DOES QUALITATIVE INQUIRY IMPLY CONFIRMATION BIAS IN EVALUATION STUDIES?

What Evaluation Studies Reveals about Core Issues in Economics and other Social Sciences: The Qualitative versus Qualitative Methods Debate Revisited

Olav A. Kvitastein

Institute for Research in Economics and

Business Administration

Address:

SNF, Breiviksveien 40,

N-5045 Bergen, Norway

Phone: +47 55 95 95 00

Fax: +47 55 95 98 74

e-mail: Olav.Kvitastein@snf.no

Abstract

The paper discusses to which extent qualitative methods produce confirmation bias in applied evaluation studies. The argument put forward is that the agenda setting nature of the overt superficial "first" question of applied evaluation studies, namely, "did it work?" demands claims about *effects*, regardless of how well the programme or project under investigation lends itself to causal investigations. The evaluation researcher does usually not have the power to influence the discourse in the aftermath of the presentation of results. Ambiguities of results are often taken by stakeholders as an opportunity to set the premises of debates. I argue that the researchers carry a responsibility for minimising the distance between the factuality of outcomes and the interest driven impression of outcomes that may characterise the public discourse following presentation of the evaluation report. Hence, research methods that facilitates causal claims should be preferred because the researcher is embedded in a research context where evaluative claims are demanded and inevitably brought into the policy decisive discourse, regardless of the foundations for such claims. Methods where no conclusion about outcomes and no evaluative claims is the only option, may produce bias simply because the responsibility for conclusive claims are handed over from researchers to stakeholders who should not be assumed to act independent of their own interests. Such interest driven evaluations tends to cause evaluations to degenerate into marketing efforts, i.e. promotion substitutes evaluation. This change implies a category-mistake that invites miscommunication. Consequently, there should be a lower bound on evaluations, indicative of when an evaluation should *not* be carried out. That is, whenever a category-mistake of this type is inevitable or highly probable, the decision *not* to evaluate should be preferred. Conversely, when valid conclusive evaluative claims are likely or at least possible, requests for evaluations should be encouraged. The article discusses how and why I can propose allegation of confirmation bias in different schools of evaluations. It concludes that a pragmatic ontology based on the causal counterfactual approach to probabilistic causation seems to be the most appropriate framework for applied evaluation research.

Introduction

This paper explores the apparent paradox that under the realm of the new public management, qualitative inquiry seems to have won a prominent place in public sector evaluations. Thus, the new managerialism, typically presumed inclined toward rational models and hard quantitative evidence, appears supportive of qualitative methods, usually taken as a mean for radical or in-dept analyses, not as a mean for evaluating outcomes. The consequences of the central place for qualitative inquiry in the evaluation setting are examined.

Despite the conventional wisdom that the relevant measurement in an evaluation of a public programme or project should in one ore other sense objective and related to the stated goals of the programme/project to facilitate assessments of the project/programme's success, this is in usual not the case. Instead, questionnaires are handed out, in-dept interviews are conducted or any other of the researcher's favourite means for gathering information is employed, regardless of how well it is suited for a situation where most participant are prepared for their status as the relevant respondents. The reliability of self-reported measures in the evaluation setting is questioned.

A governing theme throughout the paper is the agenda setting nature of the overt superficial 'first' question of applied evaluation studies, namely, 'did it work?' This question demands claims about *effects*, regardless of how well the programme or project under investigation lends itself to causal investigations. Ambiguities of results are often taken by stakeholders as an opportunity to set the premises for debates. Accountability means minimising the distance between the factuality of outcomes and the interest driven impression of outcomes that may characterise the public discourse following presentation of the evaluation report. Hence, research methods that facilitate causal claims should be preferred because the researcher is embedded in a research context where evaluative claims inevitably are demanded and brought into the policy decisive discourse, regardless of the foundations for such claims. Methods where no conclusion about outcomes can be produced may invite bias simply because the responsibility for conclusive claims are handed over from researchers to stakeholders, who can not be assumed to act independently of their own interests. Such interest driven evaluations may cause evaluations to degenerate into 'perception management'.

The article discusses how strong directional pressures and the framing effects of the New Public Management mantra mould methodologies into providing 'acceptable' measures via institutional processes that are interpretable, but could have revealed fewer demonstrations of human weaknesses provided that more powerful methodologies had been employed. In particular, one of the direct results of the framing effects of the New Public Management is over-confidence in techniques like cost-benefit analysis, a paradigm long demonstrated to be founded on faulty assumption. The article concludes that an open-system ontology that allows for a range of understandings and theorising, non-reactive or unobtrusive measures together with analysis based on the counterfactual approach to probabilistic causation in general seems to be the most appropriate framework for applied evaluation research.

Outline of Article

In line with the nature of evaluation research, the article sweeps a broad range of disciplines for the construction of arguments. The introductory part presents my view of the state of evaluation research and its relations to social science in general. The second part introduces the research questions and the third part outlines my view of evaluation as embedded in the context of the New Public Management movement. The fourth part gives a brief theoretical background for why I think confirmation bias can be a threat to evaluation studies based on qualitative methodologies. The fifth part presents my interpretations of the qualitative – quantitative debate and gives a crude overview of the various methodological positions. The discussion part sums up why I think qualitative methods are not the best choice when the quest for effects or impacts dominates and also discusses some presumed consequences of evaluation as a field of research split up in many different 'schools' of research. The concluding remarks give a brief summing up.

Evaluation and the Social Sciences

The conception of social science implied by applied evaluations, that researchers do not identify problems, they simply solve predefined task, is hard to accept for many scholars. Compliance and resistance toward decreasing availability of funding for general social research, combined with growing opportunities for finance via evaluation contracts, take many forms. The formation of professional organisations like *European Evaluation Society* and the *American Evaluation Association* and the birth professional journals dedicated to evaluation studies, are all positive adaptive institutional responses. The tension between the research communities within the established disciplines and the emerging sub-discipline dedicated to evaluation

research is, however, the other side of the coin. Even though evaluation research is well accepted as a legitimate research activity, hints to payoffs for opportunism and insinuations about biases in reported outcomes of evaluations, is not entirely absent. For the legitimacy of evaluation research, it is of importance to reveal the alleged sources of bias.

As pointed out by several prominent scholars (Aaron, Gramlich, Hanushek, Heckman, & Wildawsky, 1990; Haveman, 1987; Nathan, 1988) the status of the social sciences has been on a downward slide for more than thirty years. It may seem paradoxical that evaluations flourish in times when the social sciences in general have a downturn, and it is not all that clear that the new confidence in evaluations and hence, social science based public policies, implies a renewed vote of confidence in the social sciences. As an integral part of the managerial orientation of the New Public Management, the awake of evaluations is interpretable, but badly in need of clarifications. What used to be called a report is now an evaluation, and thus, carries a promise of being something more than just a report. Evaluation is a semantic magnet (Vedung, 2000) with a positive power that lends itself easily to a message of confidence. On the other hand, the *American Evaluation Association* is, to my knowledge, among the very few professional organisations that openly admits to and publicly discuss that negative reputation is a problem for their profession (Donaldson, 2001). Recently, some evaluators also have expressed frustration over the tension between this quest for confidence and the feeling that a widespread acceptance of relativism characterises both practical evaluation reports and the professional recommendations from leading journals in the field. With respect to methodology, the impression is that 'anything goes' (Adelman, 1996).

However, applied evaluations have a built-in propensity to reveal deep-seated problems in the social sciences. The practical, interest ridden setting of applied evaluation research, tends to unveil unpleasant questions about ideological underpinnings of the theories in use, and ambiguities of the methods employed. Practical implications of the theories that guided interventions may disclose unwarranted side effects, and the prescribed methods may fail to provide trustworthy information about outcomes. Frustration in the research communities seems to disperse in two directions, towards overconfidence or retreat. Overconfidence tends to surface as an expression of near unconditional faith in theories in support of the arguments set forth and retreat can be described as a backfire of the researcher's methodological training. When the prescribed remedies fails, the researcher renounces, not only the recommended methods, he/she rejects the entire role as a researcher and take flight into roles apparently

more favourable, such as, say, the role as a judge. The canon of opportunism in this respect, is the so-called 'fourth generation evaluation' (Guba & Lincoln, 1989) where the idea of evaluation as the search for quality, merit, worth, etc., is rejected in favour of the idea that it is negotiation which is the issue (Scriven, 1993). That is, negotiation between stakeholders with different interests or world-views is the essence of program evaluation.

I have considerable sympathy for the creativity and insight of Guba and Lincoln's work (Guba, 1990; Guba & Lincoln, 1989). With the exception of their open prejudice to an undefined group of their fellow evaluation researchers who they label 'positivists', the Fourth Generation Evaluation (Guba & Lincoln, 1989) is worth reading for anyone interested in research methodology. The 'positivists' they describe in the following way: "Convinced that there exists some single, true reality, driven by natural laws, open to discovery and harnessing by the methods of science, positivists reject all relativist views, of which constructivism is one, as not only seriously in error but pernicious and repugnant" (Guba & Lincoln, 1989:16). Clearly, by negatively attaching a historical philosophical position, or more precisely, a set of related philosophical positions, to a generalised third person, the 'positivist' evaluator, they create a middleman, made up for the sake of the argument. By doing so, they tell us that constructivists not only reveal and recognise the existence of social constructions. They also create them. Indeed, people constantly create and communicate social constructions and it is of great importance that we seek to understand and unveil the underlying processes. In particular, I recognise the need to understand social constructions since I consider most man-made abstract systems, including such systems as logic and mathematics to be social constructions. As such, they are systems or constructions of great value to, among other uses, evaluations.

