The Rise and Fall of Evaluation Standards

Petri Virtanen

ABSTRACT

This article takes a critical look at the limits of applicability of current evaluation standards. First, the role of values in evaluation theory is scrutinized building upon the typology presented in Shadish et al. (1991). Second, the article builds upon the fact that there are a number of psychological studies available which expose counter-intentional and unconscious biases in our behaviour. Is it possible to control the quality of evaluation by standards at all? Third, this results in the fact that the current evaluation standards available are not salient enough. The use of standards is highly personal and individual and practical applications vary greatly.

In the world of ethics, a number of symmetries and asymmetries exist and this does not make it easy to understand short notions put forward as sentences in various evaluation standards developed by national evaluation communities. In ethics, different paradigms and schools of thought compete, utilitarian theories with deontological theories, utilitarian and deontological theories with virtue theories, egalitarian theories of justice with libertarian theories, and so forth. It appears that current evaluation standards do not have so much to offer to the discussion of values and ethics. The problem with standards is that they do not necessarily relate to value theory.

Evaluation standards and evaluation criteria do not provide "miracle" solutions and it would be truly naive to assume that it would work out that way. Hence, there is an immanent black box present when an evaluator does his/her evaluation task - whether it be an evaluation of a programme, a project or an institution - and tries to come up with relevant conclusions from his/her evaluation object. By following the guidelines set out in various standards we cannot be sure that evaluation is automatically of good quality and ethically of "high-class".

Key words: Values, ethics, standards, evaluation

INTRODUCTION: FROM EVALUATION FEARS TO STANDARDIZATION?

Imagine yourself playing a game of chess. You know the rules about how to make the moves, you have tried it a number of times and you have managed to play the game with your friends. Does this - that you know how to make the moves - make you a good chess player, or an expert in chess? Hardly. As soon as you know the rules, you are still far from being an expert. You do not know the winning strategy.

Continue the same mind game. Imagine yourself reading through all possible available evaluation standards, guidelines or quality checklists. Does this information make you expert in evaluation practice? Do these guidelines guarantee that you can perform well as an evaluation expert? I think not. On the contrary, as soon as you are aware of the evaluation standards and understand them properly, you are far from being expert in evaluation practice.

Derlien and Rist (2001: 452-453) have asserted that a shift has occurred - a decentralization shift, affecting evaluation actors from central government level to the lower levels of government and other public sector domains. This shift is, to follow the words of Jann and Reichard (2003: 39), mainly due to the growing devolution of ministries and central departments towards lower levels of government, more independent public, private or hybrid agencies and organizations. Today, evaluators at the regional and local government level have also

emerged in large numbers and they have become new players on the evaluation field. The demand for professional guidance seems to be immense and immanent. The question of evaluation ethics relates to the question of evaluation culture and to discussions concerning the norms or ethics of the evaluation profession (Furubo & Sandahl, 2001: 6-12). In an international comparison, national evaluation societies have played a dominant role not only in making up the evaluation fabric but also in formulating the evaluation standards, ethics and norms (on the structural base of the field of evaluation, see Shadish et al., 1991; 25-28; Shadish et al. 1995).

At the moment, there are evaluation standards available, for instance, by the European Commission, the Swiss Evaluation Society, the German Evaluation Society, the Canadian Evaluation Society, the American Evaluation Association, the African Evaluation Association, the Australasian Evaluation Association, and by a number of other associations and institutions. These standards have been considered to be a guarantee of quality in carrying out evaluation studies and they are addressed to all stakeholders - i.e. evaluators, commissioners of evaluation studies and evaluands themselves. In this sense, they include a strong prescriptive element - meaning, that evaluations ought to be implemented "according to these lines", "this way", and the like. It seems that the vastness of the subject matter is currently daunting. Still, critical literature on the subject matter remains scarce. In the United States, for example, evaluation standards date back to the 1950s. The Joint Committee Standards were created to bring diverse stakeholder groups together to get a purchase on the meaning of evaluation in the context of the failures of evaluation in the early years of the War on Poverty programmes of the 1960s and 1970s. Prior to the Joint Committee Standards, the narrow view that any good evaluation need meet only the experimental design requirements of internal and external validity and the measurement requirements of reliability and validity was generally acknowledged.

Certainly, the Joint Committee Standards helped evaluators and clients see the great limitations of standardized tests and experimental design studies. Perhaps most importantly it brought together a diverse group of teachers, counsellors, statisticians, evaluators, policy makers, administrators, and so on, to pursue a common purpose of strengthening evaluation theory, practice and utilization. In this respect, the standards defined a common evaluation language and general guidelines that evaluators and educators in general could use to collaborate productively in evaluation work. From this perspective, a standard is a principle commonly agreed to by a group of service providers for guiding assessments of their work. Finally, the Joint Committee Standards earned and have maintained accreditation by the American National Standards Institute.

As a whole, this text stems from the idea that evaluation standards are here because of a fear. In human life, fear is no news. Evaluation fears and risks there perhaps are. Now, a clear distinction between risks and fatal flaws should be drawn. We may, and in most cases do, recognize that even the most carefully prepared standards and guidelines offer scope for misapplication (innocent or malicious), just as we should recognize the inherent risks of bias in ourselves as evaluators, whether individually or as a profession. What might they be and in what form do they exist? In mapping out potential evaluation fears and risks, at least three possibilities arise. First, the fear of performing morally badly or unethically, as such. This also includes the fear of poor implementation, not according to scientific standards, and the fear of being subjective (Bourdieu & Wacquant 1992). Second, the fear of politically unethical dissemination, or the fear of not being disseminated at all. And finally, the fear of cultural imperialism in terms of transforming evaluation know-how from one cultural context to another. As Bernstein (1998:7) suggests, to judge the extent to which today's methods of dealing with fears and risks hold true, we must know the whole story, from its very beginnings. This aim is not an easy one. But we begin with a clue. Evaluation fears and risks relate to evaluation standards.

