceptual framework using the process of critical evaluation mediated by agreed values and methods. Science seeks to constrain the affective influence in evaluation and ensure that fluency with the object is managed within a shared set of boundaries. Attitudes and values are predetermined, consistent, and known. The ongoing critical evaluation of the object builds knowledge constructs which represent conditions, phenomena, and objects (bearing in mind that anything can be treated as an object and these distinctions are simply accepted and agreed terms for the sake of simplification and clarity).

Furthermore, science is very much concerned with change. It is concerned with how objects interact and bring about change, about how objects can be changed through deliberate intervention, about how objects change over time of their own accord, and so on.

Evaluation, more properly critical evaluation, is undoubtedly the primary artefact of science and is well recognised as such. The scientific methodology is one that serves to constrain and manage the impact of the automatic process of evaluation and bring it under some degree of control.

A range of specialist knowledges, methods, constructs, objects and values are brought into play and values, meanings, and objects vary within and between disciplines. In this sense we can consider evaluation, but not Evaluation, transdisciplinary in that all

disciplines, not only the sciences, make use of evaluation.

With regard to Evaluation, there is a good reason to consider Evaluation scientific, especially as social change is driven by, and constructed on, values which are by default grounded in emotion and strongly held beliefs. When we look more closely at the history of Evaluation we shall see that this is exactly the effort made over time: to ensure that Evaluation is scientific (i.e. makes use of the methodologies employed by the sciences. Note, these methodologies do not exclude qualitative methods.)

With further regard to science's understanding of the role of evaluation in science, Zuckerman and Merton (1971) noted that while science depends on status judges engaged in an evaluation process that process is not exempt itself from being evaluated.

Evaluation is Ubiquitous in the Sciences

As a simple demonstration of the ubiquity of evaluation in a very wide range of sciences and related disciplines I conducted a search of a small number of journals, collections of journals, and publishing houses for the term 'evaluation'. Searching collections of journals and publishers is informative as it provides insight not only into the range of publications and articles but also the range of disciplines knowingly engaged with evaluation (to include Evaluation). (Searches such as this clearly produce a number of false positives. These will not affect the overall picture as the risk of false positives is consistent across all searches.)

The selection of journals and publishing houses was purposive and value-based and intended to give a simple snapshot of the range of primarily scientific journals and scientific journal publications concerned with evaluation in the sciences and related fields and disciplines relevant to policy and programs. The results are presented in Table One along with a set of accompanying notes.

What we can see from Table 1 and the Notes is: (1) the enormous number of articles published that refer to evaluation (4.8 million from this search alone); and (2) the enormous number of evaluation-specific articles

published and the wide range of objects subject to scientific evaluation (229,772 articles with evaluation in the title; 5% of articles in the related searches). Furthermore, a large number of these evaluations are concerned with social change.

Searching Google Scholar for publications containing 'evaluation' anywhere gives 6,220,00 results with 292,00 added in the last year (2019). Limiting 'evaluation' to the title

gives 980,000 results with 88,000 added in the last year (2019). However, limiting the search to articles containing 'program evaluation' anywhere gives only 757,00 results. When this is limited to the title it gives only 13,800 results. It is feasible that this latter set could be analysed for the range of journals to reflect the number and range of disciplines overtly and knowingly engaged with Evaluation as an object.

Table 1
Results of Searches for 'Evaluation' Across Select Number of Journals, Collections of Journals, and Publishers and Accompanying Notes

Journal/Publisher	'Evaluation' as Generic Search Term	'Evaluation' in Title	Comments
The Annual Review series (Group of journals)	18,959	127	
<i>Public Administration</i>	1449	33	
<i>The International Journal of Public Administration</i>	730,725	73	
<i>The Journal of the American Medical Association</i> (Group of Journals/Publications)	76,343	5,385	
Academy of Management journals (Group of journals)	11,639	430	
<i>Administrative Science Quarterly</i>	507	9	
<i>Nature</i> (Group of journals)	124,505	3,746	
<i>Philosophical Transactions of the Royal Society of London</i> (Group of journals)	7,497	243	
American Physical Society (Group of journals)	44,286	853	
American Association for Public Opinion Research (Group of journals)	2,878	Not searched	
American Chemical Society (Group of journals)	347,064	10,686	
Hindawi (Group of Open Access journals)	112,132	35,711	
Directory of Open Access Journals (Group of Open Access journals)	202,400	69,160	
Springer (Publisher)	3,136,112	101, 935	
Elsevier Journals (Publisher)	250 journals (not articles)	Not searched	See Note 1
Taylor and Francis (Publisher)	1,664,072	Not searched	See Note 2
Sage (Publisher)	3,258	Not searched	See Note 3

Note 1. Examples of the journals and the sort of evaluations they publish include: *Research in Social and Administrative Pharmacy*: 'Topics of interest include outcomes evaluation of drug products, programs, or services...'; *Nuclear Data Sheets*: '...devoted to compilation and evaluations of experimental and theoretical results in Nuclear Physics...';

International Journal of Child-Computer Interaction : ‘...New methods for working with children in design, evaluation and research...’; *Transportation Research Part A: Policy and Practice* : ‘...design, management and evaluation of transportation systems...’; *Artificial Intelligence* : ‘...describe a principled solution, emphasize its novelty, and present an in-depth evaluation of...’.