The Fourth Generation Evaluation constitute one of the two extremes of evaluation; on the one hand, the assumption-dependent devotees of the cost-benefit methods of neo-classical economics, on the other hand, the followers of constructivist inquiry, supposedly dependent on fewer unrealistic assumptions concerning human nature and behaviour. Both traditions have problems with the empirical contents of their analysis and they both run the risk of being victims of dominant stakeholders, simply because of predictable methodological flaws. While the Fourth Generation Evaluation approach is capable of answering questions like "Can we reach an agreement about what we did in this programme?" cost-benefit analysis is competent of answering questions like "What did it cost"? None of the two approaches can provide a trustworthy answer to questions concerning the effects of the programme/project under con-

siderations. Hence, very little can be said about goal-achievement. Consequently, the benefit side of the cost-benefit ratio is at best, dubious. Another reason for taking these two traditions as extremes is that they both seem overtly convinced of the supremacy of their positions and that they are located on opposite sides of the qualitative- quantitative divide.

In line with (Reichenbach, 1938) I agree that qualitative inquiries in usual are better suited for theory development and new discoveries than most quantitative methods that seems more correctly applied to confirm what is already indicated by other methods. In some cases, however, for instance a change in infant mortality, discoveries depend heavily on quantitative methods. My point is that even though evaluations indeed can provide an opportunity for new discoveries, the policy making side of the evaluation process unequivocally asks for effects and goal-achievements. There is little room for new discoveries and reinterpretations of the programme/projects in question, even though policy-making processes could benefit from new insights, provided they were able to absorb them. I argue, however, that recently developed quantitative methodologies are far better suited than qualitative methodologies for most evaluations, in particular in situations where the cry for documentation of impacts of public programmes or projects dominates. Substituting qualitative methods for quantitative methods as a matter of principle does not promote the reputation of applied evaluation.

Research questions

Is the use of Qualitative Inquiry a Source of Bias in Evaluation Research?

I believe that the qualitative-quantitative dichotomy in many respects is a false dichotomy that stimulates an unfortunate debate. I do, however, think there are some features of this debate that may help us to identify sources of what I have labelled *confirmation bias* in evaluation studies. The term confirmation bias signals that I hypothesise a unidirectional bias. That is, the bias goes mainly in favour of dominant stakeholders. Hence, I contend that the context of evaluations tend to influence the outcomes of evaluations and that researches who see qualitative methods as the only valid way, give away means for resisting pressure towards concordance. That is, I do not suggest any difference between the devotees of qualitative inquiry and other researcher with respect to moral courage. I simply suggest that the qualitative researcher lack the opportunity to appeal to method as an independent judgement. The lack of support from method as a 'third person' may be of importance in situations with conflict be-

tween researchers and stakeholders, however weak the distinction between the researcher and this 'third person' might be. Hence, my second thesis of inquiry is:

II. Qualitative inquiry may produce confirmation bias in evaluation studies by amalgamating the research and the researcher, and hence confusing psychological self-defensive mechanisms and professional argumentation. Thus, conflict may be experienced as insult and support for argument as praise, constituting a mechanism that increase the probability of agreement between the researcher and the dominant stakeholders.

By leaving little or no room for methodology, the researcher risks to be understood as a lawyer who continuously writes and rewrites the law he/she practices. Needless to say, this is a situation where any blame will be directed towards the person, not the rules.

Is the Devotion to 'Schools' of Evaluation a Source of Bias?

Also, the researcher's declaration of confession to a specific sub-discipline or 'school' of research may serve as a source of bias insofar that loyalty limits the range of valid outcomes. Thus, by excluding arguments that do not conform to the epistemological and ontological basis of the 'school', the researcher may cause disciplinary loyalty to become a source of bias in the direction of central beliefs held by the 'school'. The bias is hypothesised to be in the direction of 'within school consensus', regardless of whether this is in line with the interest of dominant stakeholders or not. Hence, my second thesis of inquiry is:

III. Strong commitment to a specific 'school' of thought may produce bias in a direction that ensure paradigmatic support.

Noteworthy, this thesis reflects my impression that at least some qualitative methodologies seem to be more directed towards paradigmatic support than towards traditional investigation. That is, some methodologies seem to search for more theory-laden observations than others. I do, however, accept the notion that theory-laden observations are hardly entirely avoidable regardless of methodologies. By the same token, actively seeking observations that are good candidates for confirmation and excluding observation that could lead to refutation violates the very notion of doing research.

The Scope of the Questions Proposed

The argument that qualitative inquiry may produce confirmation bias in evaluation studies is only taken to be valid insofar that a question about outcomes, i.e., the effects of a proj??ect/programme or its merit in other respects is invoked. That is, the argument may be applicable to both formative and summative evaluations, but for formative evaluations only to the extent the relative merit or worth of the project/programme is questioned. For summative evaluations where the demand for evaluative claims is explicit, problems concerning bias are indubitably of importance. For evaluations where the intention behind the evaluation is of a different nature, say, program adjustment based on analyses of implementation processes or other more intermediate concerns, problems of bias in conclusions may still apply, but is of a different nature. Thus, producing exaggerated expectations is different from exaggerating claims about outcomes. By the same token as the distinction between formative evaluations, process evaluations, summative evaluations, effect evaluation is fuzzy due to a great many different uses of these two concepts.   For the sake of simplicity it is convenient to restrict the scope of or arguments to be valid for summative evaluations only.

The Context of Public Evaluations

The Move from Bureaucratic to Managerial Control
While the twentieth century was governed by the principle of a politically neutral civil service offering impartial policy advice to the elected government of the day, the last part of the 20th century has witnessed an increasing importance of public servants in process of policy for-mulation and implementation (Plowden, 1994). The demise of Weberian bureaucracy and the belief that a stronger, more competent and vigilant bureaucracy is the result of modernity, is among the basic premises behind many newer governmental reform initiatives.
Over the last three decades the New Public Management initiatives has produced fundamental and ubiquitous institutional change in the nature of public administration in most western in-dustrial democracies. These changes have had a variety of consequences, some of them of considerable concern for evaluation practices. The paradigmatic essence of the new public management is reducing and deregulating bureaucracy, using market mechanisms and simu-lated markets to conduct government action, devoting responsibility downward and outward in organisations, increasing productivity, energising agencies, and empowering employees to pursue results, improve quality, and satisfy customers (Carrol, 1998). The 'worldview' of the

New Public Management presupposes that 'something' fundamental happened in the 1980s that changed the field of public administration. The (assumed) appearance of new forms of governance, new relationships between citizens and their government and between the public, private, and non-governmental sectors fundamentally altered the processes of policy making. The nature of this assumed change is explicitly expressed by former U.S. Vice-President Al Gore, when he claims that Americans view themselves as customers of the government rather than as citizens (Gore, 1993).

The new orientation implies the substitution of self-interests for the more complex norms of traditional bureaucracy as the baseline for the design of governance, and the substitution of the customer for the citizen as the basic individual unit of democratic society. These changes have altered the core concept of evaluation, *accountability*, from its many-valued meaning *democratic accountability* to the single-valued meaning *economic accountability*, and thus, brought new topicality to look at evaluations as an *agency problem.*

Agency theory assumes opportunism in the relationship between the leader and the subordinate, i.e. the principal and he agent. I do not believe that traditional bureaucracy was devoid of opportunism. My basic point is that, under the realm of the new public management there may be, at least at the individual level, a better payoff for opportunism, in particular for the top civil servants (Lægreid, 2000). A market-based system for evaluation contracts implies agency problems. By its reliance of self-interests as the driving forces of governance, the new public management acknowledges self-interest as a more legitimate concern for the individual than was the case under the traditional bureaucratic regime. This may be viewed as a concession to opportunism and may interfere with the choice of research strategies. To understand some implausible evaluation outcomes it is of importance to come to grip with the interplay between evaluation methodologies, the institutional mechanisms of research practices that pertain to the various methodologies, the emerging routines of evaluation practices, and the potential rewards or penalties for opportunism.

Agency theory – the Heart and Soul of the New Public Management
Agency theory (Jensen & Meckling, 1976), the analysis of principal-agent relationships, in which one person, an agent, acts behalf of another person, a principal, lays at the heart of the new public management and can be viewed as the dominant idea behind structural reforms. Hence, when consulting companies and research institutions are competing for evaluation

contracts, the contest for contracts are in line with the ideological underpinnings of the new managerialism. It is an implicit, albeit naive assumption, that competitive bidding guarantees the best quality in the evaluation process. Set apart from overlooking the differences in normative traditions between consulting companies and research institutions, this line of thinking confuses the costs of the governmental contract with the costs of the consequences of the evaluation task. Also, this line of reasoning fails to recognise that the very same contractual theory that justifies the competitive bidding process can be applied to the evaluator – evaluation management relationship. Whenever a contractual relationship can be identified, agency theory can be applied. From a methodological point of view, there are very few reasons to believe that the less costly evaluation contract produce more reliable results than the more expensive contract. It is more likely that minimum funding for the evaluation task will induce methodological shortcuts, thus undermining the trustworthiness of results. On the other hand, an abundance of research funding does not guarantee the quality of evaluations. Hence, the assumption that competition is a quality optimiser does not apply. The agency framework applied to the contractual relation between evaluator and evaluation management may, however, shed some light on the problem. It is, however, a peculiar feature with this relation, namely that it cannot be understood as a principal agency relationship *without violating the fundamental rationale for undertaking the evaluation task*. The basic rationale for evaluations is the *independence* between the researcher and the evaluation management.