What are the risks, then, in evaluation, if we are not performing in accordance with given standards? Generally, I would argue, evaluators are working in an emergent, chaotic world, where things are resolved by competition, negotiation and strategic influence rather than by universal acquiescence to a single set of rules. Above all,

there is the problem of the normative vacuum. Who sets out the rules and who is acting as the judge? This article takes a critical look at the limits of applicability of current evaluation standards. First, the role of values in evaluation theory is scrutinized building upon the typology presented in Shadish and others (1991). Second, the article builds upon a fact that there are a number of psychological studies available, which expose counter-intentional and unconscious biases in our behaviour. Is it possible to control the quality of evaluation by standards at all? Third, this results in the fact that the current evaluation standards available are not salient enough. The use of standards is highly personal and individual and practical applications vary a lot.[1]

I hope that the reader do not feel uneasy about the overall trend of the article. This is possible since people read different meanings into the same set of facts. Please note that this text is not aimed at attacking the developers of evaluations standards. On the contrary, this text tries to put forward the importance of ethics in evaluation practice. I have not tried to convey the impression that any existing set of evaluation standards (or ethics) is meant to be definitive, universal or exhaustive, for that matter. Nobody, as far as I know, who has set out to develop such standards has ever had any such pretensions. It really appears to be a question of doing the best job that can be done in the particular circumstances. In this sense, the current evaluation standards really are a child of their era.

VALUES IN EVALUATION THEORY: A CLASSICAL TYPOLOGY

In the world of ethics, a number of symmetries and asymmetries exist. Therefore, it is not easy to understand short notions put forward as sentences in various evaluation standards developed by national evaluation communities. In ethics, different paradigms and schools of thought compete, utilitarian theories with deontological theories, utilitarian and deontological theories with virtue theories, egalitarian theories of justice with libertarian theories, and so forth. That's how it goes - the list is unending since it is obvious that different traditions of philosophy and different philosophers simply disagree on which

theory to rely on. Mary Anne Bunda (1987) offers extensive discussion in this respect by analyzing three main schools, and their sub-schools, of moral thought, namely utilitarianism, duty-based theories, and right-based theories. These schools differ notably in terms of criteria for action, focus of moral decisions and options for action. According to Bunda (1987: 365-367), utilitarians focus on the consequences of an action and use the criterion of aggregate good, whereas the two deontological systems (duty-based and right-based schools) focus on the relevant principles in a situation (to be evaluated). Thinking in terms of these schools, moral behaviour is seen as an end in itself and not solely as a means to an end.

To take another example, Shadish and his colleagues (1991: 46-64) have argued that three different components form what they call evaluation theory. These components include the value component, the use component and the practice component. Therefore, values are immanent elements in today's evaluation practice. This has not always been the case though. According to Shadish and others (1991: 46-52, 455-456; see also Shadish et al. 1995), evaluators and evaluation communities paid scant attention to values during the 1970s and 1980s. Shadish and others explain this by arguing that during that time evaluators probably too naively thought that value-free evaluations were a kind of norm. Shadish and his colleagues make a distinction between descriptive, prescriptive and metatheoretical approaches to valuing. This distinction is of use when we try to understand the logic of current evaluation standards.

Descriptive valuing is - simply - a description of stakeholder values. Therefore, knowledge of stakeholder values relates to the political context of evaluation. Descriptive approaches are thus limited by how an evaluator conceives a stakeholder's moral and ethical commitments. Descriptive theories describe values without claiming one is best in comparison to other values. In practice this means that evaluators describe values held by stakeholders, and commissioners of evaluators in particular determine the criteria they use in judging a programme or project, and institutions (evaluation focus, if you like), find out if stakeholders think the subject to be evaluated was good or bad and then see what the stakeholders think should be

done to improve or to correct matters.

Needless to say, descriptive valuing is implicit in most evaluation theories, even though the word value is not used at all. Descriptive values are, in the end, quite easily constructed by being in contact with various stakeholders by organizing joint meetings, seminars, workshops, and the like. This makes descriptive valuing much more practical and perhaps (too) easy in comparison with prescriptive valuing. However, there are a number of problems in descriptive valuing. Consider goal-bound programme evaluation, for instance. Can goals and objectives be used as criteria of merit? In some cases perhaps yes, but this is definitely not always the case. The reason is that programme objectives are usually formulated through political compromise and consensus. Exclusive reliance on the goals formulated in this manner, however, may prove to be a mistake because goals can often be vague, contradictory or even latent and programmes usually have unintended, positive or negative, effects that can be just as important as intended goals.

Prescriptive ethical theories, then, advocate the primacy of particular values. To take an example, think of Rawl´s theory of justice. The concept is important but justice is still just one of several possible values available, others being, for example, human rights, liberty, freedom, equality and utility. The point is that the justification and explication of prescriptive theories usually have been carefully worked out over centuries of philosophical thinking, and the relationships of different prescriptive theories to each other and to issues in philosophy have also been explored, which means that the relative strengths and weaknesses of the approaches are quite well known today. Any evaluator carrying out research acts - or evaluation acts, to be more precise - who prioritises on perspective ethical theories has the heavy burden of justifying why.

Metatheory, as put forward originally by Scriven (1980), refers to the study of the nature of valuing and to the analysis of justification for valuing. It describes how and why value statements are constructed, underlines the structure or logic of valuing and tries to reveal the nature of justifications for values. Given this, one of the most important tasks for an evaluator is to logically construct value statements related to the evaluation focus. This logic involves selecting

criteria of merit that something must fulfil to be good, setting standards of performance about how well it must fulfil the criteria (in practice, this can be analysed by using control groups, for example), measuring performance vis-à-vis the set criteria (and putting special focus on causal relationships between causes and effects), and finally synthesizing results into a value statement. According to Scriven, valuing is not complete until a final synthesis about merit is achieved. Synthesis requires thorough and complete weighting and summation criteria - which is difficult, of course, since the measurements must be converted to a common metric at this final stage of the evaluation process.