Note 2. A scan of the titles of the journal articles and where they are published gives a good idea of both: External Government Performance Evaluation in China: Evaluating the Evaluations, in *Public Performance Management Review*; Evaluation, or Just Data Collection? An Exploration of the Evaluation Practice of Selected UK Environmental Educators, in *The Journal of Environmental Education*; How well are aid agencies evaluating programs? An assessment of the quality of global health evaluations, in *Journal of Developmental Effectiveness*; Evaluating publicly supported periodic events: the design of credible, usable and effective evaluation, in *Journal of Policy Research in Tourism, Leisure and Events*.

Note 3. This range includes textbooks, professional books, and journals all concerned, primarily, with the Social Sciences.

What is Program Evaluation?

Program Evaluation is the value-based assessment, and judgement, of some sort of deliberate social intervention, including organizational interventions, designed to bring about some sort of intended social change. The use of the term ‘some sort’ indicates both the complexity and confusion typically, one might argue inevitably, associated with these interventions. In that sense Evaluation is very much based in the Social Sciences and similar disciplines as it is concerned with social problems. (The object of Social Science is perceived social problems, however and wherever these arise.)

Evaluation is not a social change movement. It is a sort of amorphous discipline or a professional practice whose object of inquiry is interventions intended to bring about social change. When Evaluation assumes to directly and intentionally influence social change it becomes both a program in its own right and a political/activist movement and can and should be fully and critically evaluated as such with named bodies and persons assuming responsibility for the outcomes. While such a movement can exist within the field of Evaluation such a movement cannot assume to speak for the entire field without the approval and endorsement of the entire field.

As we have already seen, science requires the use of critical evaluation, and arguably this is true also for Evaluation (although this can be contested). Assuming, however, that Evaluation relies on critical evaluation there must be a set of clearly articulated values and a set of clearly articulated methods that both serve to constrain and bound the conclusions of the process of an enacted Evaluation.

There is a strong suggestion that there are, indeed, a distinct set of values about ‘good’ that are central to Evaluation in relation to the object. (Broader understandings of ‘social good’ currently assumed by some sectors of the field of Evaluation form a separate and distinct discussion and this is addressed elsewhere.) These values have been articulated historically as Effectiveness and Efficiency: (1) Does the program do what it is meant to do, does it bring about the intended change and what change, if any, does it bring about; the falsity of the behaviourist surrogate argument notwithstanding. The behaviourist surrogate argument advanced by Sociology in relation to Evaluation (Gordon & Morse, 1975) is that the goals of the program need to be known and clear for an Evaluation to be conducted. This is not the case and the argument against this in the context of Evaluation was provided by Scriven (1973). Furthermore, the behaviourist surrogate argument is untenable when we consider the process of evaluation in relation to contact with unknown objects; and, (2) is the program cost-effective (is the cost of the program appropriate in relation to the perceived benefit), and is there a cost-benefit (do the benefits outweigh the financial costs regardless of whether or not the cost is appropriate). These are core values against which any program can be assessed. However, they are not enough.

A further set of pertinent values in Evaluation are those of the policy makers, the ultimate owners of the program—those who are responsible for the program. When we Evaluate a program the values, the intents, of the policy makers are critical in order to appreciate what the pre-existing merit, worth, and significance is to the policy maker as this will impact on everything from the commissioning of the Evaluation to how the results are interpreted. It also shines a light

into the inner-mechanisms of the program. And it may well be that a program reviled by the evaluand is successful in bringing about an intended change, which may or may not be 'good', and so on. (Clearly, the reverse may also be the case.)

This is not to say the values of the policy makers will determine the merit, worth, and significance of the program per se, but they will determine the merit, worth, and significance of the program for the policy makers.

This does not mean that those values cannot be queried or challenged, and it does not mean that other sets of values cannot equally be considered. Indeed, a good Evaluation attributes meaning/value in accord with both the values of the policy makers and the values of the evaluand. These values may be, or may not be, in accord with each other and consequently the Evaluation may find that the object, the program, has different meaning/value for the different groups.

The importance of the values of the policy maker is almost clarified by Scriven (2007) when he says, "The key issue for program evaluators is the effects the program actually had, measured in terms of what they meant to those affected (and those you did not reach), whether or not you meant to have those effects..." (p. 8). In other words, what meaning did the program have for the policy maker in relation to the evaluand, and this can only be determined in relation to the values of the policy maker. Furthermore, Scriven explicates the types of value claims and these claims are very much in keeping with the values of the product/policy maker and do not always necessarily reflect the values or attributed meaning of the evaluand.

As a final note it behoves us to remember that the road to hell is paved with good intentions. A good program does not need the policy makers to hold a set of values that another community, or even the intended recipients of the program, considers good. Similarly, a program that is highly valued by the intended recipients (the evaluand) in the short term may do great harm in the longer term and may dramatically exacerbate the problem the program was intended to resolve. And so on. What this highlights is the need for evaluators not to allow their own values to

influence the Evaluation but to utilise the relevant value sets of Evaluation, the policy maker, and other stakeholders and not to privilege one set of values over and above the other (at least until the value set of Evaluation is clearly determined, agreed, and established and the methodology to manage and bound the influence of these values is equally established and agreed).