The Process of Evaluations

Psychological and Institutional Theories – Concessions of Human Fallibility

To come to grip with a fuller picture of, *a)* the processes that are candidate for producing bias and, *b)* the processes that make it understandable that biased results are accepted by governmental agencies, we have to consider processes at both the individual and the institutional level. Clearly, these processes are related, but for the purpose here, it is convenient to split the discussion into two parts. The *first part* concerns the relationship between the researcher and his/hers interpretation of data. This involves not only the peculiarities of the individual researchers but also traits that are present in the research community that embeds the researcher. The *second part* concerns the stakeholders in the evaluation process itself. That is, the evaluation management and other relevant stakeholders, the policymakers and other parts of the government agencies involved in or affected by the outcomes of the evaluation.

The Researcher an the Data – the Brain Alone Perspectives

Much of the early psychological reasoning was founded on 'the brain alone' principles. The behavioural research embraced an input-output model linked by internal conduit that makes behaviour possible but exerts no influence of its own on behaviour. Human behaviour was shaped automatically and mechanically by environmental stimuli (Bandura, 2001). Even though the behavioural perspective is much in line with the basic view of, or assumptions about the actors in economic agency theory, the behavioural perspective is of limited value for understanding the researcher at work. The next phase in the development of mainstream psychology can roughly be labelled 'the cognitive revolution'. Although the brain is still much alone, this line of theorising that coincided with the advent of the computer, filled the mind with a lot of computational operations and inventive thoughts that is easier to associate with the researcher. However, the cognitive perspective also revealed a lot of human rational frailty that give rise to concepts like 'selective perception' (Dearborn D. & A., 1958) 'bounded rationality' (Simon, 1982) and 'framing' (Kahneman & Tversky, 1979; Tversky & Kahneman, 1986), concepts that provides insights that also concerns the researcher's work processes. Bounded rationality recognises the limitations of human cognitive capacity and need for heuristics in order to cope with large amount of information. This implies that researcher's tends to develop routines to cope with information overload, routines that imprint their work regardless of their ability to account for habits or procedures. Selective perception is, among other things, relevant to the debate over 'theory laden observations' (Kuhn, 1962). 'Framing' is relevant for e.g. understanding the effects of problem formulations in tender documents, i.e. how the evaluation management may impose specific 'world views' upon the researcher, 'world views' that implies assumptions that the researchers accepts, hesitant, willingly or unconsciously accepts.

Newer psychological research has progressed beyond the 'brain alone' perspective and has long recognised that people are social beings. One line of research has gone in the direction of microanalysis of the mind in processing, representing, retrieving, and using the coded information to manage various task demands, another line of research focus on the macro-analytic workings of socially situated factors in human development, adaptation, and change (Bandura, 2001). The first line of research may be of great value for understanding some features of the research process that may produce bias, such as 'conceptual tunnel vision' e.g. the qualitative researcher's tendency to over-categorise data i.e. of assigning more data to one

category than actually belongs, or the 'pink elephant paradox', that once the idea of a pink elephant is mentioned, it cannot be erased from one's consciousness (Morse et al., 2002). The second line of research implies social cognition and is better suited for understanding the foundations for the institutional processes that may function to produce and preserve bias in evaluation practices.

Interpersonal Processes and Psychological Theories of Expectancy Confirmation

It can be safely assumed that the researcher approaches the study of individuals involved in or affected by the project/programme in question with some pre-determinate ideas or hypotheses about what reactions or experiences that are central to the evaluation. The quantitative oriented researcher may express his/hers ideas in the form of written questions in a questionnaire or as instruction in an interview guide. The qualitative oriented researcher is, however, less assertive of own capabilities to settle in advance what the right questions might be, and want to figure this out by interacting with the people involved, either via direct conversation or by other means that allow for the most direct communication. Interaction and direct involvement with those affected by the public project/programme in question, seems to be a salient feature that distinguishes qualitative inquiry from quantitative inquiry. Clearly, the personal features of the qualitative inquiry procedures are in line with cognitive psychology reasoning insofar that it recognises limited cognitive capacity. The qualitative methods literature is, however, more reluctant to incorporate newer research on expectation confirmation in interpersonal relations.

The most well-known type of expectation confirmation is the self-fulfilling prophecy (Merton, 1948) where our expectations about others, whatever their origins, tends to elicit the very behaviour that is expected. Coincidentally, this line of research has much in common with evaluation research, as they both originate from studies of the primary school system. One of the first hints of expectancy confirmation was the demonstrations that teachers that was led to expect particular levels of performance from students in their classroom, acted in ways that elicited performances that confirmed the initial expectations (Rosenthal & Jacobson, 1968). More recent research show that the processes of expectancy confirmation and disconfirmation involves a complex intertwining of cognitive, motivational, and behavioural activities in social interaction (Snyder & Stukas, 1999) but also that the phenomena follows traceable paths that makes it possible to map the mechanisms at work. Decomposition of the elements of the mechanisms into a series of steps suggests the following sequence: *(a)* per-

ceivers adopt beliefs about targets; *(b)* perceivers behave toward targets as if these beliefs were true; *(c)* targets fit their behaviour to perceivers' overtures; and *(d)* perceivers interpret targets' behaviour as confirming their beliefs (Kelley, 1992). Expectations can be categorised according to their *properties*, such as certainty, accessibility, explicitness, and importance. Increases in one of these properties should lead perceivers to increase their tendency to act on expectations in ways that increase the likelihood of confirmation (Olson, Roese, & Zanna, 1996).

Institutional theory

Institutional theory (Powell & DiMaggio, 1991; Scott, 1995) offers considerable insights for understanding evaluation processes. It is, however, not easy to comprehend institutional theory as an unambiguous coherent theory. The concept of an *institution* has been used in different ways by numerous authors, and to cover diverse phenomena. In accordance with the purpose here, we use W. Richard Scott's omnibus definition: *"Institutions consist of cognitive, normative, and regulative structures and activities that provide stability and meaning to social behaviour. Institutions are transported by various carriers – cultures, structures, and routines – and they operate at multiple levels of jurisdiction"* (Scott, 1995:33). In Scott's conceptualisation, institutions are multifaceted systems, including symbolic systems, cognitive constructions and normative rules, and regulative processes carried out through and shaping social behaviour (Scott, 1995). The view of institutions as both systems and processes facilitates discussions of the interplay between the overarching ideas from economic theory that settle what the operating notion of evaluation should be and the design of the evaluation process. It also provide for a way to interpret the cognitive mechanisms that make diverse empirical representations converge across researchers within homogeneous subgroups when methodological rules or norms are unclear or absent. Noteworthy, even though constructed and maintained by the individual, *institutions assume the guise of an impersonal and objective reality.* Institutional mechanisms require little or no conscious mobilisation of will or effort (Scott, 1995) and hence, facilitate discussions of the pitfalls of various research strategies without invoking accusations of deliberate distortions of assumed empirical representations.

Scott's brilliant summing up of institutional theory provides a way to systematise the many active mechanisms of institutional processes that is suggested in the vast literature on the subject. Table 1 gives a sketch of the *basis of compliance*, the *mechanisms* at work, the *logic*

of the particular process, the *indicator* for the process and the *basis of legitimacy* according to what Scott labels the *three pillars of institutions*.

Table 1. The Three Pillars of Institutions

	Regulative	*Normative*	*Cognitive*
Basis of compliance	Expedience	Social obligation	Taken for granted
Mechanism	Coercive	Normative	Mimetic
Logic	Instrumentality	Appropriateness	Orthodoxy
Indicators	Rules, laws, sanctions	Certification, accreditation	Prevalence, isomorphism
Basis of legitimacy	Legally sanctioned	Morally governed	Culturally supported, conceptually correct

Source: (Scott, 1995): 35)

Scott's table sweeps thousands of pages of research, and his conceptualisation has a strong intuitive appeal. Although there are no well defined boundaries that allows for precise definitions of the various concepts introduced, it is easy to grasp each and every concept as *descriptive of a social relation* that is straightforward to imagine when we think about how an evaluation process is unfolding. We can reflect about the relations between the evaluation management and the evaluator, or we can envision the relations between a group of researchers with different personalities. Scott's conceptualisation provides a rich source for speculations and reflections about what the relations could be and what they could produce of outcomes that are simply due to the nature of the relations alone. This way of using concepts to induce reflections over relations (Bourdieu & Coleman, 1991; Cassirer, 1910) has a distinct European sociology flavour, is easily applicable for discussing hypothetical configurations of relations, and is indicative of Scott's unusual synthesising capacity.

The three pillars cover diverse scholarly approaches the discussion of institutions and different understandings of institutional processes. The *regulative* pillar clearly resembles the approaches typical of economists and economic historians, the *normative* pillar mostly that of sociologists and the *cognitive* pillar mirrors the approaches most likely to be found among psychologists and organisation theorists.

Scott's (1995) typology also contains suggestions about the factors or types of repositories or 'carriers' that sustain or reproduce institutions. *Cultures* as carriers transmit schemes that inform and constrain behaviours, *social structures* carry expectations connected to networks of social positions and role systems, while *routines* carry habits, standard operating procedures and other repetitive behaviours or trained rigidities.