According to Shadish and others (1991: 47-48), all of the above-mentioned ethical approaches are inadequate. What might be added as a conclusion here is the fact that the distinction parameters between these approaches are not clear either. Metatheorizing as put forward originally by Scriven differs from the two others. One might argue, namely, that neither descriptive or prescriptive approach can be achieved by using the metatheoretical way of constructing value statements. In this respect, descriptive and prescriptive approaches are approaches in the strictest sense, whereas metatheorizing points more to a process-kind of thinking. Metatheorizing is a process of constructing values, whereas the two others guide you in where to look and where to find assistance.

In brief, a good value theory, Shadish and others argue, should actually be a sum of the above-mentioned value approaches, combining their best elements and neglecting the worst shortcomings. They argue that a better theory of this value component consists of the following elements. It should, firstly, describe all of the elements laid down in descriptive, prescriptive and metatheoretical approaches. Secondly, it should be added, that a better theory should recognize clearly that no prescriptive theory is widely accepted as best - all prescriptive ethics are unjustified and selecting one immediately involves trade-offs - and that prescriptive theories suffer from inconsistency, since today´s society is based on fostering pluralism of values, competing against each other. Therefore, thirdly, a better theory should clearly state its priorities about which kinds of values to attend and to address, and why. It is evident that a good value theory

has very precise content. If it appears that the majority of evaluation theories have little content about values and ethics, then they fail this test straight away.

The distinction between the above-mentioned value approaches - descriptive, prescriptive and metatheoretical - tells little about the specifics by which evaluators actually answer questions about values in an evaluation and how they actually implement their research acts. Keeping this in mind, this sets out high hopes for various evaluation standards and codes of ethics.

It appears that current evaluation standards do not have so much to offer to the discussion of values and ethics. The problem with standards is that they do not necessarily relate to value theory. At worst, they are only individual words and sentences without any clear logic with reference to value theory. Value theory is important in evaluation since it helps evaluators understand what steps to take to make value statements, to see the value judgements explicit in their work, to make recommendations based on ethical considerations, and so forth. Without these components, evaluators may not understand the values that permeate their work.

But why are evaluation standards inadequate? First, it would be tempting to believe what Shadish and his colleagues (1991: 49) say about the ethical competence of the modern evaluator. According to their mind, no evaluation theorist does descriptive, prescriptive and metatheorizing explicitly and systematically. Some might ironically ask: do evaluators rarely even deal with topics such as ethics, morals and values? My answer to this question is, based on my experience from doing evaluations studies over a decade, is that this sounds very much like common sense. Secondly, evaluation standards as they exist today say practically nothing on valuing and values in particular. And even if they do, the contents of values remain obscure. This means that standards remain as a list of proposed good practices. Thirdly, and related to the second point, it would be naïve to assume that ethical codes as such could exist in a way that everybody views them in the same way.

Summing up, evaluation can never be value-free. The question is, then, how should we deal with this fact, if standards and codes of ethics do not convey this state of the art in a proper manner? Before answering this question we have

to take a closer look at what evaluation actually is, how it is carried out, what really "happens" in evaluation practice. And is it possible to develop ethical guidelines for evaluation practice even if this practice is unrepeatable?

IMMANENT ETHICAL RELATIVISM IN EVALUATION

What we have learned so far is the fact that values are an immanent part of evaluation practice. Now we should ask the fundamental question: how do we behave as evaluators? Are we ethical and unbiased in our professional endeavours? Banaji and others (2003) answer these questions bluntly: we are not. They explain that psychological research routinely exposes counter-intentional, unconscious biases in our behaviour. They list four interrelated sources of unintentional ethical decision-making: implicit forms of prejudice, bias that favour one's own group, conflict of interest and a tendency to overclaim credit (see also Bernstein, 1988: 270-271, for managing risks and uncertainty from the psychological perspective). Even though Banaji and others speak of managers, it would be tempting to think of their conclusions also in terms of evaluators. Banaji and others (ibid.) claim that (their original postulates are modified to accord better with the requirements and the "world" of evaluation):

- Postulate 1: Implicit prejudice is common and persistent among evaluators since it is rooted in the fundamental mechanisms of thought. We organize our world, we grow to trust them, and they can blind us to those instances in which the associations are not accurate - when they do not align with our expectations.

- Postulate 2: In-group favouritism means that evaluators tend to do more favours for those we know, and for those who tend to be like ourselves: for example, people who share our nationality, social class, values, modes of thought, and so on.

- Postulate 3: Conflict of interest can lead to unintentionally corrupt behaviour. The evaluators' working world is rife with situations in which these conflicts lead to honest,

ethically high-class evaluators to un-consciously make unsound and unethical recommendations. Just think of pressure from the commissioner of a certain evaluation study. An evaluator can find herself/himself in a situation in which s/he is forced to make recommendations for the continuation of a programme which has high political status but which also has poor results. Conflict of interest comes into the picture here. In order to maximize her/his gain (to get the next evaluation from the same commissioner), it would be reasonable to write a positive judgement on the programme and try to explain something like that "the programme has great potential, even though that the intermediate results have so far been somewhat modest". Some other evaluator might write down the following: "Indeed, the objectives of the programme are not valid or timely, the performance of the programme team has been inadequate and poor, and this has lead to the programme's very modest net effects.

- Postulate 4: Evaluators hold positive views about themselves, their skills and their performance as a whole. Many psychological studies have shown that people consider themselves above average on a variety of measures, from intelligence to driving ability. Is the methodological expertise of evaluators any exemption?

Do evaluators have help from standards to overcome the difficulties related to these kinds of decision-making problems? A quick answer to a quick question would be, from my point of view: presumably not. Ironically, it seems that only those who understand their own potential for unethical behaviour can become the high-class evaluator that they aspire to be.

Next, we should consider evaluation as a profession and view standards in that light. The reason for this is the fact that the concept of professionalization closely relates to that of evaluation standard. According to Freidson (1994: 9-10, 173-173), the word profession refers to an occupation that control its own work and is organized by a special set of institutions sustained in part by a particular ideology of expertise and service. He uses the concept of professionalism

to refer to that ideology and special set of institutions. For him, professionalism represents a method of organizing the performance of work. Professionalism revolves around the central principle that the members of a specialized occupation control their own work, i.e. they determine the content of the work they do. Secondly, members of a given profession usually control recruitment and training, even the entrance to the labour market, and produce the criteria by which performance at work is organized and evaluated.

