

Editorial: The Fiefdom Problem

Michael Scriven

NOTE: Editorials in JMDE represent the personal views of the editor who signs them, not of the journal's editors or staff as a group. They are somewhat uncommon in scholarly journals, but JMDE is a somewhat uncommon journal. Correspondingly, you will not be surprised to hear that they are published with the thought of stimulating a discussion, or at least reactions, so please send in your considered reflections on them!

The emergence of dominant countries in world politics is marked by a history of the amalgamation of fiefdoms—mini-empires usually ruled despotically by a baron, prince, king, or maharajah. Usually the fiefdoms were too small to defend against some of their neighbors, and they were often too small for major economies of scale in production. Hence they formed alliances through marriage, trade, or mere covenants. Of course, these are fragile links, compared to complete unification, so the path to better defense, industrialization, and further expansion—as well as riches for the conqueror—lay along the latter path, which often was unilateral and of course it also resulted in an entity powerful enough to invade or dominate still larger but reluctant fiefdoms and eventually countries. The great empires, from West to East, developed in this way, and it is often said that this is the way that the present leadership in the USA is trying to go, under the smokescreen of (selectively applied) slogans such as democratization, the Monroe Doctrine, 'death to tyrants,' or 'protection of vital interests.' Whether or not that rather cynical view is correct is not the issue. The evaluation of that policy is closer

to our business, and it's clear that its merits are now considerably compromised by two new considerations: (i) the proliferation of extremely powerful, portable, and cheap weapons; and (ii) the exemplar of successful guerilla resistance to mighty armed forces. It thus seems possible that the best view of the present situation is that the way the US won the Cold War (or the USSR lost it) may be the only cost-feasible path for world leadership, as the violent alternatives simply continue to falter or fail. It might be called, "takeover by exemplifying a better way".

These thoughts about fiefdoms and their fate are occasioned by two recent events, and one persistent problem in the evaluation world. The first of the recent events is the Causal Wars that began last year, which remind us that the world of ideas is not immune to the bare-faced use of political power, misrepresentation, and ad hominem argumentation in the struggle for ideological and economic control. The other is a request to all presenters at a major series of educational workshops and seminars this past summer—not the Evaluators' Institute, by the way—that they should adhere to the definitions and structuring of evaluation provided in some online resources provided by the sponsors. This seems harmless enough—and was, I am sure, merely an effort to avoid confusion amongst the attendees—until one studies these definitions and structure. Then one discovers something that, one recollects unhappily, has now become too frequent an occurrence: a multiple and major failure to grasp the essential elements of many of the basic concepts of our field. The definitions provided for terms shared with statistics, social science methodology, or common English are quite adequate: but definitions of terms unique to evaluation reflect a severe lack of clarity about these concepts. And now one recollects that there are other foundations, organizations, and educational institutions that are prominent in the evaluation business, and deserve much credit for their support and work in that field, where the same tendency to standardize on

confused interpretations of these concepts has become part of the—conscious or unconscious—efforts at ‘branding’, that is, the effort to leave a distinctive mark on some part of the field that will demonstrate one’s own contribution.

The result of each fiefdom standardizing on their own (significantly different) usage is of course just the kind of confusion at the macro level that the standardizers are trying to avoid in their own bailiwick: a person learning or using one set of definitions will have trouble understanding and communicating with those trained to another version. We've already seen this happening quite often on Evaltalk. If combined with the kind of economic and political enforcement that has occurred in the Causal Wars takeover of most of the federal funding for educational research, where some \$500 million per annum is now (de facto) reserved for those with the ‘right views’ on the highly controversial issue of establishing causation, we will seriously undercut the possibility of progress towards an understanding of the nature of our field, and of our discoveries in it, whether it's conceived as a discipline, a profession, or a set of practices. In other words, the political cycle from fiefdom to empire is playing out again in our domain, and we should be concerned that evaluation funding restrictions, for philanthropies, will follow the federal precedent in being totally restricted to those willing to share particular variants of standard conceptual frameworks that lack adequate justification for the variation.