Table 2. Institutional Pillars and Carriers

	Pillar		
Carrier	*Regulative*	*Normative*	*Cognitive*
Culture	Rules, laws	Values, expectations	Categories, typifications
Social structure	Governance systems, power systems	Regimes, authority Systems	Structural isomorphism, Identities
Routines	Protocols, standard routines	Conformity, performance of duty	Performance programs, Scripts

Source: (Scott, 1995): 52)

The carriers provide yet another way for reasoning about likely mechanisms or processes behind outcomes. When we observe outcomes that most likely are biased in one or another direction, it is possible to make constructive speculations about what processes would be the most likely candidates for producing such a result, without the requirement that we are able to directly observe the process in question. That is, we use institutional theories to provide a framework for speculations and discussions about institutional processes at the organisational or the research community level, not as guiding devices for empirical observations.

Thus, we take agency theory as indicative of the dominant underlying epistemology of the New Public Management. In its most general form, agency theory implies a coherent epistemological understanding that leaves little grounds for choosing among perspectives that excludes e.g. general equilibrium theory, welfare theory or other understandings that conflict with mainstream economics. This is the feature of the New Public Management and of agency theory that provide for the strong framing effect of the perspective chosen. Provided that we do not accept lower methodological standards for evaluation research than for discipline research, we should start looking for sources of bias where they are most likely to be found. That is, in the human frailty as revealed by psychological and institutional theories. Given the strain of the task of conducting unbiased judgement conflict, human fallibility should be expected to show up in most evaluations.

Methodologies and Disciplines

The Qualitative versus Quantitative Methods Debate

More than two decades ago John Van Maanen stated that "the label *qualitative methods* has no precise meaning in any of the social sciences" (Van Maanen, 1979:520). The absence of a precise meaning does not, however, disqualify this loosely connected bundle of methods,

"since qualitative researchers tend to regard social phenomena as more particular and ambiguous than replicable and clearly defined" (Van Maanen, 1979). By the same token, it is not so easy to give an exact account of what methods should qualify for the label *quantitative methods*. Further, it is no reason to assume that researchers who use quantitative methods *a priori* take social phenomena to be less complex and ambiguous than the qualitative researcher. It is, however, probably not entirely wrong to say that, despite that many assertions that qualitative and quantitative methods are not mutually exclusive and in usual should strengthen analysis when combined, the tension between the devotees of each camp have not levelled off over the years. On the contrary, many new textbooks express attitudes that come close to hostility towards any use of numbers besides necessary paging. The deadly serious, humourless rhetoric employed hints to a hermeneutics of suspicion[1], indicative of a linguistic turn[2], where, however, the text to be dissected is missing.

Grounded Theory (Glaser & Strauss, 1967) had a tremendous impact after its launch in the late sixties and throughout the seventies and eighties several books intended for instruction and also straight textbooks on qualitative methods were introduced. By the mid nineties the harsh criticism of quantitative methods, a subject that had a substantial place in most of these textbooks, had become 'internalised' in many business schools, departments of sociology, political science and other social science departments to an extent that students saw qualitative methods as just a different path of methodological development. As a path of methodological training, qualitative methodology came to be viewed as equivalent to quantitative methodology, a bit more 'modern' and the preferred choice for the more 'theoretical inclined' student. It also carried the advantage that one could skip some tedious statistics classes. – I do not know of any exact figures, but it is my impression that the percentage of students that graduate from social science departments without basic training in statistics, is increasing. This may affect the future of evaluation research.

 Many of the newer textbooks do, however, give valuable contributions to the comprehension of methodologies and understanding of the limitation of methods. Some are extremely well organised, e.g. (Flick, 2002) and others open new perspectives for the evaluation researcher

---

[1] This term was coined by Paul Ricoeur (1970:27) to describe the three key intellectual figures of the twentieth century, who, in their different ways, sought to unmask, demystify, and expose the real from the apparent, namely Marx, Nietzche, and Freud, the leading figures of the school of suspicion.
[2] Attributed to the excellent, humorous phrase by the econometrician Arjo Klamer (Klamer, 2001) "It was about then that I made my linguistic turn".

by relating both research strategies and methods to established research traditions (Creswell, 1998). Many of these books could indeed be included on more curricula. The point is that qualitative methods open for insights that should *add* to the researcher's toolkit, not *substitute* other knowledge. As additions new methodological perspectives should receive a warm reception. As alternatives that expel and replace existing competence, they can be erosive.

A particular feature of many textbooks on qualitative methods is the overwhelming number of methodologies and perspectives introduced. Creswell (1998) encourage students to become familiar with the research tradition of *biography*, *phenomenology*, *grounded theory*, *ethnography* and *case studies*. Instructions for all five traditions, usually thought in different courses to students of philology and seldom offered to social science students, are all covered in a single textbook. As shown in table 3, the different perspectives introduced cover research traditions that require a wide range of training and skills. *Biography* is something that is usually understood as quite apart from what in general occupies the social scientist. The focus is on the life of an individual, a theme that is even further from the traditional tasks of the evaluation researcher. *Phenomenology*, a tradition heavily criticised for departing from

Table 3. Dimensions for Comparing Five Research Tradition in Qualitative Research

Dimension	**Biography**	**Phenomenology**	**Grounded Theory**	**Ethnography**	**Case Study**
Focus	Exploring the life of an individual	Understanding the essence of experiences about a phenomenon	Developing a theory grounded in data from the field	Describing and interpreting a cultural and social group	Developing an in-depth analysis of a single case or multiple cases
Discipline origin	Anthropology Literature History Psychology Sociology	Philosophy, Sociology, Psychology	Sociology	Cultural anthropology Sociology	Political science, sociology, evaluation, urban studies, other social sciences
Data collection	Primarily interviews and documents	Long interviews with up to 10 people	Interviews with 20-30 individuals to 'saturate' categories and detail a theory	Primarily observations and interviews with additional artifacts during extended time in the field (e.g., 6 months to a year)	Multiple sources-documents, archival records, interviews, observations, physical artifacts
Data analysis	Stories Epiphanies Historical content	Statements Meanings Meaning themes General description of the experience	Open coding Axial coding Selective coding Conditional matrix	Description Analysis Interpretation	Description Themes Assertions
Narrative form	Detailed picture of an individual's life	Description of the 'essence' of the experience	Theory or theoretical model	Description of the cultural behaviour of a group or an individual	In-depth study of a 'case' or 'cases'

Source: (Creswell, 1998)

Husserls's original intentions, is included with a focus on understanding. *Grounded theory* is introduced as a methodology with a focus on theory development, *ethnography* with a focus on cultural interpretation, and *case study* methodology is introduced as a mean for in-depth analysis of single or multiple cases.

The ambitions of a training project of these dimensions are praiseworthy, even though the realism of gaining thorough understanding of all these traditions can be questioned. Indeed, my experience from discussions with scholars from ethnography and cultural studies indicates that many of the courses offered in business schools tends to imprint the content of courses in ethnography and phenomenology with instrumental epistemologies, alien to the original theories. Case studies[3], on the other hand, are more familiar to the business student, but mostly applied as a teaching device, to a lesser extent as a research methodology.

A very informative way to organise and understand the different methodologies is introduced by Morgan (Morgan & Smircich, 1980). The different methodologies are organised along a continuum from subjective – to objective approaches to social science, differentiated by their assumed core ontological assumptions and their basic epistemological stances. His rough typology is a helpful device for discussion, although he has added some spice to the debate, in particular by using the term 'concrete' quite frequently and using Skinner (Skinner, 1953) as an example of the 'positivist'.

Combing table 3 and table 4 opens up large fields of inquiry that should inspire imagination and clarify the same phenomenon could look very different dependent on how we choose to observe it. The data analysis part of table 3 gives a rough impression of how one should go about data collection, and a less than lucid guidance to data analysis. Table 4 demonstrates that explicating core ontological assumptions and basic epistemologies clarifies both the choice of research strategies and what kind of question that will have priority in a given study.

Clearly, questions concerning the impact of a given project of programme do not have a high priority in a qualitative methods framework. On the contrary, questions concerning *understanding* of the nature a given project or programme seem to have a high priority. Also theory construction seems to be regarded an important task in qualitative inquiries.

Table 4. Basic Assumptions Characterising the Subjective – Objective Debate within Social Science

	Subjectivist Approaches to Social Science					**Objectivist Approaches to Social Science**
	◀					▶
Core Ontological Assumptions	Reality as a Projection of Human Imagination	reality as a social construction	Reality as a Realm of symbolic Discourse	Reality as a Contextual field of Information	reality as a concrete process	Reality as a concrete structure
Assumptions About Human Nature	man as pure spirit, consciousness, being	man as a social constructor, the symbol creator	Man as an actor, the symbol user	Man as an Information Processor	man as an adaptor	man as a responder
Basic Epistemological Stance	to attain phenomenological insight, revelation	to understand how social reality is created	to understand patterns of symbolic discourse	To map context	to study systems process, change	to construct a positivist science
Some Favoured Metaphors	Transcendental	language game, accomplishment text	Theater Culture	Cybernetic	Organism	Machine
Research Methodology	Exploration of pure Subjectivity	Hermeneutics	Symbolic analysis	Contextual analysis Of Gestalten	Historical Analysis	lab experiments, surveys

Source: (Morgan & Smircich, 1980)

The Measurement Question

In 1966 Eugene Webb, Donald D. Campbell and their associates (1966) made important contribution, apparently long forgotten by the evaluation research community, by introducing the term '*unobtrusive measures'*. Webb et al (1966) defined *reactivity* as *obtrusiveness*, thus making the search for unobtrusive measures the search for methods that do not affect or distort the data that are collected.