That having been said there are clearly dominant values that need to be considered because policy makers will base their decisions on these values and an Evaluation that does not take account of them lacks relevance other than from the moral position of the evaluator, and it is not the evaluator's role to make moral judgements from their own entrenched moral position.

A Plausible History of Program Evaluation

The recent history of Evaluation largely assumes it can be traced back to the New Deal era of Roosevelt 1933-1936 in the United States. While there is a good argument that Evaluation became widely acknowledged by the United States government during this time this cannot be taken as the origin of Evaluation.

To gain some idea of a more comprehensive understanding of the historical origins of Evaluation I conducted a simple search of the British Library Catalogue for the keyword 'evaluation'. This gave 1,440,427 results. Scanning the results showed an eclectic mix. I narrowed the search down to 'books' with 'evaluation' as a key word. The results contained publications as early as 1640, which are French 'economic evaluations'. They translated as 'valuations' and relate to valuing different currencies. However, as Scriven and others have noted, evaluation is the valuing of an object.

The French continued to publish on evaluation through the 1700s-1800s including evaluation relating to quantity (weights and measures) and quality. However, it was not limited to this. A reasonable assessment of these early French publications on evaluation show they were concerned primarily with money/currency, weights and

measures, and proceedings of the king's court. In other words: economics, science, and politics/policy. Thus there is every reason to suggest, bearing further investigation, that Evaluation and the use of evaluation in Science in its current form, has its roots in France from c. 1600-1800+.

Progressing into the 1900s there is a strong growth in the number and range of English language publications across a range of topics and particularly in relation to 'the Arts' (see Table 2 for an overview of the evolution of evaluation c. 1900-1960).

Between 1920 and 1940 we start to see the first recognisable modern Evaluations emerge in the United States in relation to Social Security, Education especially, 'Public Policy' (loosely framed), and other areas. (Noting that evaluation publications continue in France during this time and in other domains.) Evaluation seems to be recognisable predominantly, but not exclusively, in Education during this period. Between 1940-1950 both evaluation and Evaluation appear well established and flourishing across many domains, including the Arts.

Table 2
A Snapshot of the Evolution of Evaluation from the early 1900s to 1960

Date	Title*
1906	Manual of Descriptive Annotation for Library Catalogues ... With chapter on evaluation and historical note.
1916	Digitalis Standardization. The Physiological Evaluation of Fat-Free Digitalis and Commercial Digitalin. By G. B. Roth. II. Preliminary Observations of Metabolism in Pellagra.
1917	Bohemia : a brief evaluation of Bohemia's contribution to civilization.
1920	Evaluation of an intellectual.
1928	American Foreign Policies: an Examination and Evaluation of Certain Traditional and Recent International Policies of the United States.
1930	An Experimental Evaluation of Equality Judgments in Psychophysics.
1932	Evaluation of Types of Student-teaching, etc.
1932	Methods and Techniques used in Surveying Health and Physical Education in City Schools. An analysis and evaluation, etc. [A thesis.]
1934	Evaluation of Oils from the Manufacture of Carburetted Water Gas by their Available Hydrogen Content. Extracted from a dissertation submitted by P. J. Merkus, etc.
1937	The Evaluation of Permanent Incapacity for Work in Social Insurance.
1937	Yoga. A scientific evaluation. [With plates.]
1938	Evaluation of the Supravital Staining Method.
1944	A method for the rapid evaluation of Glauert's expressions for the angle of zero lift and the moment at zero lift.
1945	Work-Book in Educational Measurements and Evaluation.
1947	Job Evaluation for Launderers and Dry Cleaners.
1947	Student Exercises in Measurement and Evaluation for Education and Guidance.
1947	The planned state: an evaluation of the social and economic foundations of the state in the light of a comparative study of the conditions in the east Asiatic and western countries.

Date	Title*
1948	Evaluation of Choral Music. Methods of appraising the practical value of choral compositions with reference to music generally available in the United States, etc. [A thesis.]
1950	Evaluation of Indexes of Maturity for Apples.
1950	Evaluation of Citizenship Training and Incentive in American Colleges and Universities, etc.
1950	Medieval Skepticism and Chaucer. An evaluation of the skepticism of the 13th and 14th centuries of Geoffrey Chaucer and his immediate predecessors, etc.
1954	The "Isms." A history and evaluation.
1956	Evaluation in Extension ... Preliminary, June 1956.
1957	A Moral Evaluation of the American Law regarding Literary Obscenity. Dissertation ad lauream, etc.
1957	Man and Culture. An evaluation of the work of Bronislaw Malinowski.
1958	Statistical techniques in missile evaluation.
1958	Symposium on the Evaluation of Drug Toxicity.
1958	PERT Program Evaluation Research Task. Summary report. Phase 1.
1958	Pig carcass evaluation.
1958	Democracy, Social Science and Evaluation.
1959	Seminar on Evaluation and Utilization of Population Census Data in Latin America. Santiago, Chile, 30 November-18 December 1959, etc.
1960	Psychological Techniques in Diagnosis and Evaluation.
1960	ESP in Relation to Rorschach Test Evaluation.

Note. *All titles taken from the search of the British Library.

