VALIDITY OF THE AHP/ANP: COMPARING APPLES AND ORANGES

Sibs von Solms
Mondi South Africa Division
Richards Bay Mill
sibs.vonsolms@mondigroup.co.za

ABSTRACT

Determining the validity of the AHP/ANP is an important issue. This paper discusses the complex and often controversial field of validation. The debate regarding the validity of the AHP/ANP is evaluated, from which the need for a different approach is identified. The AHP/ANP is a decision-making methodology that should be located within a qualitative, subjectivist or constructivist paradigm where different measures of validating research become relevant. The goal of this paper is three-fold: First, to locate the AHP/ANP within the field of Interpretivist qualitative research; Second, to argue for the importance of validating the AHP/ANP in terms of criteria important to the decision makers rather than some *objectively* given or *normatively* defined criteria; Third, to stimulate research specifically aimed at testing the validity of the AHP/ANP in terms of decision makers' criteria.

Keywords: Consensus, Constructivism, Criticisms of the AHP/ANP, Group Decision-Making, Multi-Stakeholder Groups, Qualitative Research, Quantitative Research, Sociological Paradigms, Subjectivity, Validity.

http://dx.doi.org/10.13033/ijahp.v3i1.60

1. Introduction

At the recent ISAHP2009, the question was frequently raised as to how presenters were going to prove that their findings or conclusions were *valid*. The issue of research validity is a complex and often controversial subject [Sechrest, 2005] involving both philosophical as well as practical aspects. The paper will take as its starting point the debate regarding the validity of several aspects of the AHP/ANP, briefly highlighting both the arguments aimed at showing the AHP/ANP to be an invalid method as well as those offered in defense of the AHP/ANP including the validation examples discussed by Whitaker [2004]. This debate is, however, not the focus of the current paper, but its evaluation leads to the identification of an aspect neglected by the extant literature, i.e. that the AHP/ANP is an approach with specific strengths to enhance multi-stakeholder group decision-making. The traditional forms of research validity are tied closely to a quantitative/objectivist research paradigm and it is argued that different and more appropriate criteria are required for research and application within a qualitative/subjectivist paradigm [Guba & Lincoln, 1994]. Applying the AHP/ANP in the multi-stakeholder group decision-making field requires the formulation of a new set of validity criteria that are relevant to the problematic of such multi-stakeholder groups and the stimulation of research to test the validity of the AHP/ANP as a multi-stakeholder group decision-making (GDM) methodology. Traditional approaches emphasize the *prediction* of outcomes, but what we need are approaches that can assist decision makers to *actively make things happen*. Human development requires *action* and action requires *choice* based on

human interest. The validity of the AHP/ANP hinges on its ability to support such choice.

2. The AHP Validity Problem

2.1 Proving the Invalidity of the AHP/ANP

A steady stream of papers have been written claiming the invalidity of some technical and mathematical aspects of the AHP/ANP methodology. Saaty [2008] says that essentially, there have been five main types of criticisms of the AHP all of which have been addressed in the literature (e.g. [Saaty et al, 2009]), namely:

(1) Rank Reversal: This is the concern with illegitimate changes in the ranks of the alternatives upon changing the structure of the decision. Rank reversals are shown by critics to occur when using comparisons and relative measurement in two ways: First, when new alternatives are added or old ones deleted; and second, when new criteria are added or old ones deleted with the caveat that the priorities of the alternatives would be tied under these criteria and hence argued that the criteria should be irrelevant when ranking the alternatives. Rank reversals that followed such structural changes were attributed to the use of relative measurement and normalization. Papers relevant to this critique include [Watson and Freeling, 1982; Belton and Gear, 1983; Dyer, 1990; Barzilai and Golany, 1994; Pérez, 1995; Finan and Hurley, 2002; Pérez et al, 2006; Ishizaka and Labib, 2009].

(2) Consistency & Aggregation: The concern is about inconsistent judgments and their effect on aggregating such judgments or on deriving priorities from them. A condition that may not hold with inconsistent judgments is Pareto optimality. Another condition also inherited from expected utility theory has to do with a relation called Condition of Order Preservation (COP). A paper relevant to this critique is [Bana e Costa and Vansnick, 2008].

(3) Geometric Mean & Multiplicative Weighing: The third criticism has to do with attempts to preserve rank from irrelevant alternatives by combining the comparison judgments of a single individual using the geometric mean (logarithmic least squares) to derive priorities and also combining the derived priorities on different criteria by using multiplicative weighting synthesis. Papers relevant to this critique include [Holder, 1990; van den Honert and Lootsma, 1996; Barzilai, 1999, Forthcoming].

(4) Fundamental Scale: The fourth criticism has to do with people trying to change the fundamental scale. Papers relevant to this critique include [Ma and Zheng, 1991; Salo and Hämäläinen, 1997; Leskinen, 2000].

(5) Behavioral Aspects of Pairwise Comparisons: The fifth and final criticism has to do with whether or not the pairwise comparisons axioms are behavioral and spontaneous in nature to provide judgments.

2.2 Proving the Validity of the AHP/ANP

Proponents demonstrate the validity of the AHP/ANP in mainly three ways:

(1) Counter Arguments: In reply to the criticisms listed above, several papers were written to answer and counter the comments made. References relevant to this point include [Saaty, 1990, 1999, 2001, 2008, 2010; Harker and Vargas, 1990; Vargas,

1997; Forman and Gass, 2001; Wijnmalen, 2001; Saaty and Vargas, 2006; Whitaker, 2007; Saaty et al, 2009].

(2) Applications: The fact that the AHP/ANP is one of the most widely used MCDM methodologies is argued as proof of its validity. The point is made that if the AHP/ANP was so flawed and invalid, applications would not have proliferated as they have because the results would not have been useful. One specific reference in this regard is [Forman and Gass, 2001].

(3) Validation Examples: Whitaker [2004] presents thirteen research cases in which the validity of the AHP/ANP was demonstrated. This specific approach to testing the validity of the AHP/ANP requires a more detailed discussion.