This is a good moment to remind ourselves of the classic disaster of this type, the stupid blunders of the statisticians who casually redefined perfectly good words in the English language in such a way as to confuse millions of students and citizens for most of a century. To redefine ‘reliability’ so as to exclude its common meaning which includes validity, instead of using ‘consistency,’ was the first of a series of analogous mistakes, where ‘significance’ was next to suffer, and then

‘explanation’ as abused by factor analysts¹. The current attempt to redefine ‘evidence-based practice’ in medicine, public health, social services, education, etc., is at least one where more sophisticated arguments are being used.

Back to the fiefdom problem. The third trigger for this concern with the Balkanization of evaluation—that is, unnecessary fragmentation, confusion, and attendant hostility, with the shadow of dictatorship in the background—is of much greater importance to the world at large. In the field of international development, it has become increasingly clear that the situation with the evaluation of interventions is far from satisfactory. This areas has long been one of concern to thoughtful evaluators, because of the combination of limited external oversight with the usual strong (though tragically short-sighted) double-barreled motivation for doing superficial or zero evaluation—namely, that serious evaluation *might* make you look bad, and it uses valuable resources. This appeal to both risk-management and fiscal conservatism is always hard to beat². More detailed analysis, especially by Paul Clements, one of the faculty for our doctoral program in evaluation here, makes clear by on the ground meta-evaluation studies in Africa of the World Bank, CARE International and USAID program evaluations, that these concerns are all too appropriate³. Each maintains a fiefdom of its own

¹ Perhaps the ultimate example was to elevate a contradiction to the level of a recommended procedure, viz., ‘grading on the curve.’

² What beats it here, and in most other situations, are these four cards from the same pack: (i) evaluation is the best way to uncover early warnings that allow you to avoid the worst possible outcomes; (ii) it is a cost-effective way to get that insurance; (iii) it will sometimes produce highly positive conclusions which reward good work and can generate further rewards, including further opportunities for good work; (iv) it will often benefit you far more than it costs, in ways other than those referred to in (i) through (iii), e.g., by showing the way to plain quality improvements. Evaluations that do these things need to be designed to do them, and many are not; so these are not blanket endorsements of evaluation, but only of good evaluation practice.

³ For further information, Google will locate sites that have his doctoral dissertation (from the Woodrow Wilson School at Princeton): search on the first five words of “Development As If

operations, including their evaluations, which has its own rules and indeed culture. Despite some improvements, and—please note—*some* very good evaluations, gross errors persist. The editors hope, and intend, that this journal will provide one source of encouragement for improvement in this area, and hope to include an article by Dr. Clements in the next issue, as well as comments from country evaluators where the big development agencies operate.

Related to this example is the recurrent tendency for agencies to issue RFPs for ‘external evaluations,’ in which they overspecify the design all the way down to overspecifying the requirements for bidders⁴. Doing this of course undercuts externality to the point where it loses most of its contribution to credibility and seriously attacks validity. A tempting way to extend the fiefdom, of course, and nearly as bad as sole-sourcing the contract to a friendly consultant. In other words, how to make an external evaluation into an internal one.

What else can be done to avoid both the linguistic confusion and the Balkanization of research—and the funding of research—on evaluation? We might be able to learn something from what happens in philosophy, the field where nothing is taken for granted, all concepts are up for reformulation, and very different interpretations of the key ones are taught at different colleges, depending on which school of thought is dominant amongst the resident faculty. Doesn’t this just show that one can’t hope to prevent multiple interpretations of key concepts? I believe the main lesson to be learnt is more fundamental: one must treat the definitions of key existing concepts as an extremely serious matter, not a matter of casual linguistic

Impact Matters: a comparative organizational analysis of USAID, The World Bank, and CARE based on case studies of projects in Africa”

⁴ A particularly egregious example is the latest UNICEF RFP to evaluate their “programme of cooperation” to strengthen humanitarian responses. Here the requirements on the bidders amount to requiring membership in the inner circle of ‘acceptable’ players, i.e., the fiefdom country club.

convenience (which is true only with neologisms). Conceptual schemes, and the definitions that go with them, are powerful instruments of analysis and hence persuasive support for particular interpretations, not minor precursors to it (a point well made in *Zen and the Art of Motorcycle Maintenance*, by the way).