In the 1950s-1960s evaluation appears to have expanded in terms of remit and scope and Evaluation became established as a state mechanism with its own department. Not in the United States, but in India with the development of the Programme Evaluation Organization (1952). This appears to be one of the earliest, although not the only, instance/s of the use of the term Programme Evaluation and perhaps the first [modern] government office dedicated specifically to this purpose. It is worth noting that at around the same time (1955) the United States Department of Defence set up a covert paramilitary mission to the Kingdom of Laos: The Programs Evaluation Office. Also, in 1957/8 the U.S. Navy Special Projects Office developed and established the Program Evaluation and Review Technique.

Throughout this latter period the use of evaluation is widespread and objects as

diverse as statistical techniques in missile testing, 'Isms', ESP, the law and literary obscenity, pig carcasses, and the work of Bronislaw Malinowski were being evaluated and the role of evaluation in democracy and social science was being written about. From the 1960s onwards not only was the use of evaluation widespread but Evaluation quite literally exploded across the policy world.

I put forward that modern Evaluation has, thus far, two distinct eras. The first began in the 16th century France and was concerned primarily with money, weights and measures, and proceedings of the king's court (economics, science, and policy). The second phase did indeed emerge in the United States as early as the 1920s and was born in the fields of Education, Social Security, and Employment/Work, all of which are strongly interconnected. The embedded institutional origins of Evaluation in direct relation to

government first emerged in India in the form of a specialist government office. What we must not lose sight of is that evaluation as an artefact was simultaneously developing equally rapidly across a whole range of activities, practices, and disciplines (see Table 2).

A Disciplinary Perspective of the Modern History of Program Evaluation

The argument that Evaluation is a distinct discipline is well rehearsed but neither universally acknowledged nor accepted. More widely accepted is that Evaluation is a practice of the Social Sciences—although this is debatable and it depends on what is classed as a Social Science. Further to this, Evaluation was demonstrably central the founding doctrines of the Social Sciences in the early 1900s (Weber, 1949). As an aside it is acknowledged that there are a number of distinct schools of Evaluation that have emerged as independent entities (e.g., Empowerment Evaluation, Randomised Evaluation, Realist Evaluation, and Utilization Focused Evaluation), but they do not constitute disciplines in their own right either individually or collectively and for the most part they are contained within, or constrained by, other disciplinary boundaries with regard to the philosophical underpinnings and methods that bound and constrain the associated value set/s.

In the absence of due consideration of the contribution of various disciplines to Evaluation, and the role of those disciplines, historic and ongoing, in the practice of Evaluation there can be no valid historical understanding of Evaluation (except that it is limited to substantial constraints and objects, see for example the work of: Ayob & Morell,

2016; Hogan, 2007; King & Alkin, 2016; King & Alkin, 2019).

The *Annual Review* series of journals provide annual reviews of key topics and developments within distinct disciplines. The reviews are conducted by eminent scholars and practitioners within the disciplines and reflect and represent both the historic and current state of understanding of the topic within the discipline. As such they provide something of a 'level playing field' for assessing any one topic, or range of topics, both within and across the disciplines covered by the *Annual Reviews*.

Over the years, I have conducted regular searches of the series for papers on 'evaluation' and papers with 'evaluation' in the title. This reveals an accepted and documented history of the understanding of the topic within and between disciplines. For the purposes of this essay I conducted a fresh search for reviews with 'evaluation' in the title. I exercised some leeway in selecting the papers and made an effort to select only those papers that commented directly on social intervention and particularly those papers that also had 'program' and/or 'policy' and/or 'intervention' in the title.

Table 3 lists the *Annual Review* journal of publication, the year of publication, and the title of the selected paper. Authors have not been listed as it is the topic that is of interest as represented by the title. The largest number of relevant titles was found in [The Annual Review of] Public Health with 16 titles between 1980-2019, next was Psychology (including two sub-disciplines) with nine titles between 1976-2019, followed by Economics (including one sub-discipline) with seven titles between 2009-2018, then Sociology with four titles between 1975-2012, Statistics with two titles between 2014-2018, and finally three disciplines grouped together in the category 'Other' with three titles between 1992-2019.

Table 3
List of Titles relating to Evaluation in the Annual Review series of journals

Annual Review of	Year	Title
Public Health		
Public Health	1980	Economic Evaluation of Public Health Programs
Public Health	1981	Primary Prevention of Ischemic Heart Disease: Evaluation of Community Interventions
Public Health	1991	The 20-Year Experiment: Accounting for, Explaining, and Evaluating Health Care Cost Containment in Canada and the United States
Public Health	1992	Selected Methodological Issues in Evaluating Community-Based Health Promotion and Disease Prevention Programs
Public Health	1994	Relapse Prevention for Smoking Cessation: Review and Evaluation of Concepts and Interventions
Public Health	1996	Economic Evaluation of HIV Prevention Programs
Public Health	1998	Evaluating the Cost-Effectiveness of Clinical and Public Health Measures
Public Health	2000	Preference-Based Measures in Economic Evaluation in Health Care
Public Health	2006	Process Evaluation for Community Participation
Public Health	2010	Evaluability Assessment to Improve Public Health Policies, Programs, and Practices
Public Health	2012	Economic Evaluation of Pharmaco- and Behavioral Therapies for Smoking Cessation: A Critical and Systematic Review of Empirical Research
Public Health	2014	Evaluation of Systems-Oriented Public Health Interventions: Alternative Research Designs
Public Health	2017	Evaluating the Health Impact of Large-Scale Public Policy Changes: Classical and Novel Approaches
Public Health	2017	An Overview of Research and Evaluation Designs for Dissemination and Implementation
Public Health	2017	Public Health Surveillance Systems: Recent Advances in Their Use and Evaluation
Public Health	2019	Innovations in Mixed Methods Evaluations
Psychology		
Psychology	1976	Program Evaluation
Psychology	1980	Evaluation Research
Psychology	1983	Evaluation Research: A Methodological Perspective
Psychology	1986	Program Evaluation: The Worldly Science
Psychology	1993	Program Evaluation
Psychology	2000	Evaluation Methods for Social Intervention