Both quantitative and qualitative methodologies presuppose the participation of individuals as research objects. Thus, regardless of the data collection methods being interviews, surveys, rating scales or open questions in questionnaire based studies, people have to volunteer to participate in the endeavour of data collection. For this process to succeed, there must be consent among the participants that it is worthwhile to offer the time and effort necessary to help the researcher to gather the data that is needed for the study. In the psychology literature this spurs a variety of questions, some of ethical nature e.g. (Atwell, 1982), others of meth-

---

[3] Noteworthy, however, Creswell (1998) has no reference to Eisenhardt (1989), the elsewhere most cited article on the theory building aspect of case studies.

odological character e.g. (Rosenthal, 1976). It is undoubtedly unethical to force people to participate in a study and it is questionable to engage people in studies where they disagree with the purpose of the study or would not have participated had they knew the purpose of the study. Clearly, attitudes toward a study may also affect how people choose to relate and react to various questions. Thus, the way a given study relates to its study objects is a methodo-logical problem and a potential source of bias. For evaluations this is a crucial question. In many projects/programmes it is obvious that those who participates, also benefits from the program. When asked to respond to questionnaires, interviews, personal or by phone, they will know how to respond in ways that are in their own interests. Hence, measured against a relevant control group it is very likely that a difference may show up, a difference that in many cases would be taken for granted as an effect of the project/programme in question. It is very difficult to untangle the reactivity component from the answer respondents had provided given that they did not know how to answer in their own best interests. Thus, even with the best psychometric methods it is hard to separate the method effect, i.e. the reactivity of meas-ures, from the 'true scores' differences. With qualitative methods, the problem of reactivity is likely to become even more severe. In the qualitative case it is the sole judgement of the re-searcher that decides what is the reactivity component and what is the respondents true expe-rience. What the respondent says because it serves his own interest or because he/she thinks it may please the researcher is a well-known source of ambiguity. What the respondent really thinks of the project/programme in question provided that he has given it any thought at all, or is just reacting to the researcher's point of view, is often not all that clear. In the case where the respondents answer coincides with what the researcher anticipate, it is reasons to assume that accept the information on face value. The measurement question in applied evaluation is not much discussed. I contend that the reactivity of measures in the evaluation context poses a formidable challenge. Contemporary practices seem to overlook that although many meas-ures, both quantitative and qualitative, may provide considerable insights for many purposes, their reactivity in the evaluation context render many measures *useless for effect studies*.

Discussion

Methodologies, Ideas and Basic Beliefs

In accordance with many other scholars e.g.(Morgan & Smircich, 1980), the article maintains that qualitative- quantitative divide is a misnomer that covers up a variety of underlying

ideological[4], epistemological and ontological questions. When such questions are clarified, the link between methodologies, i.e., strategies of inquiry and method, i.e., techniques of investigations, becomes more transparent and explicit. The concept of 'ideology' as used here refers to ideas that sticks to the mind regardless of the extent to which they have been proven erroneous. It does not necessarily have to be associated with a specific political ideology and it will be used interchangeable with the term 'idea'.

The debate over qualitative versus quantitative inquiry in evaluations research has mainly evolved around an assumed strong tie between paradigms and methods. The notion of a paradigm is in most cases directly inspired by Kuhn's (Kuhn, 1962) concept, although discussions seems to reveal different interpretations of Kuhn's conceptualisation. Mostly, the term paradigm seems to be taken as synonymous with 'philosophical world view'. This expansion of the concept unfortunately serves antagonism and fuels debates more effectively than the more unpretentious idea of paradigm as "the psychological phenomenon related to believing that the description or explanation is correct, and the sociological phenomenon surrounding the co-ordinated enterprise of instrumentation, graduate education, textbook writing and reading, and 'problem-solving' according to the suggestions of the theory's agenda", called 'normal science' by Kuhn (Poslby, 1998:202). A second misinterpretation of Kuhn is the assertion of a close connection in general between a single paradigm and an entire academic discipline, an interpretation that depart substantially Kuhn's from intentions but is well suited for justifications of sharp delimitation between disciplines or sub-disciplines. Also, a small and rather insignificant sub-field may look more prominent when it is introduced as a new paradigm, alternative to a larger, established discipline. At any measure, it is hard to be confident that it is the question of qualitative versus qualitative inquiry that constitutes a paradigmatic divide. It seems more to be the case that the debate over paradigms is used a rhetoric device for the sake or the argument. Adding descriptions of the epistemological and ontological beliefs presumed to be essential to the various 'philosophical world view' a debate where polarisation is a likely outcome is generated since a 'straw-man' extreme and representative of historical well-known philosophical positions but characteristic of only few living researchers, is constructed. Many evaluation researchers characterise this debate as 'unfortunate' (Worthen, Sanders, & Fitzpatrick, 1997). Kuhn's paradigms, as portrait by e.g., Guba & Lincoln (1990;

---

[4] By the term 'ideology' we refer to the pre-Marx concept of ideology that was first coined by Destutt de Tracy in 1796 to refer to the 'science of ideas'.

1989) are rigid in their hard-core structure and contribute little to justify a determined devotion to either quantitative or qualitative inquiry.

Qualitative Methods and the Researcher – Perils and Potentials

Creswell's five research traditions (Creswell, 1998) (table 3) include *biography*, *phenomenology, grounded theory, ethnography* and *case studies*.

Biography is hardly ever used in evaluation, although it may be relevant for 'historical' cases. However, as a method, biography is usually associated with subjectivity and there are countless examples of storytelling blamed for lacking 'authenticity' or being otherwise biased from one or other point of view.

Phenomenology seems to have lost the meaning as given by Husserl (1965). In psychology '*phenomenological*' is typically used as an interchangeable word for '*subjective*' (Jennings, 1986) and most psychologists conceive of phenomenology as the study of private responses to a given situation, an opportunity for studying the individual's 'unique point of view'. When the modern-day phenomenology researcher talks about 'essence' he/she usually have this 'unique point of view' in mind. Husserl, in his mission to restore philosophy's original, more prominent position among the sciences against the increasing dominance of the natural sciences, used the term 'essence' for a fact or entity that is universal, eternally unchanging over time, and absolute (Jennings, 1986). Hence, when 'essence' to Husserl was things like 2+2=4 that has a timeless definite reality, something *not* unique to the single individual is the focus of phenomenology according to Creswell (1998). For Creswell phenomenology is 'understanding the essence of experiences about a phenomenon' (table 3). According to Creswell 'long interviews with up to 10 people' is what it takes to achieve such understanding (table 3). *Ethnomethodology*, (Garfinkel, 1967) an outgrowth of the phenomenological movement seems to be a more refined development of phenomenology that steers clear of much of Jennings' criticism. The *common* knowledge is the focus for ethnomethodology. "Ethnomethodological studies analyse everyday activities as member's methods for making those activities visibly, rational and reportable for all practical purposes, i.e., 'accountable' as organisation of commonplace everyday activities" (Garfinkel, 1967:vii).

Grounded theory (Glaser & Strauss, 1967) implies that the researcher approaches phenomena with as little *a priori* bias or prejudice as possible, thereby allowing core theoretical concepts to emerge in the process of undertaking fieldwork. Hence, contrary to the traditional Anglo-American approach where theory is given the priority and hypotheses derived from theory are tested against data, theory is not only informed, but also constructed from existing

data. This approach is unfamiliar to the Anglo-American research community, but common in French sociology, also within quantitative methodology, in particular in *correspondence analysis* (Greenacre, 1993) where they let 'data speak' to the researcher. Grounded theory emphasises procedures for formulating new knowledge claims that are founded upon under-theorised empirical situations. The relation between its procedures and established theory is, however, unclear. Procedures are strongly inductive, and as theory become developed, increasingly less suited as means for confirming that data fits theory (or vice versa?). Grounded theory researchers are usually not devotees of the qualitative methodology. In fact grounded theory suggests how to combine qualitative and quantitative inquiry in *order to produce theory*.

Ethnography (Manilowski, 1922) the core method of anthropologists, and the preferred methodology for a growing number of evaluation scholars, seems to be in a confused state after a series of intense debates over belief ascription (Jones, 2000; Weeks, 2000).

Belief ascription, a common practice both in social science and everyday life and a central *methodological* feature of ethnography is problematic in the sense that is easy to do, but it is hard to provide *evidence* for the manners of thinking and feelings we ascribe to others. It is probably impossible, or at least very hard to relate to other people without making belief ascription. It is also layman knowledge that is easy to reach a wrong conclusion about other's thoughts and feelings. The ambitions of ethnographers is, however, to gather evidence of *unconscious* beliefs, a task that is undoubtedly more demanding. The two general ways of gathering evidence can be labelled 'the behavioural strategy' and the 'environmental strategy' (Jones, 2000). The 'behavioural strategy' typical assumes that certain sorts of things can *cause* the observed behaviour, and hence, what is observed is evidence of the assumed or ascribed belief that is the cause. The 'environmental strategy' relies on observing external conditions that can be thought to produce certain regular, observable patterns. In most cases, the two strategies are used in combination to improve the reliability of conclusions. Ethnography is clearly in line with much 'folk psychology' but in contradiction with most logical reasoning, first and foremost because causes have to be assigned to effects beforehand and the observed is then granted the status as evidence of the presumed cause. The list of potential alternative explanations may in many cases, indeed, be long.