2.3 Validity Examples

Whitaker [2004] points out that there are two ways in which the AHP/ANP can be validated, i.e. as a *predictive* tool or as a *decision-making* tool. The former lends itself to fairly easy validation due to the availability of both a known of real world data and an AHP/ANP predicted outcome for comparison. In the latter case, the success of a decision may not be known for some time and is, further, always a matter of interpretation, sometimes influenced by events happening that were – or could – not be known at the time of making the decision [Whitaker, 2004]. An additional problem is that in the absence of a known correct outcome, even a decision that can be shown to be a success, may well be sub-optimal because it could be argued that another decision-making tool could have led to an outcome that may have outperformed the AHP/ANP decision. Whitaker [2004] presents 13 validation examples – 10 for AHP and 3 for ANP – all falling in the first (predictive tool) category. This list is clearly only a small sample of many validation exercises done over the years but it can be assumed that it constitutes a fairly representative sample of such validation examples. This assumption is borne out when compared to the examples listed in [Saaty, 1997]. Whitaker [2004] states that the objectives of collecting theses examples are: (i) to promote the use of the AHP/ANP theory as a valid theory; (ii) to show the wide variety of scales against which validation can be accomplished, and (iii) to collect a wide variety of validation examples in one place. This is both a worthy and commendable endeavour and invites an ongoing review of the principles of AHP/ANP validation.

2.4 Evaluating the Proofs

When evaluating the Validity Examples two points are important, points that can be linked to the split between *Predictive* and *Decision-Making* tools. First, clearly, the AHP/ANP is meant to be more than a tool to *predict* known outcomes – however, varied these may be. Second, it is similarly clear that research on validating the AHP/ANP as a *decision-making* tool – in the thorny world of subjective preferences, values, and perspectives where known outcomes are not available – is surely needed. This two-way split will be used here as a guide to presenting the evaluation of the Validity Examples.

2.4.1 Predicting Known Outcomes and Generalizibility

Validity is subdivided into discussions on Internal and External Validity where *internal* validity refers to whether the research conclusions are adequately supported by the data, and *external* validity refers to whether the findings of a given research project are generalizable to all similar cases [Babbie, 2007].

Within the GDM research literature (e.g. [Laughlin, 1996]) two types of tasks are identified. A task with a demonstrable correct answer is called an *Intellective* task while a *Judgmental* task is one without such a demonstrable correct answer. In many cases tasks are neither purely Intellective nor purely Judgmental, for example the 'Moon Survival' task which is intermediate between *Intellective* and *Judgmental* because the correct result exists (due to an expert panel rating), but although this correct result is available it cannot be *demonstrated* within the groups while in session. In [Whitaker, 2004] the Validation Examples all have correct (known) outcomes but are not equivalent to Intellectual tasks as the correctness cannot be *demonstrated* by the decision-makers until afterwards. This is important for the validation of the AHP/ANP in that the GDM literature indicates, further, that the decision outcomes differ in terms of the type of task performed [Gigone and Hastie, 1997; MacCoun, 1998]. Kirchler and Davis [1986] point out that for Intellective tasks in GDM outcomes are normally a '*truth* wins' situation, while in Judgmental tasks this changes to '*majority* wins'. In the latter cases, if group member status is very varied, outcomes change further to become '*power* wins'. This has an effect on the generalizibility of the Validation Examples in two ways. First, the changes in outcome types indicate that generalizing from one task type to a domain including all task types may be suspect (cf [Harper et al, 1992]). Second, and more important, is the context within which the GDM debate takes place. The Validation Examples in [Whitaker, 2004] all seem to have taken place in, what Flood and Jackson [1991] call, *Unitary* contexts. The Flood & Jackson *Unitary-Pluralist-Conflictual* continuum will be discussed below; now it is important to realize that in Unitary contexts no pressure exists to argue for, or maintain allegiance to, specific values, preferences or views and it seems logical to expect 'majority wins' outcomes, while in the Conflictual contexts pressure clearly exists to strongly side with a particular viewpoint and this could lead to a change to 'power wins' outcomes. Applying the AHP/ANP in multi-stakeholder GDM as envisaged here and elsewhere ([von Solms and Peniwati, 2001; von Solms, 2009]) would seem to require validation over a wider front – including GDM in pluralist and conflictual contexts – over and above the unitary contexts covered in the extant literature (e.g. [Saaty, 1997; Whitaker, 2004]).

2.4.2 Research: Rigor vs Relevance

Whereas the previous discussion related mainly to the Validity Examples, the current point is relevant to all three Validity Arguments – including the *Validity* Examples, the *Counter* Examples and the *Applications* argument – and involves the question regarding the *practical* value of the AHP/ANP as opposed to its *formal* mathematical foundation. Forman and Gass [2001] argue for the point that the many successful applications of the AHP/ANP are proof of its validity. Interestingly, many critics acknowledge the successful application of the AHP/ANP before documenting their criticisms (e.g. [Pérez, 1995; Salo and Hämäläinen, 1997; Barzilai, 1999]). Some critics explicitly state that their critique is aimed *not* at the value of the AHP/ANP to decision makers, but specifically at mathematical issues of its underlying theory. For example, Barzilai and Golani [1994:57] say (italics in the original):

> "This is an *axiomatic* (*i.e. mathematical*) framework, which enables us to gain insight into problematic aspects of the underlying structure and to identify a consistent variant of the AHP. The framework does not address behavioural issues and is therefore independent of the way in which decision makers express their preferences."

In other cases the mathematical focus is implicit. It is clear, for example, that the first four criticisms mentioned in [Saaty, 2008] are relevant to mathematical and

theoretical issues of the AHP/ANP. It is, therefore, natural to find that the responses (e.g. [Saaty, 1990; Harker and Vargas, 1990; Vargas, 1997]), too, focus on defending these *formal* aspects. The fifth Criticism mentioned in [Saaty, 2008] does not, on first blush, seem to relate to a mathematical issue. In the more elaborate discussion in [Saaty et al, 2009], however, it becomes clear that the criticism leveled at the AHP/ANP is that its process of pairwise priority elicitation is not aligned with the mathematical calculations (e.g. the Eigenvector and Normalizations) performed thereon. This criticism – presented inter alia in [Watson and Freeling, 1982; Belton and Gear, 1983; Dyer, 1990] – argues that the decision makers do not understand the pairwise comparison process because what they are (behaviorally) doing is not in line with what the underlying mathematics require. This point will not be elaborated here, suffice to say that this criticism also focuses on a formal aspect of the AHP/ANP and as such lies outside of the focus of this paper.