Annual Review of	Year	Title
Psychology	2008	The Renaissance of Field Experimentation in Evaluating Interventions
Clinical Psychology	2012	Development, Evaluation, and Multinational Dissemination of the Triple P-Positive Parenting Program
Organizational Psychology and Organizational Behavior	2019	Measurement Development and Evaluation
Economics		
Economics	2009	Improving Education in the Developing World: What Have We Learned from Randomized Evaluations?
Economics	2010	Structural Estimation and Policy Evaluation in Developing Countries
Economics	2018	Econometric Methods for Program Evaluation
Resource Economics	2009	Behavior, Environment, and Health in Developing Countries: Evaluation and Valuation
Resource Economics	2010	Providing Safe Water: Evidence from Randomized Evaluations
Resource Economics	2011	The New Economics of Evaluating Water Projects
Resource Economics	2018	Advances in Evaluating Energy Efficiency Policies and Programs
Sociology		
Sociology	1975	Evaluation Research
Sociology	1984	Evaluation Research: An Assessment
Sociology	1987	Network Approaches to Social Evaluation
Sociology	2012	Toward a Comparative Sociology of Valuation and Evaluation
Statistics		
Statistics and Its Application	2014	A Systematic Statistical Approach to Evaluating Evidence from Observational Studies
Statistics and Its Application	2018	Statistical Modelling for Health Economic Evaluations
Other		
Ecology and Systematics	1992	Definition and Evaluation of the Fitness of Behavioral and Developmental Programs
Environment and Resources	2010	Evaluating Energy Efficiency Policies with Energy-Economy Models
Nutrition	2019	Evidence Collection and Evaluation for the Development of Dietary Guidelines and Public Policy on Nutrition

Given the direct relevance of the titles to Evaluation across this range of disciplines it is inappropriate to comment on less than all of them for the purposes of this essay, so I

haven't. What does emerge from assessing the titles alone is: (1) a number of titles demonstrate the use of Evaluation in relation to a discipline relevant object; (2) a number of

titles demonstrate the contribution to the development of Evaluation by that particular discipline; and (3) the importance of both economic and methodological considerations.

Equally interesting is the temporal spread within each discipline. Public Health has reviewed Evaluation consistently and constantly since 1980. Psychology similarly, though not to the same extent and the contribution is arguably different. Sociology had a small spread of reviews with large gaps also since 1975. Economics does not start reviewing the topic until 2009, roughly the time when psychology and sociology began to tail off their contribution. Statistical reviews don't start until 2014. While psychology and sociology are much concerned with Evaluation as an object the other disciplines seem more concerned with Evaluation in relation to economics, cost-effectiveness, implementation, and policy implications.

This is, as already noted, little more than a glimpse of what promises to be a very rich historical tapestry indeed and I suggest that much could be learned from a substantive review of these reviews especially if additional titles were included.

Similarly to the *Annual Review* series the *Philosophical Transactions of the Royal Society of London*, the world's first and longest-running scientific journal/s, provides an overview of key topics and developments across scientific disciplines rather than within them. I conducted a search for articles with 'evaluation' in the title, and again exercised some leeway in selecting the papers, and made an effort to select only those papers that commented directly on social intervention. The results are presented in Table 4.

Table 4
Selected articles from a search of *Philosophical Transactions of the Royal Society of London* (Only articles with 'evaluation' in the title and that reflect on social intervention selected)

Year	Title
1973	Advanced technology in the hospital laboratory: Equipment for clinical chemistry and its evaluation.
1976	Evaluation of existing techniques: Is 'the pill' safe enough to continue using?
1977	Evaluation of land-use and land-treatment practices in semi-arid western United States.
1977	The evaluation and exploitation of the West African Sahel.
1977	The evaluation and exploitation of semi-arid lands: Australian experience.
1980	Methods, equipment and techniques for rural health care and their evaluation - The people.
1981	Quantification of biological risk assessment and evaluation of risks to health from chemicals.
1990	Land resources inventory and productivity evaluation for national development planning.
2008	An age-structured model to evaluate the potential of novel malaria-control interventions: a case study of fungal biopesticide sprays.
2010	New directions in management strategy evaluation through cross-fertilization between fisheries science and terrestrial conservation
2011	Evaluation of vaccines against enteric infections: a clinical and public health research agenda for developing countries.
2014	Veterinary and human vaccine evaluation methods.
2015	Using prediction markets to forecast research evaluations.
2015	Evaluating the potential impact of enhancing HIV treatment and tuberculosis control programmes on the burden of tuberculosis.