Case studies, (Yin, 1994) the well-known in-dept analysis of a single case, certainly has its followers in evaluation research. Where it is in fact, only one case to study, this may also be the obvious choice. However, case studies do not necessarily follow the quantitative- qualitative divide, both types of data may be gathered as evidence (Jick, 1979). Excellent studies like

Allison's study of the Cuban missile crisis (Allison, 1971) and Pressman and Wildavsky's study of implementation processes (Pressman & Wildavsky, 1973) have been carried using the case study methodology. – It is common to distinguish between various types of case studies, such as *descriptive*, *exploratory* and *explanatory* (Yin, 1994). Of these types, the latter is clearly the most ambitious one, as it aspires at causal explanations. Exploratory case studies are sometimes considered as a prelude to other research, often of quantitative nature. Stake (1995) further makes a distinction between, *intrinsic*, *instrumental* and *collective* case studies. Intrinsic refers to studies where the researcher has an interest in the case, instrumental refers to the situation where the case is used to understand more than what is directly open to the researcher, and collective refers to a group of related case studies.

I do not intend to give anything near a comprehensive account for the many varieties of qualitative inquiries, and I believe I can be rightfully accused for overlooking important nuances. I do, however, contend that they have something in common that distinguish them from quantitative methods. *First* and foremost, the dependency upon the researcher's personal judgement is strong, as there are few options for re-evaluating a personal interpretation. *Second*, the rules and recommendations for carrying out research are indistinct and vary substantially across textbooks. *Third*, the emphasis on in-dept understanding and discoveries of new aspects and dimensions is very different from the emphasis on *verification* that is at the core of impact or effect oriented evaluations.

The *first aspect*, the dependence on the researcher's personal judgement, implies that the search for in-dept understanding is risky in the sense that interpretations of discoveries that are the outcomes of interpersonal processes are not error free. In particular, beforehand expectancies have a tendency to turn out to be confirmed. Also, this tendency may interact with the reactivity of measures, in particular when respondents know in advance how to provide the right answer, e.g., how to answer in their own best interest.

The *second aspect*, indistinct rules and recommendations make it hard for other researchers to replicate studies, and hence may reduce the assumed reliability of studies. It also tends to make standard features of research, such as the distinction between type I error and type II error, to disappear. I find it hard to believe that qualitative researchers can entirely avoid all situations where the consequences of an erroneous conclusion are so severe that there is a need for more conservative or careful judgement than in other situations. When a doctor dis-

covers symptoms of a severe disease, he will most likely order further diagnostics even though he knows that the probability that the patient has this particular disease is very low. In the case of evaluations, we can make the parallel that the failure of a programme/project may be conceived as a more severe condition than a success. Thus, but the qualitative researcher has no means for adjusting his chances of declaring success when this is not the case, or, vice versa, of proclaiming a failure when that is not the case.

The *third aspect*, the problem that *verification* is mainly overlooked, is in conflict with the popular notion of evaluation, as perceived by most policymakers. This is not necessarily a consequence of the choice of methods in itself. As demonstrated by Mohr (1999) causal claims can be established within a qualitative inquiry framework. It is more the deliberate choice of focus that makes this problematic. The public discourse surrounding an evaluation project is inevitably an evidence seeking process. Thus, any written conclusion whether being it a new insight or discovery or any other description of the project/programme under investigation can be interpreted as a finding, a result or an outcome of the project. In general, the discourse following an evaluation report may have consequences that by far exceeds the intentions of the researcher and misguided interpretations can be quite embarrassing, in particular when there is no evidence for conclusions about outcomes that is promoted in the debate. Also, the quality of the public discourse may severely damaged when statements of the following kind are put forward:

"Reviews of past projects five years after programme completion report that

- *43% of participating SME's had increased their turnover,*
- *53% had accessed new markets and*
- *42% had created new jobs"* [5]

Statements like these are as commonplace as they are meaningless. In the case of the first statement nobody seems to question whether the remaining 57% in the study had *decreased* their turnover or whether other, comparable firms, not included in the study, had *increased* their turnover by *more* than 43%. By the same token, whether more than 53% of other, comparable SME's, not included in the study, had accessed new markets or whether the 42% of firms that had created new jobs had created many jobs or, say, only hired one new employee each are seldom questioned.

In the example above, it is easy to see how quantitative information is construed and transformed in a favourable direction, even when it is void of any meaning. Verbal statements are just as vulnerable to distortions and ambiguous statements more than unambiguous ones.

Ontological Commitment and the Luhmannian Trap

It is generally accepted that the term social sciences cover several basic disciplines, like economics, sociology, anthropology, organisation studies and so forth. Moreover, within or between these basic disciplines a variety of sub-fields or sub-disciplines have emerged over the years. Each of these up-and-coming sub-disciplines is viewed as having an inner drive towards maturity, although some of them may turn out to be just fads. The development towards maturity implies a number of institutional processes, including more formalised arrangements as conferences and the establishing of new journals for the field. Clearly, the ambitions of the sub-fields vary, from the more modest to the full-blown ambitions of professionalism with formal educational requirements and defined boundaries for membership.
As evaluations research becomes divided into increasingly distinct sub-fields, the debate over methods in evaluation research deserves new attention. Studies based on Transaction Cost Economics, Social Action Theory or Fourth Generation Evaluation can all be labelled evaluations. The world-views inherent in the sub-disciplines usually reflect those of a parent discipline. The distinctive characteristic of methodological peculiarities of the parents disciplines in usual are carried over to their respective sub-disciplines and in most cases worsened by the sub-disciplines status as an ongoing theoretical project. As theoretical project sub-disciplines tend to have a defensive drive towards maturity, e.g. the development of core concepts and research methodologies

Membership in a research community that constitutes a discipline or sub-discipline entails *commitment*. By ontological commitment we mean that this commitment is different from solidarity or other feelings of belonging to the research community. Ontological commitment implies a shared body of concepts and other entities that are assumed to exist in the discipline or sub-discipline. That is, ontological commitment implies that the acceptance of what is assumed to exist as important, is the sign of membership.

---

5 European Commission, 2000. SME: Taking the opportunity. In Community Research (Ed.) *Innovation & Research – European support for small companies* European Commission SME Helpdesk: Office for Official

Thus, academic disciplines and sub-disciplines can be regarded as specialised communication circuits that are thematically and conceptually differentiated from the general communicative circuit of society. Some of them have become so independent that they can be regarded as an 'autopoeitic' social systems (Luhmann, 1995), i.e. a self-referential social system that constitutes itself by blocking external communication. The legal system is the standard example of such self-referential, self-producing system. Applied to disciplines and sub-disciplines this means that your mastering of the 'internal language' and the ability to demonstrate familiarity with this 'hidden jargon' of your research community is the key to success. These aspects of a scholarly disciplines and sub-disciplines are of importance for understanding how rules for acceptance are established.

The basic question in any evaluation is to provide trustworthy evidence about the merit or worth of the programme or project in question. The reasons for using phrases like *evaluation research* the reasons for engaging *research institution* to do the job, and the reasons for discussing *research methods* whenever *evaluation methods* are discussed, have to do with the reliability and validity of conclusions. We invoke our right to appeal to the legitimacy of science by demonstrating that we apply re-examinable *scientific methods*. That is, we use what is commonly known as the most reliable methods within the fields we are working. Clearly, this line of reasoning carries obligations that restrict our range of approaches to the way we answer questions about the outcome, worth or merit of the efforts to improvement we are asked to evaluate.

A question is an expression of intellectual anxiety and an answer an attempt at resolution of that anxiety (Myhill, 1951). From a logical point of view, we can distinguish between two kinds of questions, *formal* and *inform* questions. A formal question carries with it the form of its answer, i.e., the social context is such that the criterion of the acceptability of the answer is known and agreed upon by both questioner and answerer in abstraction from the answer itself (Myhill, 1951). The purest kind of formal question is the question of the truth or falsity of a mathematical theorem within a known system. The criteria for being a proof within the system are exactly specified and agreed upon by both questioner and answerer. In empirical research formal questions are a less pure kind of formal questions. Criteria of confirmation are

less specifiable than for the mathematical proof and hence, not so easily agreed upon by the research community.

An informal question is one where the form the answer is not known either by the answerer or the questioner in abstraction from the answer itself. The dictum that the meaning of a proposition is the method of its verification, does not apply to propositions which answer informal questions, since part of the meaning of such questions is to question what the form of its answer should be (Myhill, 1951:58). Hence, part of the meaning of the question "What are the merits of this particular project" is "What form of answer is best suited to resolve the anxiety expressed by this question"? That is, to the same extent a question is formal, to the same extent the questioner will be prepared to state precisely the kind of evidence it would take to convince him/her of the of the truth of any proposed answer. Hence, a formal question asks for the matter of its answer but provides the form while an informal answer asks for both (Myhill, 1951). The problem is that is mainly the informal questions that provide new insights, and the form of the answer has to be agreed upon by the research community. With a considerable number of 'schools' or sub-disciplines that establish their own, internal standards for conduct, what is considered the truth of a proposed answer in one camp, may be very different from what is considered acceptable in another camp. In my view, evaluation research is heading in the direction where the boundaries between disciplines and sub-disciplines make it difficult to establish how evidence for outcomes should be verified. This development does not necessarily enrich the field of evaluation research.

The Mandatory Assessment of Costs and Benefits

Under the realm of the New Public Management, *cost-benefit analysis* fits inside the established framework of understanding of society as an economy. The framing (and priming) effect of the New Public Management assumes cost-benefit analysis as a natural extension of any evaluation. Also, cost-benefit analysis is the preferred *method* of evaluation for many governmental agencies. Due to traditions, general trust in the discipline of economics, and the fact that the educational background of many bureaucrats is economics, many tender documents specifically require cost-benefit analysis to be integral part of the evaluation. That is, whatever kind of evaluation the researcher wants to carry out, whatever 'school' the evaluator subscribe to and whatever methodology is qualitative or quantitative, the final report *have to* include an assessment of benefits and costs in the form of a cost-benefit analysis. This re-

quirement should be expected to have some impact upon the choice of evaluation methods. It is, however, surprisingly many evaluation reports that manage to carry out a cost-benefit even though they do not include convincing attempts at assessing the impacts of the proj??ect/programme in question. The fact that this can be done so smoothly, cast serious doubt about such assessments of the cost and benefits of a given project/programme, in particular there are reasons to be sceptic about the calculations of the benefit side.