What does need consideration is the *Rigor-Relevance* debate in the Research Methodology literature as it clearly has an impact on the current theme. Research Methodology discussions of research *rigor* and *quality* are normally done in terms of Validity, Reliability and Generalizability [Mouton and Marais, 1990; Babbie, 2007]. Evaluating the validity of the AHP/ANP, hence, brings the question of the Rigor-Relevance debate to the fore.

Baldridge et al [2004] indicate that the controversy regarding the practical relevance of academic research is a familiar topic in management articles, books and special issues of academic journals and while this debate has many elements, one of the core questions is whether academic quality (rigor) and practical relevance are mutually exclusive or mutually reinforcing. Some argue that academicians and practitioners hold *irreconcilable* different views of what constitutes academic quality and relevant research. Others view the relationship between relevance and academic quality as mutually reinforcing – arguing that if academic theories are not sound or well tested, they have little to offer practice. A similar debate between *Pure* Research and *Action* Research rages around the question of Rigor vs Relevance related to the dual goals of Action Research - i.e. the production of *Action* and *Knowledge* [Susman and Evered, 1978]. Eden and Huxham [1996] argue that Action Research is exactly about providing research output that meets the *dual requirements* of rigor and relevance. Rigor and Relevance are often – particularly from a Positivist perspective - seen as being in an inverse relationship and that increased relevance inevitably is achieved at the expense of methodological rigor [Susman and Evered, 1978]. Greenwood and Levin [2007] say that the operating assumptions in the conventional social sciences are that greater relevance and engagement automatically involves a loss of scientific validity.

The dual objectives of Action Research lead to the critique that the inclusion of an *Action* component dilutes (contaminates?) the spirit of true research. A strong argument against research aimed at solving practical problems has been a life-long theme of Fred N Kerlinger. Kerlinger [1979:280] says:

> "It is thought that research in sociology and psychology should be aimed at solving amongst other things problems of prejudice and discrimination, repairing learning deficiencies, improving learning and teaching, ... In general, however, such practical and worthy goals are not the purpose of scientific research. The purpose is theory, and it has no other purpose."

Elsewhere, Kerlinger [1977:6] states:

> "Scientific research never has the purpose of solving human or social problems, making decisions, and taking action."

The critique of Action Research - that the action component is *contrary* to the true spirit of research - is countered by Action Research proponents by arguing that it is exactly the action component that is one of the reasons for Action Research to be regarded as *superior* to traditional research [Eden and Huxham, 1996; Greenwood and Levin, 2005, 2007]. Particularly, C West Churchman and Ian Mitroff argue that a stance of disinterested objectivity is not only wrong, but the cause of much human tragedy and what is needed is ethical involvement in practice. In Mitroff's [1994:96] own words:

> "…humankind certainly faces a fundamental problem. (Ethics and management are inseparable in the ways that West and I conceive of them.) All the scientific studies, models, and words in the world, … fail to produce significant ethical actions (that is, implementation) to stop a situation that by any standard is morally outrageous. … we can probably generalize safely the following sad conjecture: in most cases, scientific studies, models, words and so forth fail to produce ethical actions to eradicate some important human problem."

In sum: The criticisms of the AHP/ANP, the rebuttals, and validity examples almost exclusively address formal, mathematical aspects of AHP/ANP theory (cf [Millet, 1997]). An important question presents itself at this point: Should the AHP/ANP be evaluated against the formal mathematical elegance of the underlying theory (i.e. rigor) or against the practical relevance of the method in real world decision-making? The answer to this question must, however, be delayed until after the introduction and brief discussion of traditional views on validity and how the quantitative-qualitative debate has influenced – and changed – these views.

3. Validity and the Philosophy of Science

3.1 Traditional Validity

Traditional discussions of validity usually involve a distinction between *Internal* and *External* validity. Internal validity is defined as referring to the fact that a study generates accurate and valid findings of the specific phenomena being studied. Therefore, we refer to a project as having produced internally valid results if the constructs were measured in a valid manner, the collected data are accurate and reliable, the analyses are relevant for the type of data, and the final conclusions are adequately supported by the data [Mouton and Marais, 1990; Babbie, 2007]. *External* Validity is defined as referring to a further stage in the research process, whether the findings of a given project are generalizable to all similar cases, and it would therefore be correct to regard *external validity* and *generalizability* as synonymous [Mouton and Marais, 1990; Babbie, 2007]. In many studies and research designs, there may be a trade-off between internal and external validity [Schram, 2005]. When measures are taken or procedures implemented aiming at increasing the chance for higher degrees of internal validity, these measures may also limit the generalizability of the findings. One way of enhancing internal validity is to perform highly controlled experiments including random sampling and Experimental vs Control Group designs. This situation leads, however, to an artificial situation divorced from the real world and this, in turn, leads many researchers to call for *ecologically* valid experiments.

Researchers criticize the lack of external validity in many laboratory-based studies with their focus on artificially controlled and constricted environments and the resultant lack in resembling real-world conditions [Schram, 2005]. The traditional criteria for validity, reliability and generalizability find their roots in the Positivist or Foundationalist tradition and the quantitative approaches to research [Guba & Lincoln, 1994]. McTaggart [1998] points out that the dominant discourse of validity hinges upon the combination of two key quests: the quest for *generalization*, and the quest for *causality* (which in other terms is a quest for *prediction* and *control* of events). On this view, research is valid if the researcher makes defensible *general* causal inferences, establishing - in the terms of Cook and Campbell [1979] – 'molar causal laws'.