Year	Title
2015	Impact evaluation to communicate and improve conservation non-governmental organization performance: the case of Conservation International.
2015	Integrating impact evaluation in the design and implementation of monitoring marine protected areas.
2015	International funding agencies: potential leaders of impact evaluation in protected areas?
2016	Re-evaluating the health of coral reef communities: baselines and evidence for human impacts across the central Pacific.
2016	System-level approach needed to evaluate the transition to more sustainable agriculture.
2016	The National Eclipse Weather Experiment: use and evaluation of a citizen science tool for schools outreach.
2017	Evaluation of natural sounds in urban greenery: potential impact for urban nature preservation.
2018	Evaluation of mechanistic and statistical methods in forecasting influenza-like illness.

Once again it is clear that not only evaluation but evaluation of social interventions is wide-spread and far reaching and carried out from within scientific disciplines. What is interesting here is the increasing number of evaluations of social interventions and the range of what are either social interventions or topics immediately relevant to wider range of social policy and programs.

The point being, there is a well-documented history of Evaluation within and across relevant Social Science and other disciplines that will shed objective light on the modern history of Evaluation and the contributions these disciplines have made and continue to make. Furthermore, these disciplines have a high impact on policy makers and knowing the disciplinary position on Evaluation also sheds light on influences on, and the workings of, policy makers and policy making.

Program Evaluation as Evaluation Science

The notion of Evaluation as Applied Social Science renamed as Evaluation Science was mooted as early as 1990 (Rossi). Recent developments have seen the book chapter, *Evolving Methods in Evaluation Science*, in a publication by The National Academies Press (Olson, 2014), the establishment of the Office of Evaluation Sciences (2015) housed by the

U.S. General Services Administration, the creation of the post of Director for the Office of Drug Evaluation Sciences at the United States Food and Drug Administration's Office of New Drugs (October, 2019; see *Regulatory Focus*, 2019), and the awarding of the Nobel Prize for Economics (2019) to a team of highly respected evaluators who developed and established the school of Randomised Evaluation, which serves as the engine for the emerging discipline of Development Economics. The point here being that the notion of Evaluation Science is at least 30 years in the making and is clearly evident in the world of United States' policy.

Patton (2018) articulates Evaluation Science as, "...systematic inquiry into how, and how well, interventions aimed at changing the world work. Evaluation science involves systematic inquiry into the merit, worth, utility, and significance of whatever is being evaluated by adhering to scientific norms..." (p. 184). This explanation is simply an explanation of science, it is what science does. The single object that differs somewhat obliquely from a pure explanation of what science does is the phrase "...how well, interventions aimed at changing the world work" (p. 184). However, the interventions Patton refers to are actually aimed at addressing and resolving perceived and interpreted social problems rather than changing the world. This understanding is one that is consistently repeated throughout both science and Evaluation. Consider, for

example, Trochim's (1998) explanation of Evaluation as, "...a profession that uses formal methodologies to provide useful empirical evidence about public entities...in decision making contexts..." (p. 248).

Patton continues that evaluators have access to a body of knowledge about, "...ways of applying knowledge to design and improve interventions, both based on empirically and theoretically validated patterns of successes across interventions and evaluations" (p. 188). To some extent this may indeed be the case, but there is little empirical evidence, if any, to suggest that this occurs outside of boundaries of recognised and distinct disciplines whose objects are specific [social] problems. The counter argument is that the majority of knowledge applied to designing and improving [social] interventions comes not from the proprietary Evaluation literature but from relevant discipline specific literature. Of course, it very much depends on what we mean by 'evaluators' and how we assume that particular group is constituted.

What we know, if anything, from the last 100 years of Evaluation, is that it is remarkably difficult to develop a [social] policy and implement a program that will consistently produce the desired results with the desired effect. There are any number of reasons for this ranging from the way in which the problem is both perceived and constructed through the nature of social institutions to failures in the fidelity of implementation to poorly designed programs that set out with good intentions to a lack of relevant disciplinary theory and evidence to corruption to a failure to take account of varying context. This was articulated, demonstrated, and explored in some depth by Tharp and Gallimore (1979), which remains one of the definitive Evaluation reflections on this/these point/s.

Patton notes that a 'reservoir of knowledge' exists within the Evaluation community that can be drawn upon to design, implement, and evaluation interventions. What we need to recognise is that a further, perhaps more extensive, reservoir of relevant knowledge exists within the disciplinary literatures, as indicated and demonstrated earlier.

There is also the argument that program evaluators have the knowledge, skills, and expertise to advise on the design,

development, and implementation of interventions to address social problems. This is a bold statement and it supposes a level of knowledge and skill that goes far beyond the knowledge and skill required to Evaluate a social intervention. Furthermore, the inherent value of Evaluation as a craft is that it not only adheres to a set of relevant disciplinary specific values and methodological practices - to the object, either a program in this case or an object relating to a program—but also uses these to methodologically, logically, and philosophically constrain the evidence according to both those [disciplinary] values and the values held by the policy maker and the evaluand (and if appropriate other stakeholders, but there are limits). Put another way, the inherent worth of an evaluation carried out as Evaluation is that the evaluator has no inherent 'personal', value-based, vested interest in the design, development, and implementation of interventions to address social problems. The value bases are disciplinary first and policy maker and evaluand second. (This raises questions about the role of institutional values, naturally, but this is best addressed in a more advanced exploration of the institution per se.)