In other words, professions are self-regulating forms of social control, i.e. professionals have been represented traditionally as independent of significant formal control by non-professionals and responsible largely to their own professional associations and to fellow professionals. In a word, professionals are in principle characterized by high performance standards, mastery of common ways of doing things that meet those standards, belief in a mission apart from financial success, career progress through an increase in skills and respect of peers, a shared language, knowledge base, training and concepts, participation in conferences to exchange knowledge and peer review of performance. Freidson (ibid.) argues that new regulatory models are being constantly developed. These include various forms of codes of ethics. The downside to all this is, as Moss Kanter (2003) puts it ironically, that professionals in some fields want to maintain the right to set all standards and to function without any constraints, becoming arrogant and insular rather than customer-focused.

Does evaluation meet these criteria of a modern profession? We might also ask whether an evaluator is an amateur if s/he is not a professional in the sense Freidson is suggesting? Shall we use the word professional, anyway, even if evaluation as an intellectual field is not a profession? These questions are really difficult to handle since the words professionalization, and profession are all extremely ambiguous words.

What are the barriers without which evaluation would become a profession, and how evaluation standards relate to the question of professionalization? There are number of potential barriers. Take the multiple sub-fields of evaluation activity in terms of policy sectors and various methodological approaches, paradigms

and schools of thought, for example. The critical questions, then, are: Is evaluation an organized occupation, presupposing esoteric competencies of various kinds? Do evaluators have concern for the quality of their work? Do they have an exclusive right to perform a particular kind of work? Do they control training and access to it? How do evaluators view themselves vis-à-vis societal power, and so on?

The French sociologist Pierre Bourdieu provides us with a potential solution to understanding the role of ethical codes in the formation of an evaluation profession. A field, in Bourdieu's sense, is a social arena within which struggles and manoeuvres take place over scarce resources and stakes and access to them. (Bourdieu, 1991: 229-231.) The point, then, is that the question of ethical codes and standards in relation to evaluation actually relates to the question of control and power. For Bourdieu, a field is a structured system of social positions - occupied either by individuals or institutions (prominent evaluators, members of the board in national and transnational evaluation associations, national associations as such, if you like) - the nature of which defines the situation for their occupations. These positions stand in a relationship of domination, subordination or equivalence to each other by virtue of the access they afford to the goods or resources that are at stake in the field. Given this, we might conceive evaluation standards as a form of gate-keeping - acknowledge them, obey them, align your activities accordingly, and you will have the opportunity to enter the field of evaluation, as the story goes.

What is the mechanism, then, that guarantees the gate-keeping? According to Bourdieu, it is symbolic violence. Bourdieu argues that it is the imposition of systems of symbolism and meaning upon groups or classes in such a way that they are experienced as legitimate. Therefore, evaluation standards are at their worst if they are considered to be fundamental in a "pedagogic" sense in evaluation practices - meaning that standards are to be implicitly or explicitly identified as the natural or "primordial" necessity that the persons or institutions entering the field of evaluation should "obey".

If we follow Bourdieu further we find an even nastier world. He argues that those who dominate the given field and who have close interest in the existence and persistence of this field and in the specific profits it secures for them also acquire the opportunity to impose their own interests as the interests of those whom they represent. At worst, this ends up in a sort of esoteric evaluation culture, comprised of problems that are completely alien or inaccessible to ordinary evaluators managing to solve their evaluation tasks in single evaluation studies. Even worse, this kind of esoteric evaluation culture can be comprised of concepts and discourses that are without referents in the experience of ordinary people, politics, policies, civil servants, and administrative affairs, and of distinctions, subtleties and niceties that pass unnoticed by the uninitiated and which have no raison d'être other than the relations of conflict or competition between the different individuals and organizations or between the 'tendencies' and 'trends' of one and the same institution (i.e. evaluation association, if you like).

Admittedly, Bourdieu offers a quite awesome and fearful picture. Nevertheless, it would be foolish to proclaim that power or control mechanisms are absent in the field of evaluation. Now we should finally give an answer as to whether evaluation is a profession or not. As far as I am concerned, definitively not in the way Freidson (1991) is suggesting, and there are number of reasons for this, as was explained earlier. Nevertheless, different members of the evaluation profession can advance markedly different ideas and still remain bona fide members of the growing evaluation family - "ideologically" at least.

The existence of an "evaluation culture" very much depends on the degree to which the beliefs, values and knowledge are shared about the process and product of evaluation practice. Evaluation standards can both contribute in their own way to spreading knowledge about professional conduct and at a certain superficial level about the overall quality measures needed to optimize the usefulness of evaluations. Next, we turn to evaluation practice. How does the landscape of ethical relativism look from there?

CONSTRUCTIVE EXPLANATIONS AND PRACTICAL CONTEXTS

What is evaluation, how is it carried out, what is the process of evaluation, and is this process controllable? Are evaluators trying to reach absolute excellence in their evaluation projects instead of merely trying to meet some regulatory standards? There are no easy answers to these questions, as Schwandt (2003) explains. According to his view, evaluation is "...a modernist practice that aims to help us live more intelligently in the world". Please note that this aim is universal in evaluation, as Schwandt points out, and different evaluation paradigms do not have different positions on this. Evaluation is, to convey the ideas of Bourdieu again, a fusion of theoretical construction and practical research operations. The latter is achieved by clarifying the facts of the matter, explaining causal relationships, establishing empirical generalizations between variables and marshalling and arraying evidence in support of claims about effectiveness, efficiency, outcomes and sustainability - this is how it goes if we speak of project, programme or institutional evaluations (Virtanen & Uusikylä, 2004).