3.2 Different Paradigms

Methodological Monism represents the culmination of the Enlightment project and represents the search for a universal, mathematically formulated science as the model of *all* science and knowledge. This has led to the view that (correctly) all research must be based on the methodology dominant in the Natural Sciences. Auguste Comte (1798-1857), the father of Positivism, in his famous *Law of Three Stages* postulated that societies and sciences move inevitably through three consecutive stages. First, a *Theological* stage in which effects are seen to be caused by supernatural agents; second, the transitional *Metaphysical* phase and finally culminating in the third or *Positive* phase in which man seeks to determine the *laws* which regulate effects. Comte believed that the natural sciences were already in the third phase but that the social sciences lagged far behind, and that what was needed was a social physics that would give the person exactly the same role that an atom plays in physics or a planet plays in astronomy, namely, absolute determination without freedom or cavil. John Stuart Mill (1806-1873) – a one-time disciple of Comte - urged social scientists to emulate the methods of the natural scientists and promising that, if his advice were followed, this would lead to the rapid maturation of these fields, as well as their emancipation from the philosophical and theological structures that limited them. Guba and Lincoln [1994] argue that these messages were taken to heart by subsequent social scientists to the extent that scientific maturity is commonly believed to be proportional to the level of quantification or mathematization in a given field. This view of science focuses on verifying or falsifying *a priori* hypotheses most usefully stated as mathematical (quantitative) propositions. Muckler and Seven [1992] say that such formulaic precision has enormous utility when the aim is the prediction and control of natural phenomena and that there exists a widespread conviction that only quantitative data are ultimately valid or of high quality – for example in [Sechrest, 1992].

Recently, strong counter-pressures against the hegemony of methodological monism and quantification have started to emerge. Muckler and Seven [1992] point out that the acceptance of subjectivity as inevitable pre-dates Plato in the views of Protagoras; is found in the philosophy of Descartes; while Kant insisted that all knowledge begins with human experience and cannot be separated from that experience. Burrell and Morgan [1979], in their classical work, clearly juxtaposes two approaches to social science methodology which they call *Sociological Positivism* (an Objectivist approach to Social Science) and *German Idealism* (a Subjectivist approach to Social Science). Tsoukas [1994] discusses Pepper's four *World Hypotheses* and shows that two of these, the *Mechanistic* and *Contextualist* systems, are comparable to the Burrell and Morgan [1979] paradigms of Positivism and Idealism respectively. Flood and Jackson [1991] suggest a continuum of contexts ranging from the Objectivist

Unitary, through the Subjectivist *Pluralist* to the Conflictional *Critical*. Guba and Lincoln [1994] present a very similar topology of social scientific paradigms, listing four, i.e. Positivism, Post-positivism, Critical Theory and Constructivism. Here, the split between Positivist and Post-positivist is mainly to accommodate the difference between Verificationist and Popperian Falsificationist modes of inquiry. Although originally aimed at different issues, the views of Burrell and Morgan [1979], Flood and Jackson [1991], Tsoukas [1994], and Guba and Lincoln [1994] have in common the argument that research and knowledge acquisition must be seen as divided at least into two classes, i.e. the *Quantitative* and *Qualitative* research, broadly identifiable with *Positivism* and *Phenomenology*, respectively. Qualitative inquiry is distinct and separate from quantitative inquiry. Each tradition reflects a commitment to different styles of research, different epistemologies and different representations [Guba and Lincoln, 1994]. According to Bogdan and Bilken [1982], qualitative inquiry evolved in response to the inadequacies of quantitative inquiry. Quantitative research focuses on causal relationships described in terms of observation statements, verifications and prediction, while qualitative research focuses on exploration of human behavior and the search for understanding through people's actions. Guba and Lincoln [1994] state that human behavior, unlike that of physical objects, cannot be understood without reference to the *meanings* and *purposes* attached by human actors to their activities. Therefore, data derived from qualitative inquiry are used to clarify the experience and understanding of a complex phenomenon *from the participant's point of view*.

3.3 Qualitative Validity

Although many researchers posit the distinct attributes of each tradition, there are those who contend that qualitative methods are not clearly differentiated from quantitative methods. The defining attributes of one mode of inquiry can also be found in the other. Some qualitative researchers align themselves with quantitative researchers by accepting and redefining their criteria of merit for use in qualitative research and use complicated measures such as computer applications in attempts to ensure objectivity, validity, and objectivity. In contrast, Trochim [2006] points out that, depending on their philosophical perspectives, some qualitative researchers reject the framework of validity that is commonly accepted in more quantitative research in the social sciences. They reject the basic realist assumption that there is a reality external to our perception of it. Consequently, it doesn't make sense to be concerned with the ‘truth’ or ‘falsity’ of an observation with respect to an external reality (which is a primary concern of validity). These qualitative researchers argue for different standards for judging the quality of research. For instance, Guba [1981] proposes four criteria - Credibility, Transferability, Dependability, and Conformability - for judging the soundness of qualitative research; explicitly offers these as an alternative to more traditional quantitatively-oriented criteria - Internal and External Validity, Reliability and Objectivity - and argues that these better reflect the underlying assumptions involved in much of qualitative research. Shrader-Frechette [1990] says that ever since the failure of Carnap's enterprise, epistemologists like Sellars and Quine have realized that the positivist goal of finding a specific rule or method to guarantee the rationality of science or knowledge is incapable of being achieved (cf [Feyerabend, 1978]). This implies that different approaches need to be evaluated in terms of their stated, or implied, epistemological stance rather than against one (superior?) set of epistemic standards. Although qualitative and quantitative methods are fundamentally different, there is a propensity to evaluate all approaches against criteria suitable for quantitative research only. According to Ryan-Nicholls and Will [2009], the traditional definition of rigor highlights important attributes of rigor arising in a predominately positivist discourse.