Of course, much of the above argument depends on how 'program evaluator' is defined. There are many within-discipline program evaluators who do not consider themselves program evaluators per se, and many program evaluators per se whose first loyalty is to their discipline. As such, we need to take care to give full and due consideration as to what we mean when we use the term program evaluator (shortened to evaluator in most cases).

A Conclusion of Sorts

The conclusion I draw is that Evaluation expertise is first and foremost located within disciplines rather than as a distinct practice. It seems to me that the key to a robust community of Evaluation with its own norms, standards, and practices is to draw on the extant literature across a very wide range of disciplines and collate and examine this broader literature. The assumption that there is an innate and unique reservoir of

knowledge, skills, and experience within a self-proclaimed group set within a large and varied field made up of any number of disciplines seems lacking. Acknowledging the wider field has the potential to position the smaller Evaluation community—that group concerned primarily with Evaluation rather than working from within an established discipline and conducting Evaluation—as an honest broker with the stewardship of a wider body of knowledge, skills, experience, and understanding. In respect of this the question emerges as follows: What are the objects of the proprietary Evaluation literature and are there any objects not found elsewhere in other disciplinary literature? (This should not be confused with an armchair exercise of collating and summarising definitions.)

With regard to Evaluation Science, two considerations emerge. Firstly, evaluation is ubiquitous within disciplines and within the sciences in particular. It has been well studied within various science disciplines and is well utilised by any number of disciplines including for the purpose of Evaluation. Further to this there is the suggestion that the majority of Evaluations are conducted from within existing disciplines rather than from within a distinct discipline called 'Evaluation'. This renders it difficult to allocate Evaluation as a proprietary object to any one distinct discipline. Even more difficult to suggest that there is a distinct discipline with either Evaluation or evaluation as its object of concern. Can we say that there is a distinct discipline of Evaluation that studies Evaluation? That is not to say it is not possible to study evaluation, as has been done, and it is not to say it is not possible to study Evaluation, as has been done. What appears to be the case, however, from a broad and long historical survey of the extant literature is that the study of Evaluation has, again, taken place from within distinct disciplines such as Public Health, Psychology, Sociology, and latterly Economics, and it is within those disciplines that the greatest advances have been made. Similarly, if we look at studies of evaluation it is within psychology that the greatest advances have been made. And if we look at the discipline of Education we see extensive study, analysis, and development of both evaluation and Evaluation taking place within the discipline.

Second, Evaluation Science is a relatively long-standing construct. It has been definitively operationalised by Life and Medical Sciences to conduct Evaluation (allowing for some expansion of the definitional boundary of Evaluation). Judging by its existing operationalisation The Evaluation Sciences utilise disciplinary specific methodologies appropriate to the object and is arguably the engine of at least one new discipline, Development Economics. We also see Evaluation Science being re-incorporated into disciplines and their objects such as Public Health (Table 3), Conservation (Table 4), Nuclear Physics, and Artificial Intelligence (Table 1). Given this, to assume that a group within a field can assume the role of Evaluation Scientists, or that a self-defined group can rebrand its efforts as Evaluation Science, risks putting that group in direct conflict with an existing and already established community of practice. That is not to say that Evaluation is not scientific, nor that it cannot be scientific. But is there any need for the community of Evaluators, or a self-selecting community of Evaluators, to engage with a fashionable motif? It may be there is a good reason on the one hand, but on the other hand it may indicate remarkably poor judgement given the wider community of Evaluators will by default already include those engaged with Evaluation Science but not all Evaluators will be engaged with Evaluation Science nor will they all be recognised as Evaluation Scientists.

Finally, none of the above should be taken as a critique of the ongoing efforts to professionalise Evaluation—a separate and much needed critique—nor should it be taken to suggest that those engaged with Evaluation are in any way lacking in necessary and sufficient disciplinary skills, expertise, knowledge, and perspective. What I question is: (1) whether or not Evaluation can realistically be considered a distinct discipline—as opposed to a practice exercised from within an existing discipline relevant to the object—and if it can then what is/are its object/s; and (2) is there any value to the wider Evaluation community in assuming that Evaluation as a whole is Evaluation Science and if that argument is to be made how is it supported and validated?