Constructively speaking, I will also take two steps myself: first, I will propose to a few leading organizations engaged in teaching, supporting, and propagating evaluation, that we need to hold a small conference of interested parties on a double topic, which we might call “Finding Common Ground”. The agenda would cover: (i) *standardizing terminology* where possible, the reasons for doing this, *and* the limits of such attempts; and (ii) finding *compromise positions* on major conceptual issues, such as the one about causation. This is a natural marriage of goals, since the difference between common definitions and common analyses is only a gradual one.

Second, I will take care, in the doctoral program that I run, to stress the existence of, the case for, and the need to tolerate, alternative conceptual schemes and definitions besides the ones for which I argue—although not to treat this as a matter for arbitrary decision, but rather as something that requires serious justification. That’s a tough distinction to make. I hope others will join in this conscious effort, or write to *JMDE* explaining why they think this is an undesirable strategy—or one in need of major extensions.

ENDNOTES 1. The most important potential relevance of this editorial is to the problem of evaluation in Europe today; and probably in Africa tomorrow. We'll try to carry some news about the conflict between the urge to brand, a.k.a nationalism, and the urge to communicate.

2. No good evaluator would read the above without noting that it can also be seen as an attempt by someone who invented a fair number of the terms in the evaluation vocabulary to extend his own fiefdom. While I do think that people who invent terms have some obligation to argue against careless shifts from their original meanings, they also have an obligation to be open-minded about *serious arguments* for modification or clarification of the original definitions. I make an effort in the *Evaluation Thesaurus* not to 'brand' the dozen or so terms I have introduced, like meta-evaluation, impactee, and the formative/summative distinction, with any claim to authorship, hoping thereby to free others to suggest modifications to the definitions. And I'm now inclined to think that the arguments, notably by Michael Quinn Patton and Eleanor Chelimsky, for adding a third category to formative and summative have merit, although I originally took those two types to be exhaustive. In an essay in Alkin's *Evaluation Roots* (Sage, 2004) I suggest one might use "ascriptive" to identify certain evaluations—for example, an evaluation done by a military historian of Napoleon's use of cavalry—that are aimed at neither improvement of an evaluand, nor macro-decisions about it⁵, but

⁵ Macro-decisions include decisions to continue or cease funding, purchase, or export. The contrast is with micro-decisions, e.g., whether to improve the programming of the organization's database. The latter but not the former are serviced by formative evaluation.

simply at determining/ascribing merit, worth, or significance ‘for its own sake’.⁶
There, I’m not incorrigible; how about you?”

Example: here’s one of the World Bank’s definitions:

Meta-evaluation—The term is used for evaluations designed to aggregate findings from a series of evaluations. It can also be used to denote the evaluation of an evaluation to judge its quality and/or assess the performance of the evaluators.
Meta évaluation Évaluation conçue comme une synthèse des constatations tirées de plusieurs évaluations. Le terme est également utilisé pour désigner l’évaluation d’une évaluation en vue de juger de sa qualité et/ou d’apprécier
Metaevaluación Este término se utiliza para evaluaciones cuyo objeto es sintetizar constataciones de un conjunto de evaluaciones. También puede utilizarse para indicar la evaluación de otra evaluación a fin de juzgar su calidad

Comments by MS. The definition treated as primary—the one in the first sentence—is a simple confusion of meta-evaluation with meta-analysis. The second definition is correct and of course quite different. Arguably, the former will not result in an evaluative conclusion, but in an analytic conclusion of the following (non-evaluative) kind: “The evaluations studied lead to the conclusion that on balance, the new meningitis vaccine is not unduly risky for those with compromised immune systems.” A meta-evaluation always leads to an evaluative conclusion, of the form “This evaluation is sound/unsound/clear/unclear/credible/not credible.”

⁶ The example is exotic, because most examples of ascriptive evaluation can be used for formative or summative purposes by users in contexts other than that of the originator, which tends to blur the distinction.