Rigorous research uses logic and accepted systems, and therefore strengthens the arguments researchers make when attempting to demonstrate the significance of findings. Rigor is part of the iterative, self-correcting nature of research in which the data collection, systematic analysis, and findings are open to evaluation and replication. A dilemma arises when quantitative researchers attempt to apply this definition - and others like it - to determine the rigor of qualitative work. When compared with predictive quantitative research designs that are usually focused on measurement and are meticulously organized around notions of equivalence, qualitative inquiry appears to lack the requisite rigor. Yet, as Ryan-Nicholls and Will [2009] go on to demonstrate, the problem is not the rigor of qualitative work but the application of inappropriate mechanisms of evaluation.

Within the qualitative tradition, *Phenomenological Research* describes the subjective reality of an event, as perceived by the study population. Although focused on the subjective it is very much still an emic approach. *Participatory Action Research* (PAR) – an emic form of qualitative investigation – is aimed at allowing the participants to be *fully* involved. For our current purpose the differences between traditional (Positivist) research approaches and PAR are illuminating. Dick and Swepson, [1994] point out that Action Research must be attended by validity measures relevant to the unique circumstances of this form of research. McTaggart [1998] argues that the key commitment in PAR is to the *negotiation* of both the research and action among all participants and validity can, therefore, only be achieved if there are appropriate *communicative structures* in place which allow participants to continue to *associate* and *identify* with the work. Greenwood and Levin [2005] feel that Action Research should be grounded in Pragmatism and say that validity claims are identified as 'warranted' assertions resulting from an enquiry process and conclude that co-generated knowledge is deemed valid if it generates warrants for action by the *participants* in the process. The action researcher should be under no illusion that he will be fully involved, with other participants, in the problem situation and in a social process that will change that situation. He must, therefore, take some responsibility for any practical outcomes that arise and seek to ensure that these represent perceived improvements *in the eyes of the participants*. Susman and Evered [1978] argue that in traditional scientific research the emphasis is on *predicting* outcomes but that in action research the point is to *actively make things happen*. Central to this position are the ideas of the voluntary human agent and that behind every action is individual choice based on human interest. Choice is central to taking action and action is central to human development.

3.4 Validity and Truth

Ryan-Nicholls and Will [2009] posit that the search for truth is much more elusive in qualitative vis-à-vis quantitative research. In the former, truth value is *subject*-oriented rather than *defined* by the researcher. Truth is found in the *discovery* of human phenomena or experiences as they are lived and perceived by subjects as opposed to the verification of *a priori* conceptions of such experiences.

Within the field of Epistemology many truth-theories exist, two of the long-standing views are the *Correspondence* and the *Coherence* model of truth [Bowden and Swartz, 2004]. The former - that can be traced back to Plato and Aristotle - is a realist theory that claims the truth of a proposition resides in its *correspondence* to a fact in the real world. The latter - linked to inter alia Leibniz, Spinoza and Hegel – claims that the truth of a proposition arises out of the relationship (*coherence*) between that proposition and other propositions that have been accepted before [Dowden and

Swartz, 2004]. Hammond [1996] discusses these two theories in terms of human judgment and decision-making and argues that judgment and decision theories differ in terms of their emphasis on a coherence or correspondence interpretation of human judgment competence. Coherence focuses on judgment competence with respect to logical, mathematical, or statistical criteria. What counts as a good judgment or decision from the coherence perspective is whether or not it is congruent with what some well-established and internally consistent set of rules or axioms would have produced. On the other hand, correspondence focuses on judgment competence with respect to empirical accuracy. What counts as a good judgment or decision from the correspondence perspective is how well judgments map onto eventual observable outcomes in the world. Hastie [2001] points out that *Judgment* Research is focused on the process of how as-yet obscure events, outcomes and consequences in the real world could be inferred. As such, judgment quality is based on accuracy, i.e. *correspondence* to facts or conditions in the real world. On the other hand, *Decision-making* Research is focused on preferential choice and action and as such measures decision quality by comparing decisions and behavior against prescriptions of rational, normative models, often taking the form of test for *coherence* of expectations, values, and preferences [Hastie, 2001]. In application to managerial decision-making, it becomes clear that different types of judgments and decisions may evoke a preference for one or the other type of competence. There is strong tendency for prescriptive and normative decision theories to emphasize coherence competence, leading therefore to an interest in biases (i.e. deviations from axiomatic predictions) in decision-making. Descriptive theoretical approaches, on the other hand, tend to emphasize correspondence competence, leading to an interest in the factors that influence how accurate (i.e. close to actual outcomes) a decision maker is [Hammond, 1996]. Gigone and Hastie [1997] argue that insufficient research attention has been given to the problem of accuracy in group decision contexts – a finding that highlights the emerging importance of correspondence competence. Their work, however, emphasizes the fact that only a small number of judgment tasks can effectively be measured in terms of accuracy. Only tasks that have either an objectively demonstrable correct answer, e.g. length of the Nile River, or the consensus judgments by a panel of experts, are of this type. The majority of real-world multi-stakeholder decisions would not be amiable to accuracy determination making the use of a correspondence approach to judgment accuracy interesting but of little value. Gigone and Hastie [1997] argue that coherence approaches will make use of principles of rational judgment and although they cannot address group judgment *accuracy* they can provide legitimate evaluations of group judgment *quality*.

Taket and White [1997] discuss the Correspondence Theory of Truth in respect to community Operations Research (OR). They are adamant that in such circumstances the diverging views of different participants are the critical issues and that these views can only be addressed from a post-structuralist (or post-modern [Taket and White, 1994]) perspective. Taket and White [1997:99] say:

> "In relation to representing issues or situations we are forced to reject the idea of a correspondence theory of truth, that reality is 'out there', waiting to be discovered and represented accurately. We accept instead the possibilities of specific local, personal and community forms of truth."

They agree with Rorty [1989] that a distinction must be made between the *world* and *truth* 'out there' and while the world exists independently of the human mind, truth cannot so exist. In Rorty's [1989:5] words:

> "The world is out there, but descriptions of the world are not. Only descriptions of the world can be true or false. The world on its own – unaided by the describing activities of human beings - cannot."