References

- Ayob, A. H. and Morell, J. A. (2016). The historical path of evaluation as reflected in the content of Evaluation and Program Planning. *Evaluation and Program Planning*. <http://dx.doi.org/10.1016/j.evalprogplan.2016.05.006>
- Chilisa, B. (2015). A Synthesis Paper on the Made in Africa Evaluation Concept. African Evaluation Association. <https://afrea.org/wp-content/uploads/2018/06/MAE2-Final-31st-august.pdf>. Accessed December 2019
- DiMaggio, P. J. and Powell, W. W. (1991). Introduction. In P. J. DiMaggio and W. Powell (eds.) *The New Institutionalism and Organizational Analysis*, pp. 1-38. Chicago: University of Chicago Press.
- Ferguson, M. and Bargh, J. (2008). The Constructive Nature of Automatic Evaluation, in Musch, J. and Klauer, KC. (Eds.) (2008). *The Psychology of Evaluation: Affective Processes*. In *Cognition and Emotion*. Mahwah, N.J.; London: Lawrence Erlbaum Associates, 2003; 2008. Pgs. 173-194.
- Fingerhut J. and Prinz JJ. (2018) Grounding evaluative concepts. *Phil.Trans. R. Soc. B*. 373
- Gordon, G. and Morse EV. (1975). Evaluation Research. *Annual Review of Sociology*, 339-362.
- Gerth, H. H. and Mills, C. W. (1953). Character and Social Structure: *The psychology of social institutions*. Harcourt, Brace and Company. New York.
- Hogan, R. L. (2007). The Historical development of Program Evaluation: Exploring the past and present. *Online Journal of Workforce Education and Development*, Volume II, Issue 4.
- Judd, C. H. (1936; 1926). *The Psychology of Social Institutions*. The Macmillan Company. New York.
- Kantor, J. R. (1919). Psychology as a science of critical evaluation. *Psychological Review*, 26(1), 1-15. <https://doi.org/10.1037/h0074274>
- King, J. A. and Alkin, M. C. (2016). The historical development of evaluation use. *American Journal of Evaluation*, 37(4), 568-579.
- King, J. A. and Alkin, M. C. (2019). The Centrality of use: Theories of evaluation use and influence and thoughts on the first 50 years of use research. *American Journal of Evaluation*, 40(3), 431-458.
- Musch, J. and Klauer, KC. (Eds.) (2008). *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*. Mahwah, N.J. ; London : Lawrence Erlbaum Associates, 2003; 2008.
- Nobel Prize for Economics 2019. <https://www.nobelprize.org/prizes/economic-sciences/2019/press-release/>. Accessed November 2019.
- Office of Evaluation Sciences. <https://oes.gsa.gov/about/> Accessed November 2019.
- Olson, S. (*Rapporteur*) (2014). Evolving Methods in Evaluation Science. In, *Evaluation Design for Complex Global Initiatives*. National Academies Press; Washington, DC. Chapter 9, pgs. 79-86. <https://www.nap.edu/read/18739/chapter/10> Accessed December 2019
- Patton, M. (2008). Evaluation Science. *American Journal of Evaluation* 2018, Vol. 39(2) 183-200.
- Programme Evaluation Organization. <http://planningcommission.gov.in/reports/peoreport/peo/peoabout.htm>. Accessed November 2019.
- Program Evaluation and Review Technique. https://en.wikipedia.org/wiki/Program_evaluation_and_review_technique. Accessed November 2019/
- Regulatory Focus* (2019). OND Reorganization Continues Apace, Posted 25 November 2019 By Zachary Brennan. <https://www.raps.org/news-and-articles/news-articles/2019/11/ond-reorganization-continues-apace> Accessed December 2019
- Rossi, P. (1990). Forward. In H T Chen (Ed.), *Theory Driven Evaluations*. Newbury Park, CA: Sage.
- Scott, W. R. (1995). *Institutions and Organizations*. Thousand Oaks, CA: Sage.
- Scott, W. R. (2004). Institutional theory. in *Encyclopaedia of Social Theory*, George Ritzer, ed. Thousand Oaks, CA: Sage. Pp. 408-14

- Scriven, M. (1973). "Goal-free evaluation". *School evaluation*: 319–328.
- Scriven, M. (2007). The Logic of Evaluation. In H.V. Hansen, et. al. (Eds), *Dissensus and the Search for Common Ground*, CD-ROM (pp. 1-16). Windsor, ON: OSSA.
- Tharp, R. G. & Gallimore, R. (1979). The ecology of program research and development: A model of evaluation succession. In: L. B. Sechrest (Ed.), *Evaluation Studies Review Annual*, Vol. 4., 39-60. Beverly Hills, CA: Sage Publications.
- The Programs Evaluation Office. https://en.wikipedia.org/wiki/Programs_Evaluation_Office. Accessed November 2019.
- Trochim, WMK. (1998). An Evaluation of Michael Scriven's "Minimalist Theory: The Least Theory that Practice Requires". *American Journal of Evaluation*, 19 (2), 243-249.
- Weber, M. (1949). *On the methodology of the social sciences*. Translated and Edited by Edward A. Shils and Henry A. Finch. The Free Press of Glencoe, Illinois. (Chap 1 and 2 particularly).
- White, L., Barque-Duran, A. and Pothos, E. (2016) An investigation of a quantum probability model for the constructive effect of affective evaluation. *Philosophical Transactions of the Royal Society A*. 374
- Winkielman, P., Schwarz, N., Fazendeiro, T. and Reber, R. (2008). The Hedonic Marking of Processing Fluency: Implications for Evaluative Judgment. In Musch, J. and Klauer, KC. (Eds.) (2008). *The Psychology of Evaluation: Affective Processes in Cognition and Emotion*. Mahwah, N.J. ; London : Lawrence Erlbaum Associates, 2003; 2008. Pgs. 195-225.
- Zuckerman, H. and Merton, RK. (1971). Patterns of evaluation in science: Institutionalisation, structure and functions of the referee system. *Minerva* 9:66-100.