The problem is, as we shall see in more detail when we consider the role of standards in evaluation practice, that carrying out evaluations is really and literally concerned with practice and the practical. The term practical signifies the way that we human beings are in the world and this demands particular decisions and actions of interest related with value-rational questions such: How should I act in this situation? What should be done? What are the consequences of my actions? Are my actions ethically high-class? These questions are highly context-bound, and they cannot be answered once and for all. According to Schwandt (2003: 355): "In grasping (or reading) the meaning of the practical in this sense, we do not employ the scientific grammar of theories, factors, variables and contexts but look to the ways in which we use language in an instructive and directive way in everyday life."

Thus the practical entails the arts of perception, it requires situatedness, it simultaneously involves warrants, values, emotions and commitment - all dimensions that should be included in evaluation standards and ethical guidelines. Do standards make it possible to understand appropriate orientations toward a situation and subsequent actions? Do they have the capacity to identify how real things are different from their representation in theory? Do they facilitate evaluators to reason with theory and evidence in the case at hand while taking into account the many aspects of the real situation - in which standards should be applied - that are not captured in standards in written form?

Yet again, we must remind ourselves and the eager developers of evaluation standards and ethical guidelines that evaluation in action deals with real, situated practices - that is real decision-makers, real participants, real stakeholders, real actions, real people. This is why we must focus on evaluation practice. The keys to the lock of evaluation standards and ethical codes can be found there. And, in addition to this, the limits of their applicability.

In order to grasp the above-mentioned evaluation practice, we have to resort to constructionism as put forward by Czarniawska (2001), for instance. The constructionist approach to understanding evaluation practices seeks not only to render faithfully the logic of practice but to show how it arises. This approach aims at making the logic of practice symmetrical with the logic of theory, showing actually how the logic of theory is constructed. It is essential to understand that, from this perspective, the world of evaluation has no "essence" to be discovered. Worlds are not given - or being captured by standards and guidelines, for that matter - they are constantly made and remade. This means that meaning and knowledge (including evaluation values and norms) are constructed and not "found" in things and events. These worlds are constructed in the minds of evaluators in concrete places at specific times, under the constraints present in those times and places, and they build new constraints for other places and new times.

Evaluation standards and ethical guidelines provide advice that is not salient enough to be evaluated from a constructive perspective. The use of standards and guidelines is highly personal and individual; they vary from one situation to another and do not transcend time and place. Therefore, evaluation standards and guidelines do not provide answers, but plenty of questions to be answered. A constructionist perspective instructs the developers of evaluation standards

not only to present the outputs of their development work (standards sui generis), but to study and also to reveal the process of construction itself. This postulate is one of the major concerns in understanding the usefulness of the evaluation standards and ethical guidelines. It seems that they are applied without a thorough understanding of their applicability. This means that evaluation standards and ethical guidelines cannot easily contribute to the development of the logic of theory and the logic of practice in evaluation. Their logic of representation - as written artefacts - is not easy to understand and to conceive. At worst, they are abstract and rhetorically controversial; at best, they can serve as quality checklists. The key conclusive idea from this section is the fact that evaluation standards and ethical guidelines as normative and ethical tools are impossible since they do not grasp the evaluation practice the way that evaluators grasp it. They do not provide ethical advise in moving and changing situations, which an evaluator encounters in carrying out her/his evaluation mission.

IS EVALUATION ETHICS POSSIBLE AT ALL, AFTER ALL?

"To be sure, the principles eventually come along, always well dressed and exuding an air of confidence, but always after the smoke has cleared, after the decision is over, offering the wisdom of hindsight for future travellers, the difficulty being, of course, that the future will be different and this trip has already been taken. (...) the one thing all things have in common on my accounting, is that things will always be different, sometimes slightly, sometimes greatly different. Difference, idiomaticity, singularity, exceptionality are the marks of ethical life, indeed of life itself. (Caputo 2003: 169)

This section finally asks the question: how do evaluation standards "fit in" to postmodern ethics and how do they enhance evaluation culture and the dissemination of evaluation information, if at all? In addition, we might ask whether evaluation ethics is actually possible? Of course, the concept of postmodernism is obscure. It can be seen as an explanatory social theory that has produced a number of middle range models and sub-theories of culture, science and epistemology, class, social action, gender and family relations, values and economic life. At the outset, ideally, ethics is a code of law that prescribes correct behaviour 'universally'. Values are, then, essentially supposed to be priorities. They are supposed to be choices that are present in everything we do, and do not do, everything we resist and are willing to tolerate. In terms of the evaluation profession and professional evaluators, values should be builders of integrity, responsibility, optimism, self-esteem and definers of who we are. In any case, values in evaluation activities become manifest and alive through action - and only through action. The ethical capital strengthens only by action and inaction.

However, in the words of Bauman (1995b: 10-13, 32), "...Human reality is messy and ambiguous - and so moral decisions, unlike abstract ethical principles are ambivalent. It is in this sort of world that we must live (...) a morality without foundations. If Bauman is right, no logically coherent ethical code can "fit" the essentially ambivalent condition of modernity. Moral phenomena are today inherently "non-rational" in the sense that they are not regular, repetitive, monotonous and predictable in a way that would allow them to be represented as rule-guided. This kind of reasoning does not leave much room for any codes of ethics in evaluation practice, at least as they are available today.

Bauman argues that morality and ethics are not universalizable today in a way they perhaps used to be. How does one control and guide something that is not controllable nor guidable? Following on from this, modern evaluators and evaluation communities practise moral parochialism under the mask of promotion of universal evaluation ethics, if Bauman is right.

This does not mean that morality is absent in evaluation practice, on the contrary. Values are there, we are moral beings, but ethical codes seem to be irrational. In today's world we implement evaluation studies with morality without an ethical code. No universal standards, therefore, "...and looking over one's shoulders, to take a glimpse of what other people 'like me' do...", to use the expression by Bauman (1995b: 53). To be moral does not mean to be good, but to exercise one's freedom of authorship as a choice between 'good and evil'. It would

be comforting to think that most of the people, evaluators among them, most of the time, can do very well without a code certifying its propriety. Bauman stresses that people need these kinds of codes and its authorizations so seldom that they hardly ever have a chance to discover its absence. Most people follow most of the time the habitual and the routine; we behave today the way we behaved yesterday and as the people around us go on behaving (Bauman, 1995a: 11-13).