For Taket and White [1997], in methods purporting to aid decision structuring in multiple party participative situations all views must be accepted as valid and there would be no problem with conflicting visions because each vision is true in a different world. They distinguish *consent-giving* from *consensus* - where the former is less restrictive in that it allows cooperation amongst individuals without their necessarily agreeing on all aspects [Taket and White, 1994].

An approach opposing the 'single truth' view of the Enlightenment or Modernity is that of Constructivism [Schwandt, 1994]. The Philosophical Foundation of Constructivism is expressed in four statements; (i) knowledge is not passively accumulated, but rather, is the result of active cognizing by the individual; (ii) cognition is an adaptive process that functions to make an individual's behavior more viable given a particular environment; (iii) cognition organizes and makes sense of experience, and is not a process to render an accurate representation of reality; and (iv) knowing has roots in both biological/neurological construction, and social, cultural, and language based interactions. Social Constructionism grew out of the more general notion of Constructivism and both these perspectives can be located in the domain of post-modernism. Whereas the terms of pure (or Cognitive) Constructivism tend to isolate the thinking process within an individual's mind, Social Constructionism calls upon the active participation of others [Gergen, 2002]. Social Constructionism builds on Constructivism's view of knowledge as not objective, universal, or permanent, but it emphasizes knowledge as communal rather than individually created. The notion of Social Constructivism actively involves social interaction and collaboration in the learning and knowledge-making process and, hence, requires a particular definition of (true) knowledge. This definition is that knowledge is negotiated (through language) until *consensus* is reached on the group knowledge. Within groups the social construction of knowledge relies on consensus between different subjects which is seen as the ultimate criterion to judge knowledge [Gergen, 2002]. This form of truth is also called the *Consensus Theory* and holds that truth is whatever is agreed upon, or might come to be agreed upon, by some specified group. Rescher [1993] argues that a view positing *agreement as a prerequisite for truth* is evident – in different forms – throughout philosophical history, from Aristotle, Rousseau, Kant and Mill to Pierce, Habermas and Rorty. In Habermas' [1984] version of the theory, truth is not dependent on actual consensus. Rather, it is what would be agreed to by all investigators who follow principles of equal, undistorted, unconstrained communication and who adopt a discursive orientation. Thus, Habermas [1984] argues that the pragmatic meaning of a truth claim is that it could be verified or made good in a discourse oriented toward *rational* consensus and not arrived at under conditions of *distorted* (dominated) communication.

3.5 Validity and Subjectivity

The first major difficulty a view of decision-making validity from within a subjectivist paradigm seems to face, is the question of how a 'better' decision is identified – how do we know which decisions are better than others? The philosophical and sociological perspective from which the question is approached will lead to very different answers. Within the Positivist philosophy and its related sociology of Functionalism, objective reality is emphasized. Clear goals can be identified and agreement on means to achieve these goals is easy to obtain. 'Better' is measurable in terms of profit maximization or efficiency increases. The second

perspective, that of an Interpretive sociology and an Idealist philosophy emphasizes the subjective nature of reality. Different actors will perceive things differently based on each individual's own *Weltanschauung* and according to the way that he makes sense of reality. Accordingly, 'better' for one may well be seen as 'worse' by another. This perspective excludes the possibility of *objectively* evaluating different views by positing that nobody can think or evaluate anything other than from within his own worldview [Checkland, 1981; Flood and Jackson, 1991]. Ulrich [2002] points out the ethical problem of a 'we-know-what-is-best-for-everyone' view and warns against an 'expert' minority planning *for* instead of *with* the so-called 'lay' majority; while Anderson et al [1998] show that stakeholders are increasingly demanding an active role in decisions that affect them.

A debate that illustrates the issue at hand is that regarding empirical testing of the effectiveness of the Dialectic Inquiry (DI) approach [Cosier, 1982; Mitroff, 1982]. Cosier et al [1978] argue that the empirical evidence for the effectiveness of, what they call, the *Dialectical Inquiry Systems* (DIS) methodology is not as convincing as its theoretical and philosophical support and the limited field studies reported by its proponents have not demonstrated that the DIS technique leads to improvement in objective performance or more effective plans. Cosier et al [1978] carry out two controlled laboratory and field tests to compare results from DIS to results from an alternative planning approach, the Devil's Advocate (DA), but the results fail to show DI as unequivocally superior to DA [Cosier, 1982]. Mitroff and Mason [1981] argue that the critique is problematic as it assesses DI as an approach to *well-structured* problems rather than, as it was intended, an approach to *ill-structured* problems [Mitroff, 1982]. According to Mitroff and Mason [1981] DI is a Hegelian Inquiry system, based on a synthesis of multiple completely antithetical representations that are characterized by intense conflict because of the contrary underlying assumptions. The DI method cannot be tested by presenting decision makers with two *already formulated* worldviews, plans or policies, as true dialectic entails two or more groups actively participating both in *formulating and examining* of a problem from markedly different points of view [Mitroff and Mason, 1981; Mitroff, 1982; Flood and Jackson, 1991]. The similarity between the view of Mitroff and Mason [1981] and the current conceptualization of multi-stakeholder GDM is clear – both involving decision-making situations in which the problem needs formulation and debate from radically diverging points of view. The two empirical studies in [Cosier et al, 1978] must, from necessity, use objectively known outcomes to be able to test prediction accuracy. The first, based on Brunswikian Social Judgment Theory, involves subjective judgment on the part of the participants but the objectively correct outcome is known to the experimenters. Similarly the second experiment involves three contrived markets, again unknown to the participants but known to the investigators. The critique of Mitroff and Mason [1981] and Mitroff [1982] is that, although the evaluation involved subjective and uncertain *judgments*, the problems were not *formulated* by the participants themselves and this – in their view – is the essence of ill-structured problems; situations where different participants see *different problems* rather than situations where the participants may differ in their opinion as to the best way to solve a *given problem*. In their evaluation of this debate Flood and Jackson [1991] come to the conclusion that approaches like DI *cannot* be subject to empirical verification in the manner attempted by Cosier et al [1978] because the philosophy underlying DI-like approaches implies that ill-structured problems are, by definition, situations where *no* objective correct answers exist, but rather that different participants see different realities and the solution cannot be reached by convergence

on the *correct* reality but rather through debate, learning and the emergence of a *shared* – inter-subjective – reality.