In its general form, cost-benefit analysis has an appealing logical structure that resembles sound accounting principles. The benefits refer to the changes in the allocation of resources brought about by a project or programme by comparing the situation before and after the installation of the project/programme. Given some norm of calculating social welfare, the two situations can be compared. For policy maker it is of importance to know not only to which extent a given project or programme is successful, but also if there is a reasonable proportionality between costs and the results achieved. Usually, this comes down to the simple question whether the benefit cost ratio is greater than one or not.

More recent research has revealed that cost-benefit analysis is in such a troubled state that its usefulness in evaluations should be questioned. Clearly, the reason why so many public agencies demand cost-benefit analysis to be an integral part of *any* evaluation report is, at the surface level, a need for clarification of the costs associated with demonstrated outcomes. At a more subtle level, it may be questioned if it is the accountancy-like logic that constitutes a rhetorical beauty well suited for policy discourses that is the reason. It is disturbing that a technique, well proven to be seriously flawed and void of any scientific merit, still has its camp of devotees and is even regarded as a requirement for a complete evaluation. As stated by Lewis A. Kornhauser, "*Cost-benefit analysis influences, if not controls, many public decisions of great importance,*" but "*its justifactory foundations remain as best suspect and at worse in ruins*" (Kornhauser, 2000):1037).

The Institutionalisation of Evaluations

There are many instruction manuals and textbooks on the market, whose sole mission seems to be to encourage standardisation of the routines of evaluation. Institutional theory implies that there is an inherent danger the procedures of evaluation could turn into what James March coined 'procedural rationality' (March, 1988). That is, the sequence of actions that constitutes an evaluation process seems to emulate a rational process so that the process itself

became the fulfilment of the evaluation task, the outcome of the evaluation a by-product of less importance. I believe this transition has already taken place in many governmental agencies. Clearly, when the focus of attention is shifted from outcome to simply having carried out the routine, issues about methodology will become less prominent. It is not the results that matter any longer, but that the evaluation is, in fact, carried out. Routines carry habits and standard operating procedures and cognitive processes of 'taken for grantedness' that leaves evaluation managers surprisingly uncritical. The worst case scenario is a situation where well paid experts function as storytellers based on exclusive and well reputed insight and most project/programmes accomplish a positive cost-benefit ratio based on good guesstimates. The warnings with respect to methodology, that it is the impression that 'anything goes' (Adelman, 1996) is a hint that we may not pay sufficient attention to institutionalisation processes that develop a wrong track.

The Quest for Effects

I have questioned the worth of qualitative methods when it is known that most of the real attention anyway will be directed toward evidence for effects or impacts of the programme/project under scrutiny. I have also argued that many self-reported measures are of limited value for impact analysis due to their reactivity to the context they are obtained within. That is, respondents do in many cases know how to react as participants in a study. Hence, it is clear that unobtrusive or non-reactive measures are needed for assessments of impact. Such measures can be established from archival records or other measurements that can be obtained independent of the ongoing project/programme under investigation. With such measures at hand, there is reliable ways of answering the inevitable question about effects and providing trustworthy evidence for the extent to which a project/programme did have or did not have the intended impact upon those problems initially targeted.

The *causal-counterfactual approach* (Pearl, 2000) to evaluations is a rather recent invention. In essence, the inspiration for this approach can be found in several shortcomings of the established quasi-experimental approach (Cook & Campbell, 1979). First and foremost, there has been a longstanding agreement among most scholars that the only appropriate way to talk about the effect of a public intervention is to speculate about the hypothetical situation that would have prevailed in the absence on the intervention. This line of reasoning introduces the term *counterfactual* and the interpretation and quantification of impact or effect of the intervention as the difference between the *factual* and the *counterfactual*. The development in

possible world semantics (Lewis, 1973) provided the logic for counterfactual reasoning, and work in statistics along the same line of reasoning (Rubin, 1974; Rubin, 1978; Rubin, 1990) made the technical solutions available. The technical/statistical solutions also facilitated a way out of those problems that had laid earlier analysis of the quasi-experiments on shaky grounds, in experimental language, the non-random assignment of treatment, and the effects of cases (individuals, firms) being singled out for treatment for special reasons, the so-called selection effect.

Lewis' original (Lewis, 1973) formulation of the counterfactual theory of causation was spelled out under the assumption of determinism, and must be modified to allow for chancy causation. This work led to the more general notion of the causal counterfactual. *"Where c and e are distinct events, e causally depends on c if and only if, if c had not occurred, the chance of e's occurring would have been much less than it actually was (given that c occurred)"* (Lewis, 1986). This more common sense notion of causality as related to events in the past, laid the ground for joining the logical structure of causation with the probabilistic notion of causation. In a number of works, mostly in the field of medical statistics Paul R. Rosenbaum and Donald B. Rubin (Rosenbaum & Rubin, 1985; Rosenbaum, 1995; Rosenbaum & Rubin, 1983) developed their framework for the so-called *observational studies* where the properties of random assignments is emulated by means of matching procedures or other techniques. Throughout the 1990's this line of reasoning was followed up, criticised and refined by a number of statisticians (Robins, 1989; Robins, 1997) a number of econometricians (Heckman, Ichimura, & Todd, 1998; Heckman, 1995; Heckman, Ichimura, & Todd, 1997; Heckman & Smith, 1995; Heckman & Smith, 1997) and sociologists (Winship & Morgan, 1999). The causal-counterfactual approach follows the language of experimentation, similar to the literature on quasi-experimentation. Contrary to the quasi-experimental tradition, the causal-counterfactual approach is capable of dealing with the two dominating sources of possible bias in quasi-experiments. Outcomes for the treatment and control group may differ even in the absence of treatment, and the effect of treatment may differ for the treatment and control group (Winship & Morgan, 1999). The causal-counterfactual approach provides a general framework that facilitates statistical analysis in practical settings, not merely discussions of potential and pitfalls.

Concluding Remarks

I believe that Reisenbach's (1938) distinction between context of discovery and context of justification, and that the different contexts call for different methodologies, has considerable virtue for evaluation research. When the research task is explicitly stated as a case for justification, the relevant methodology should be applied. Hence, I do not maintain that qualitative inquiry is of less worth than quantitative methodology, I simply make the claim that there is a risk that qualitative methodology in the evaluation context may tend to provide good answer to questions that the evaluation task did not ask for. By doing so, the likelihood that the evaluation management or other stakeholders may want to reinterpret answers to fit the question they ordered and paid for, increases.

I also believe that qualitative inquiry implies psychological challenges that are not sufficiently incorporated in the methodology literature, and that qualitative methodology tends to reduce the distance between research and researcher to an extent where criticism of results are hard to separate from disapproval of the researcher. Thus, in heated discussions the researcher leads the debate towards the fallacy of *argumentum ad hominem*, in soccer jargon, the fallacy of taking the man instead of the ball. Hence, bias may be the outcome of a process where the researcher is not defending his results, but himself.

The prevalent devotion to various more or less distinct 'schools' of evaluation may actually reduce the risk of evaluation myopia but may also invoke the *relativist fallacy*, i.e. the rejection of a claim by asserting that the claim might be true for others but is not for him/her. Strong ontological commitment might, under given circumstances, mean that results that is considered acceptable within one discipline or tradition, is not acceptable, and hence, not true within another discipline or 'school'. This is in particular true for findings that threaten the basic conceptual or logical assumptions of a tightly connected 'school' of research.

As expressed by Wolfson (Wolfson, 2001:95) "… cost-benefit analysis is inseperable from the free market principle, and further, it attempts to apply these principles to the conduct of public affairs. Like the marketplace, cost-benefit analysis take individual preferences as its guide, and it assumes that government should serve these preferences, not direct or educate them". Used in connection with qualitative methods or other methods that do not provide any basis for assessing benefits, cost benefit may still be a sound way of calling the responsibility for costs and benefits to attention. Such use will be fully in accordance with "The Statement

of Principles on cost-benefit analysis" from the American Enterprise Institute (AEI), that states that "Benefit-cost analysis should be required for all major regulatory decisions, but agency heads should not be bound by a strict benefit-cost test. Instead, they should be required to consider available benefit-cost analysis and to justify the reasons for their decisions in the event that the expected cost of a regulation far exceed the expected benefits"[6].

Under the realm of the New Public Management, the need for an analysis that do not attempt to assess the benefits of a project/programme can also be seen as an example of the *petitio principii*, or begging the question. Since the principles of welfare economics prevail, we perform a cost-benefit analysis that confirms that these principles are valid, as assumed by the foundation of the New Public Management.

---

[6] The statement is signed by Kenneth J. Arrow, Robert W. Hahn and Robert N. Stavins

References

Adelman, C. (1996). Anything Goes. *Evaluation, 2*(3), 291-305.

Allison, G. T. (1971). *The Essence of Decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown.

Atwell, J. (1982). Human rights in human subjects research. In A. Kimmel (Ed.), *Ethics of human subject research* (pp. 81-91). San Francisco: Jossey-Bass.

Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology, 52*, 1-26.

Bourdieu, P., & Coleman, J. S. (Eds.). (1991). *Social theory for a changing society*. Boulder, CO, US: Westview Press.