Today, there is relatively little consensus on moral issues. The problem is to find what counts as evidence in moral issues and how it can be weighed, and how persons can reach a consensus in cases of disagreement, ignorance or doubt. The process nature of moral issues is very important to recognize; answers are not (and cannot be) given but rather made and shared - in many occasions and cases.

Let's elaborate Bauman's somewhat abstract thoughts further. Basically, as competent evaluators, we want to think of ourselves as ethical persons. This is a point on which it is easy to agree with writers such as Lingis (2003), Caputo (2003) and Schwandt (2003), to take some examples. I believe that this is a key part of our human nature. It would also be tempting to believe that we would like to live according to accepted standards of professional and social behaviour. By mature, evaluation ethics, which is a rational discourse dealing with the ought - elaborating the experience of being obligated, and therefore addressing the question of how to deal justly with other actors within the evaluation community, commissioners of evaluation, the evaluands and the general public.

Evaluation ethics makes the recognition of practical necessity into an intellectual recognition of obligation. For this to be rational, understandable, and what is most important, shared, it is constructed as a specification of a more general obligation which is binding on all rational actors of the evaluation community. This is perhaps the reason why evaluation standards have been received favourably by the evaluators. We like to think of ourselves as moral beings, concerned with standards of professional behaviour based on a sense of right and wrong. This is how far we can go. The standards perhaps "work" in principle, but their practical applicability is uncertain.

Basically, to think in terms of evaluation ethics is to think in terms of principles. Long ago, Aristotle himself warned us to be wary of the principles of ethics. According to Aristotle, when it comes to ethics we are not to expect too much precision and we need to understand our lives with general schemata and learn to settle - in terms of principles - for rough ethical outlines (see Caputo, 2003: 169-170). For Aristotle, ethical life is steeped in the concreteness and singularity of situations that are always slightly unprecedented and unrepeatable. This is very close to Bauman's (1995a) postmodern morality and ethics. According to one view inspired by the Aristotelian thinking, ethics is today only possible without principles, constituting a kind of ethics without ethical principles. Caputo's view conveys the idea of criticizing those modern moral theories that propose rigorous methods and guidelines to derive concrete moral conclusions from principles or rules. This approach assigns ethical priority to the particular situation, which is radically singular, and to situated judgement rather than mechanistic application. This means that principles cannot serve as fixed norms determining what human beings are and may become.

Do we have a solution to our fundamental problem - why the evaluation standards do not function as they are supposed to function - here? What does Caputo mean by arguing that we need ethics without ethical principles? According to him, at worst ethics is principled irresponsibility (principles without responsibility), and therefore what is needed is unprincipled responsibility (responsibility without principles). At worst, ethics is a mechanistic application of superficial ethical principles, an action dominated by principles where principles can even serve as a cover for acting irresponsibly. Even worse, says Caputo (2003: 172-173), principles can get us off the hook. They can provide us with the opportunity to put the blame on someone or something else. In the case of controversial principles, the hesitant evaluator can choose the principle that suits him the best. Thinking in terms of Caputo, evaluation standards as ethical guidelines are impossible. One of the most difficult tasks we face is to be prepared to encounter and face that which we cannot be prepared for. Briefly, we cannot construct a standard that would be applicable in all those situations where evaluation ethics would be needed. Time and again, the situations

in which the evaluator finds herself/himself, have always a certain unique dimension - these situations are different. We just cannot always subsume it under principles.

Let us elaborate this scepticism a bit further. Consider ethical blame. Lingis (2003: 206-207) stresses that while ethics admits that education and force may be required to impose ethically approved behaviour, its natural own method is blame. Ethics sanctions the blame of the rationally convinced community. According to Lingis (ibid.), ethical blame isolates an action or an omission, which it identifies to be voluntary: the actor could have dealt with the issue otherwise, in an ethically correct way. Then the question to ask is: what is ethical blame, sui generis, in the field of evaluation and within the evaluation community? Who is judging and punishing who? What forms of ethical blame are effective? National evaluation societies do not easily have, or even would like to have, this kind of role, as far as I know. How do the evaluands get the chance to cast ethical blame on the evaluators who have performed their activities in an unethical manner and caused them harm? Or, when evaluators find themselves in a situation in which there are controversial ethical claims prevailing, which ethical principles are the most important? It appears that the appeal to ethical reasons is preaching to the converted. It presupposes that the individual actor in the field of evaluation is already committed to wanting to know the reasons and rationale for actions and to want to act accordingly.

Then, if the standardized professional ethics is out of the question, the ultimate question would follow: if we do not need a deontological, virtue or consequentialist ethics for the whole range of our professional lives, do we really need instead an ethics, an art of living? From where can this be obtained? Now we are probably almost there. Perhaps the only effective ethics is not professional ethics, after all? The solution might be individual ethics, a personal world view incorporated in professional activities, whatever they might be. Bell (2003) speaks of ethical ambition - individual strategies in maintaining one's integrity while seeking success in the field of evaluation by devoting oneself passionately to one's profession, taking risks, showing courage, advancing mutual relationships and gaining wisdom. Paulo Coelho (1999) conveys the same idea in his famous novel The Alchemist by

describing his shepherd implementing a life strategy of speaking in the language of enthusiasm, of things accomplished with love, passion and purpose, as a part of a search for something believed in and desired.

Where does the above-mentioned reasoning lead us, then? To a discussion of what is the role of individual values of a single evaluator in terms of the collective values of the evaluation profession? On the other hand, if the creative class (modern evaluators included) favour individuality, meritocracy, openness and diversity, at least to a certain degree it is an individuality, meritocracy, openness and diversity of elites, limited to highly educated and well-off people in research institutions, ministries, universities and consultancies.

CONCLUSIONS

In the famous words of Popper (2002: 36-37), "...knowledge cannot start from nothing - from tabula rasa - nor yet from observation. The advance of knowledge consists, mainly in the modification of earlier knowledge (...) the significance of the discovery will usually depend upon its power to modify our earlier theories." Thinking in terms of refuting existing knowledge comes about as the landscape of values in evaluation is still to a major extent considered to be "unexplored territory", to use the words of Schwandt (1997).