4. A New Validity Problem

4.1 Is the AHP/ANP a Useful Decision-Making Methodology?

A criticism leveled at the AHP relates to whether it is useful in GDM situations. Aronson et al [1997] claim that their research indicates that there is *no* advantage in GDM when using AHP instead of unstructured applications and that claims for the success of the AHP is a situation they call 'much-ado-about-nothing'. This statement belies the many claims made in favor of AHP as a GDM method – for example [Dyer and Forman, 1992; Peniwati, 1999; Forman and Gass, 2001; Saaty and Peniwati, 2008], inter alia. In replying to this critique two aspects are relevant. First, the Aronson et al [1997] paper does not include any data regarding how the AHP was used exactly; no information is given regarding the hierarchies used, the priorities obtained, or the means of aggregation applied. The exact way in which the AHP was applied cannot be determined and a direct critique of the Aronson et al [1997] findings cannot be made. The emphasis is on testing the sharing of *common* and *unique* information based on the theoretical underpinnings of the Information Sampling Theory [Stasser and Titus, 2003] but the fact that the research detail is not available makes it impossible to evaluate whether, on the one hand, AHP really does not enhance information sharing – a fact that would have serious implications for the AHP/ANP as a GDM methodology – or on the other hand whether the lack of decision support ascribed to the AHP is caused by the way in which the AHP was applied, instead. Second, this research is based on positivist/functionalist premises. In order to establish when and to what extent errors occur, they assume an objectively correct result exists against which individual decisions can be compared. As such the outcomes of this research are of no use in multi-stakeholder GDM as envisaged in this paper. Clearly, no objective truth can be assumed under such conditions as it not only belies the pluralistic characteristics of multi-stakeholder groups but also makes the use of such groups redundant – an *expert panel* would be able to find this true decision without input from stakeholders. In contrast, a Pluralist approach to multi-stakeholder GDM is argued elsewhere [von Solms, 1999] and promoted here. Banville et al [1998] show that, although Stakeholder engagement and MCDM have many points of mutual reinforcing, there is a paucity of research and joint applications of these two fields. The criticism found in [Aronson et al, 1997] and comments made in [Harper et al, 1992] alert us to the fact that the AHP/ANP must be validated from a totally different perspective, i.e. as a GDM tool in a pluralist, subjectivist and qualitative context. To this task we now turn.

4.2 AHP/ANP and the Paradigms

The first issue to be discussed is where the AHP/ANP fit in relation to the Objectivist-Subjectivist paradigms [Burrell and Morgan, 1979]. Saaty [1997] argues that AHP is a *new paradigm* that some people find hard to accept. Science and reason improve the understanding of human existence and the human environment but the facts and understanding obtained through science and reason is fundamentally related to human values, needs, and to the judgments that serve these values and needs [Saaty, 2001]. The view that excludes the myriad of intangibles and qualitative human nature from decision-making, by insisting that only tangible and concrete objects, captured through science, can be measured or included in decision-making, is rejected. The world is replete with intangibles, and insistence on measuring only

tangibles excludes many of the goals and criteria important to political, social and environmental problems [Saaty, 1998]. According to Saaty [1998] the problem of measuring intangibles is the most important concern facing anyone who wants to grapple successfully with the mathematics of decision-making in all these categories of problems. The measurement of intangibles cannot be done - in the traditional scientific manner - on an absolute scale with an arbitrary unit of measurement, but must be done in relative terms because an intangible, by definition, does not have a scale with a unit [Saaty, 1998, 1999]. He states the problem [Saaty, 1998:14] thus:

> "Comparison requires the use of judgment. Even informed judgment is subjective. … Since relative judgment is more general and includes absolute judgment, it is not difficult to see that what we once though was an objective approach to reality is simply something we agree on, and that all – I mean all – interpretation of reality relates to our own sensory abilities and to our values and goals."

More recently, Saaty and his colleagues (e.g. [Saaty, 2008, 2010; Saaty and Peniwati, 2008]) have frequently reiterated the fact that human decision-making is by nature subjective and that this subjectivity must be incorporated into MCDM tools. Forman and Selly [2001] agree and point out that the fact of *multiple* criteria – even if the criteria represent tangibles - necessitates value judgments as to their relative importance, leading inevitably to subjectivity. Saaty [2010] argues that the AHP/ANP is similar to the natural sciences (physics specifically) in that both are descriptive rather than normative, but stresses that some significant differences also exist. He specifically mentions that, unlike physics which attempts to find general laws that transcend time and space, the AHP/ANP is applied to each problem *separately* because it is based on subjective judgments that differ from case to case and from one person to another.

High quality decisions in complex, ambiguous situations often require multiple perspectives, the expression of contrary viewpoints and the evaluation of multiple alternatives. Groups are often reluctant to focus on divergent objectives because of the complexity of processing multiple information streams [Clarke et al, 2000]. The AHP is particularly strong in this area as it was developed specifically as a multiple criteria decision-making methodology, allowing multiple perspectives, the expression of contrary viewpoints, and the evaluation of multiple alternatives [Saaty, 2001, 2008, 2010]. Petkov and Mihova-Petkova [1997] argue convincingly that AHP performs well as a systems method (cf [Saaty and Kearns, 1985]), in all three of the problem contexts – unitary, pluralist, and coercive [Flood and Jackson, 1991] - and that it enhances the expression of divergent viewpoints. They also link AHP and SSM [Checkland, 1981], a systems methodology applicable to the pluralist context, showing that both approaches allow subjectivity and multiple worldviews [Petkov and Mihova-Petkova, 1996; Petkov et al, 1998; Petkov et al, 2007]. Rosenhead [1996] disagrees and explicitly identifies the AHP as an Operations Research (OR) methodology - i.e. a Hard Systems approach - in opposition to Problem Structuring Methods (PSM) like SSM. The demarcation between AHP and the PSM approaches, according to Rosenhead [1996], are the latter's transparency of method, restricted mathematization and their focus on *supporting* judgment rather than *representing* it. Petkov and Mihova-Petkova [1997], on the other hand, argue that it is important in practice to *both* support and measure human judgment and this strengthens the case for AHP as a systems methodology.