Carrol, J. D. (1998). Book Review. *American Review of Public Administration, 28*(4), 402-407.

Cassirer, E. (1910). *Substanzbegriff und Funtionsbegriff*. Berlin: B. Cassirer.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation Design & analysis issues for field settings*. Boston: Houghton Mifflin Co.

Creswell, J. W. (1998). *Qualitative Inquiry and Research Design: Choosing Among Five Traditions*. London: Sage.

Dearborn D., & A., S. H. (1958). Selective perceptions: A note on the departmental identification of executives. *Sociometry, 21*, 140-144.

Donaldson, S. I. (2001). Overcoming our Negative Reputation: Evaluation Becomes Known as a Helping Profession. *American Journal of Evaluation, 22*(3).

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Research, 14*, 532-550.

Flick, U. (2002). *An Introduction to Qualitative Research*. (2 ed.). London: Sage.

Garfinkel, H. (1967). *Studies in Ethnometodology*. Englewood Cliffs NJ: Prentice-Hall.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine.

Gore, A. (1993). *Creating a Government that Works Better and Costs Less: Report of the National Performance Review*. . Washington D.C.: U.S. Government Printing Office.

Greenacre, M. (1993). *Correspondence Analysis in Practice*. London: Academic Press.

Guba, E. G. (1990). *The Paradigm Dialog*. Newbury Park: Sage.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, Calif.: Sage Publications.

Haveman, R. H. (1987). *Poverty Policy and Poverty Research.* Madison: University of Wisconsin Press.

Heckman, J., J. , Ichimura, H., & Todd, P. (1998). Matching As An Econometric Evaluation Estimator. *Review of Economic Studies*(65), 261-294.

Heckman, J. J. (1995). *Randomization as an instrumental variable*. Cambridge, MA: National Bureau of Economic Research.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies, 64*, 605-654.

Heckman, J. J., & Smith, J. A. (1995). *Ashenfelter's dip and the determinants of participation in a social program : implications for simple program evaluation strategies*. London, Canada: Dept. of Economics University of Western Ontario.

Heckman, J. J., & Smith, J. A. (1997). *The sensitivity of experimental impact estimates : evidence from the national JTPA study*. Cambridge, MA: National Bureau of Economic Research.

Husserl, E. (1965). *Phenomenology and the crisis of philosophy : Philosophy as rigorous science ; and Philosophy and the crisis of European man*: Harper & Row.

Jennings, J. L. (1986). Husserl Revisited. The Forgotten Distinction Between Psychology and Phenomenology. *American Psychologist, 41*(11), 1231-1240.

Jensen, M., & Meckling, W. (1976). Theory of the firm: Managerial behavior, agency costs, and ownership structure. *Journal of Financial Economics, 3*, 305-360.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*, 602-611.

Jones, T. E. (2000). Ethnography, belief ascription. and epistemological barriers. *Human Relations, 53*(1), 117-152.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: Ana analysis of decision under risk. *Econometrica, 47*, 263-291.

Kelley, H. H. (1992). Common-sense psychology and scientific psychology. *Annual Review of Psychology, 43*, 1-23.

Klamer, A. (2001). Making sense of economice: from falsification to rhetoric and beyond. *Journal of Economics Methodology, 8*(1), 69-75.

Kornhauser, L. A. (2000). On Justifying Cost-Benefit Analysis. *The Journal of Legal Studies, xxix*(2), 971-1004.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lewis, D. (1973). *Counterfactuals*. Cambridge, MA: Harvard University Press.

Lewis, D. (1986). Postscript to 'Causation'. In D. Lewis (Ed.), *Philosophical Papers: Volume II* . Oxford: Oxford University Press.

Luhmann, N. (1995). *Social systems*. Stanford, Calif.: Stanford University Press.

Lægreid, P. (2000). To Civil Servants Under Contract. *Public Administration, 78*(4), 879-896.

Manilowski, B. (1922). *Argonauts of the Western Pacific*. New York: E.P. Dutton.

March, J. G. (1988). Bounded rationality, ambiguity, and the engineering of choice. In J. G. March (Ed.), *Decisions and Organizations* . Oxford: Basil Blackwell.

Merton, R. K. (1948). The self-fulfilling prophecy (1982). In A. Rosenblatt & T. F. Fieryn (Eds.), *Social Research and the Practicing Professions* . New York: Abt Books.

Mohr, L. (1999). The Qualitative Method of Impact Analysis. *American Journal of Evaluation, 20*(1), 69-84.

Morgan, G., & Smircich, L. (1980). The Case for QUalitative Research. *Academy of Management Review, 5*(4), 491-500.

Morse, J. M., Hupcey, J. E., Penrod, J., Spiers, J. A., Pooler, C., & Mitcham, C. (2002). Issues of Validity: Behavioral Concepts, Their Derivation and Interpretation. *International Journal of Qualitative Methods, 1*(4).

Myhill, J. R. (1951). On the Ontological Significance of the Löwenheim-Skolem Theorem. In M. White (Ed.), *Academic Freedom, Logic and Religion* (pp. 57-70). Pittsburg: The University of Pennsylvania Press.

Nathan, R. P. (1988). *Social Science in Government: Uses and Misuses*. New York: Basic Books.

Olson, J. M., Roese, N. J., & Zanna, M. P. (1996). Expectancies.. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles* . New York: Guilford.

Pearl, J. (2000). *Causality : models, reasoning, and inference*. Cambridge: Cambridge University Press.

Plowden, W. (1994). *Ministers and Mandarins*. London: Institute for Public Policy Research.

Poslby, N. W. (1998). Social Science and Scientific Change: A Note on Thomas S. Kuhn's Contribution. *Annual Review of Political Science*(1), 199-210.

Powell, W. W., & DiMaggio, P. J. (Eds.). (1991). *The New Institutionalism in Organizational Analysis*. Chicago: University of Chicago Press.

Pressman, J. L., & Wildavsky, A. (1973). *Implementation*. Berkeley: University of California Press.

Reichenbach, H. (1938). *Experience and Prediction: An Analysis of the Foundatins and Structure of Knowledge*. Chicago, Illinois: University of Chicago Press.

Ricoeur, P. (1970). *Freud and Philosophy: An Essay on Interpretation*. New Haven: Yale University Press.

Robins, J. M. (1989). The analysis of randomized and nonrandomized AIDS treatment trials using a new approach to causal inference in longitutional studies. In L. Sechrest, H. Freeman, & A. Mulley (Eds.), *Health Service Research Methodology: A Focus on AIDS* (pp. 113-159). Washington DC: US Public Health Service.

Robins, J. M. (1997). Causal inference from complex longitudinal data. In M. Berkane (Ed.), *Latent Variable Modeling and Applications to Causality: Lecture Notes in Statistics* . New York: Springer-Verlag.

Rosenbaum, P., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods the incorporate the propensity score. *American Statitician, 39*(33-38).

Rosenbaum, P. R. (1995). *Observational studies*. New York: Springer-Verlag.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Rosenthal, R. (1976). *Experimenter effects in behavioral research*. New York: Irvington.

Rosenthal, R., & JAcobson, L. (1968). *Pygmalion in the Classroom*. New York: Holt, Rinehart & Winston.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology, 66*, 688-701.

Rubin, D. B. (1978). Bayesian inference for causal effects. *Annals of Statistics, 6*, 34-58.

Rubin, D. B. (1990). Formal models of statistical inference for causal effects. *Journal of Statistical Planning and Inference, 25*, 279-292.

Scott, W. R. (1995). *Institutions and organizations*. Thousand Oaks: Sage Publications.

Scriven, M. (1993). *Hard-won lessons in program evaluation*. San Francisco: Jossey-Bass.

Simon, H. A. (1982). *Models of bounded rationality*. Cambridge, Mass.: MIT Press.

Skinner, B. F. (1953). *Science and human behavior*. New York: Macmillan.

Snyder, M., & Stukas, A. A. J. (1999). Interpersonal Processes: The Interplay of Cognitive, Motivational and Behavioral Activities in Social Interaction. *Annual Review of Psychology, 50*, 273-303.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks: Sage Publications.

Tversky, A., & Kahneman, D. (1986). Rational Choice and the Framing of Decisions. *Journal of Business, 59*(4), 251-278.

Van Maanen, J. (1979). Reclaiming Qualitative Methods for Organizational Research: A preface. *Administrative Science Quarterly, 24*(4), 520-526.

Vedung, E. (2000). *Public Policy and Program Evaluation*. New Brunswick: Transaction Publishers.

Webb, E., Campbell, D., Schwartz, R., & Sechest, L. (1966). *Unobtrusive measures*. New York: Rand-McNally.

Weeks, J. R. (2000). What do ethnographers believe? *Human Relations, 53*(1), 153-171.

Winship, C., & Morgan, S. L. (1999). The Estimation of Causal Effects from Observational Data. *Annual Review of Sociology, 25*, 659-706.

Wolfson, A. (2001). The costs and benefits of cost-benefit analysis. *The Public Interest, Fall 2001*.

Worthen, B. R., Sanders, J. R., & Fitzpatrick, J. L. (1997). *Program Evaluation Alternative: Approaches and Practical Guidelines*. (Second Edition ed.). New York: Longman.

Yin, R. K. (1994). *Case Study Research, Design and Methods*. (2nd ed.). Newbury Park: Sage Publications.

Aaron, H. J., Gramlich, E. M., Hanushek, E. A., Heckman, J. J., & Wildawsky, A. (1990). Social Science Research and Policy. *The Journal of Human Resource, 25*(2), 297-304.