Keeping the limits of constructing a new theory of valuing in mind, however, we might underline certain observations that we came across previously. Evaluation standards and evaluation criteria do not provide "miracle" solutions and it would be truly naive to assume that it would work out that way. Hence, there is an immanent black box present when an evaluator does his/her evaluation task - whether it be an evaluation of a programme, a project or an institution - and tries to come up with relevant conclusions from his/her evaluation object. By following the guidelines set out in various standards we cannot be sure that evaluation is automatically of good quality and ethically of "high-class".

In an article co-authored by myself (Virtanen and Laitinen, 2004) I have argued that ethical standards reflect a world where formal demands presumed by principles are carried out. These

demands include, among others, generality and generalization, recognition, finality and a demand for coherence. A demand for coherence in connection with ethical standards therefore means, for example, fairness adjusting to each individual situation set as a general standard. If that standard is not met, for example, because the evaluator has not had all the information needed, has then a mistake been made in applying the standard, leading to some kind of a value sanctioning of the evaluator? Or if that fairness is applied without the knowledge of the standard in question, then at least in theory it can be claimed that the general standard of fairness is not a general standard after all. For example, one ethical standard of evaluation can morally prima facie obligate the evaluator to destroy the data collected through the evaluation but, according to a law of a certain country or the regulations of an certain organization, that data may have to be signed over to the subscriber if so demanded by it.

First, there should be a comprehensive set of values or a process of defining values. After that it is possible to come to those principles by which we may try to reach the defined values. Now the question is: from where do such guidelines and standards originate? Why should we be so certain that such guidelines and principles should be preserved? How can we know how such standards or principles should be applied? How can such standards be justified? Who has the authority to define such standards and why? Are standards purely a set of ideals for some professionals of a certain type? How can standards be combined with different values? To use the term coined by Descartes 400 years ago, how are we able to resist the temptations of the Evil Demon in evaluation, i.e. how are we able to map out misbehaviour and by whom, and who controls the non-standardized use of the evaluation standards (on this Cartesian concept see Blackburn, 1999: 18-20)?

Secondly, the ideal of professional ethics is that there are rational principles or even norms guiding the practices. Is an evaluator obligated to follow these norms in a way which differs from his own every day morals or the general morals of the surrounding society? I presume not. Given this, from the perspective of applied ethics, the question of evaluation ethics is much wider than just a question of addressing professional principles.

Therefore, there can be no guarantee of success. In a world of facts we are in a state of security but in the world of ethics and values we are in a world of uncertainty. Standards do not solve that. There a number of cultural barriers that should be taken into account in discussing mainstreaming aspects of various evaluation standards and applying some evaluation standards in a different cultural context than in which they were first developed. The British drive on the left and Americans on the right - each has hit upon an equally and ethically good solution to the essential problem of coordinating traffic. Given this, evaluation standards may solve some, even many ethical problems - such as not knowing what is "good" and what is "not good", but what is important is to understand that they do not resolve all ethical issues in evaluation. One unsolved issue is, to give an example, the question of the social responsibility of the different actors.

Third, it appears that practical experience of using the standards is so far ambivalent and the structures and application models of the current standards are controversial. Practical experiences of the use of available evaluation standards suggest that they can contribute to spreading knowledge about professional conduct in the field of evaluation. In addition, in some cases they have been used for educational purposes in training and as a benchmark for quality in carrying out evaluation studies. Standards have also fostered a common language between evaluators and people commissioning evaluations. However, we do not know very precisely how these standards have been utilised. It seems that in Europe at least, the use of these standards is limited to the members of national evaluation societies. There are number of difficulties involved in developing evaluation standards: the concepts involved may appear unclear, and standards may not be valid in certain political environments and policy areas. Therefore, evaluation standards need to be specified to be applicable. Ultimately, in order to be effective, standards have to be based on discussion, consensus and agreement (on this, see also Stern 2003). Fourth, the language and terms used around the theme of evaluation standards are obscure. As the famous Ludwig Wittgenstein once explained: language is our

prison - and this is the case also in terms of evaluation standards. So what would be useful would be a conceptual taxonomy of all related terms and concepts starting from evaluation standards, a code of ethics, ethical guidelines, quality standards, best practice ethics, and so forth. In addition, it appears that the approaches taken by various stakeholders (commissioners of evaluations, national associations) to develop standards also differ. The commissioners focus on quality aspects ("meeting minimum quality requirements") and national societies focus on the role of evaluators and the evaluation practice itself. By the way, focusing on minimum quality standards is a paradox - all quality gurus since Deming, Juran and Crosby have preached for total quality and for total excellence. Minimum criteria are far from the total excellence.

Fifth, to understand the nature of evaluation standards, guidelines, or ethics, you have to resort to philosophy and logic in particular. The valuing component of evaluation theory should be stretched beyond its limits towards other components of evaluation theory, namely the use and practice components as put forward by Shadish et al. (1991; 1995). A clear distinction between the rules of definition (evaluation standards, if you like) and the rules of strategy (actual decisions made during an evaluation process) should be made. The rules of definition indicate what is allowed, what is possible, and what is not possible. The problem with the rules of definition is the fact they do not make explicit which decisions are good, bad or better than others. Rules of definition do not guarantee anything.

Sixth, a new typology is needed in terms of evaluation ethics. In practice, this means that a clear distinction between ethical codes (i.e. values that are guiding principles for the evaluator), codes of practices (describing the values selected for the evaluation community as a whole and individual evaluators), and codes of conduct ("this should be done", "avoid doing this", and so on) should be made. As a whole, the underlying logic in setting evaluation standards as ethical codes should comprise the following elements: a professionalism postulate, ethical demands of correct action by evaluation community and normative statements on how to behave in various contextual situations. The underlying logic in setting evaluation standards

as quality guidelines, then, should take into account the following elements: the supply side of evaluation products is not homogeneous, there are a number of phases in an evaluation project during which quality errors and process delays may occur, there should be minimum standards set by evaluation commissioners and finally check-lists of excellent quality evaluations for evaluators should be provided. This means that we - as competent evaluators - need new forms to enhance morally correct evaluation studies. This could be achieved by organizing joint workshops, participatory discussions, training and by publishing and circulating best (or worst) practice cases. Also worst - or perhaps, failure cases - are useful in seeing the merits of integrating a value approach to evaluation studies. In this respect, a word of wisdom from Popper (2002: xi-xii) is appropriate: "...criticism of our conjectures (tentative solutions, unjustified anticipations, guesses -PV) is of decisive importance: by bringing out our mistakes it makes us understand the difficulties of the problem which we are trying to solve (...) As we learn from our mistakes our knowledge grows..."

A number of unsolved questions remain, though. As a conclusion, it appears that the following sequence of key questions should be further scrutinized: What are the key risks of bad practice for which we have good empirical evidence or for which the context in which evaluation is practised in this country provides strong incentives? Which of those risks can effectively be addressed by a rule? What kind of rule is most likely to minimize the risk of bad practice or the incentive for bad practice? The most effective rule might not necessarily be one which directly addresses the practice in question, but one that affects the context or underlying mechanisms of evaluation in such a way as to minimize the incentive for bad practice. And, finally, how strong can we feasibly make that rule while still allowing for all the legitimate differences in value systems or practice contexts which can reasonably be expected within our national evaluation industry? This is where the issue of values really comes in. We evaluators should continuously strive for better performance and for better achievements. Why? A new social world has now come into being. We evaluators must try to make sense of it. For the task of evaluators is not only to explain the world in

the best possible way. We must also interpret it as well. In the best possible way, taking into account the demands from various stakeholders and implementing high-class evaluation ethics. This can be achieved without any of the current evaluation standards. The evaluator's ethical ambition does not need to be built on hyperbolic evaluation standards, but on deliberate individual, ethically sustainable choices and the understanding of their consequences.

NOTE

REFERENCES

Banaji, M., R., Bazerman, M.H., Chough, D. (2003). How (Un)ethical Are You? Harvard Business Review December 2003. Pp.56-64.

Bauman, Z. (1995a). Life in Fragments. Essays in the Postmodern Morality. Oxford: Blackwell Publishers.

Bauman, Z. (1995b). Postmodern Ethics. Oxford: Blackwell Publishers (3rd edition).

Bell, D. (2003). Ethical Ambition. Living a Life of Meaning and Worth. London: Bloomsbury.

Bernstein, P.L. (1998). Against All Gods. The Remarkable Story of Risk. New York: John Wiley & Sons.

Blackburn, S. (1999). Think. A Compelling Introduction to Philosophy. Oxford: Oxford University Press.

Bourdieu, P. (1991). Language & Symbolic Power. Cambridge: Polity Press.

Bourdieu, P., Wacquant, L.J.D. (1992). An Invitation to Reflexive Sociology. Chicago: Chicago University Press.

Bunda, M.A. (1987). Alternative Systems of Ethics and Their Application to Education and Evaluation. In W.R. Shadish et al. (eds.), Evaluation Studies Review Annual (pp. 363-374). Vol. 12. Newbury Park, CA: Sage.

Caputo, J.D. (2003). Against Principles: A Sketch of an Ethics Without Ethics. In E. Wyschogrod et al. (eds.), The Ethical (pp. 169-180). Oxford: Blackwell Publishing Ltd.

Coelho, Paulo (1999). The Alchemist. London: HarperCollins Publishers.

Czarniawska, B. (2001). Is it Possible to Be a Constructionist Consultant? Management Learning, 32(2), 261-266.

Freidson, E. (1994). Professionalism Reborn. Theory, Prophecy and Policy. Cambridge: Polity Press.

Jann, W., Reichard, C. (2003). Evaluating Best Practice in Central Government Modernization. In H. Wollman (ed.), Evaluation in Public-Sector Reform. Concepts and Practice in International Perspective (pp. 36-55). Northampton, Massachusetts: Edward Elgar Publishing.

Lingis, A. (2003). The Immoralist. In E. Wyschogrod et al. (eds.), The Ethical (pp. 197-216). Oxford: Blackwell Publishing Ltd.

Moss Kanter, R. (2003). On the Frontiers of Management. Boston, Maasachusetts: Harvard Business Books.

Popper, K.R. (2002). Conjectures and Refutations. The Growth of Scientific Knowledge. London & New York: Routledge. (Originally published in 1963).

Popper, K.R. (1980). The Logic of Scientific Discovery. London: Unwin Hyman. (Originally published in 1934).

Schwandt, T.A. (2003). 'Back to Rough Ground!'Beyond Theory to Practice in Evaluation", Evaluation 9(3), 353-364.

Schwandt, T.A. (1997). The Landscape of Values in Evaluation. In D.J. Rog et al. (eds.), Progress and Future Directions in Evaluation: Perspectives on Theory, Practice, and Methods (pp. 25-39). New Directions in Evaluation Number 76, Winter 1997. San Francisco: Jossey-Bass.

Scriven, M. (1980). The Logic of Evaluation. Inverness, CA: Edgepress.

Shadish, W.R. Jr., Cook, T.D., Leviton, L.C. (1991). Foundations of Program Evaluation. Newbury Park, CA: Sage.

Shadish, W.R., Newman, D.L., Scheirer, M.A., & Wye, C. (Eds.). (1995). Guiding Principles for Evaluators. San Francisco: Jossey-Bass.

Stern, Elliot (2003). The Rationale and Challenges for Evaluation of Public Policies. Paper presented in the 5th European Conference on the Structural Funds in Budapest, Hungary, 26 June 2003.

Virtanen, P., Uusikylä, P. (2004). Exploring the Missing Links Between Causes and Effects, Evaluation 10(1). Accepted for publication.

Virtanen, P., Laitinen, I. (2004). Beyond Standards? The Applicability and Limits of Evaluation Standards. Submitted for publication to Scandinavian Journal for Spatial Development.