4.3 AHP/ANP and Truth Theories

Cooksey [2000] typifies the AHP as a *coherence* method for he sees it as a *prescriptive* or *normative* approach relying on mathematical axioms. Contrary, Saaty [2008, 2010] and Whitaker [2004] argue strongly that the AHP/ANP is *not* a normative theory but is an attempt to develop priorities directly and naturally without insistence on any prescribed notions of rationality. From this it could be deducted that the AHP/ANP should be seen as a *correspondence* method, but this too would be problematic as the reality to which decisions must correspond is a constructed reality based on subjective evaluations, priorities, and values. The dilemma is clarified if we remember that Cooksey [2000] – like Hammond [1996] and Hastie [2001] – work in a paradigm of human judgment and decision-making in which *judgments* are evaluated in terms of correspondence to objectively known outcomes and *decision-making* in terms of coherence to normatively imposed standards. This paradigm is inadequate to judge the AHP/ANP as a decision-making methodology because it does not allow for the possibility of a descriptive approach where the success of a decision is measured against the subjective evaluations by the decision-makers involved [Banville et al, 1998].

4.4 AHP/ANP and Consensus

The inclusion of stakeholders, other than management, in decision-making is a frequent issue in contemporary management literature, particularly in the environmental management and sustainable development fields [Anderson et al, 1998; McEwan, 2001]. At the core of stakeholder decision-making is a controversial decision. Controversy typically arises, because those with a stake in the decision consequences disagree about the preferred decision alternative. Stakeholders can disagree because of differences in values or differences in beliefs about the decision consequences, or both [Anderson et al, 1998; Slovic, 1999]. Multi-stakeholder groups will, by their very nature, represent major difficulties in debate and reaching consensual group choices which could be seen as just and balancing the aspirations of all stakeholders [Slovic, 1999; Edmunds and Wollenberg, 2001; von Solms, 2009]. Particularly problematic is the use of consensus as choice aggregation in multi-stakeholder groups. Consensus was shown to be less effective and desirable in *competitive social contexts* [Tjosvold and Field, 1983]. Edmunds and Wollenberg [2001] point out that a pure Habermasian deliberative approach is inadequate because, although it argues for deliberation within an equitable environment, methods for real-world participative decision-making do not always ensure adequate protection of stakeholders against domination by other, powerful, stakeholders. The deliberative approach must, therefore, be bolstered with some form of protective mechanism. A different choice aggregation method is required that would lead to a fair group outcome even under divergent value systems and inequalities of power while allowing fair trade-offs on salience and relative importance of decision aspects. Elsewhere it is argued that the 3-phase variation of the AHP can serve this purpose [von Solms and Peniwati, 2001; von Solms, 2009].

4.5 AHP/ANP and Its Own Validity

It seems opportune to locate the use of the AHP in multi-stakeholder groups within a Social Constructivist-like approach [Schwandt, 1994; Banville et al, 1998] without committing fully to all the tenets of Social Constructivism. Its emphasis on active learning in a social context with the positive valuing of all participants' views seems a useful approach. The validity of decision outcomes, within this tradition, would then be measured in terms of the agreement of participants (consensus) as to their own (subjective) observation and evaluation of the outcome rather than against any

objectively given outcome (correspondence) or normatively imposed standards (coherence).

A large number of criteria by which decision-making can be evaluated are found in the literature. We will, here, briefly discuss four criteria that seem appropriate for GDM in a multi-stakeholder situation. These criteria are: (1) Decision Quality [Ettling and Jago, 1988; Christensen and Fjermestad, 1997; Brower, 2000]; (2) Process Fairness [Rubin, 1984; Peniwati, 1996; Lind and Tyler, 1988; Kim and Mauborgne, 1995, 1997]; (3) Participant Satisfaction [Rubin, 1984; Ettling and Jago, 1988; Lind and Tyler, 1988; Christensen and Fjermestad, 1997; Brower, 2000]; (4) Commitment to Implementation [Ettling and Jago, 1988; Brower, 2000; Greenwood and Levin, 2005]. Other criteria, found in the literature, will not be discussed, due to space limitations.

Process Fairness and *Participant Satisfaction* are very important in multi-stakeholder situations. The premise of research by Peniwati [1996] was that the quality of a decision outcome is determined by the quality of the process (method) used. This assumption holds true if the method is more than a social decision scheme but also provides for adequate social influence [Pavitt, 1993]. Peniwati [1996] adapts a set of GDM quality criteria from Rubin [1984] and includes, inter alia, *Learning*, *Fairness* - both in terms of the group members and other stakeholders - and rate the AHP as High or Very High on these issues. *Learning* represents social influence and, thus, confirms that the AHP is seen as providing adequate change for social interaction (cf [von Solms, 2003]). Procedural Justice Theory [Lind and Tyler, 1988] constitutes an alternative to Exchange Theory. The latter focuses on decision *outcomes* while the former is based on the hypothesis that for participants in decision-making processes, the procedures used to arrive at decisions are significant determinants of *satisfaction* separate from the effect of outcomes [Kim and Mauborgne, 1995, 1997]. Procedural Justice Theory hypothesizes that the perceived fairness of the process is constituted by two aspects, i.e. *voice* and *dignity*. *Voice* enhances procedural fairness when a procedure allows those subjects, concerned with the outcomes, a chance to express their views freely, while *Dignity* refers to procedures that treat group members in a dignified and respectful fashion, i.e. that all member contributions are honored and considered [Lind and Tyler, 1988]. *Commitment to Implement* the decision outcome is an important proof of the participants' acceptance of the validity of the decision-making process. Greenwood and Levin [2005:54] are adamant:

> "Validity, credibility, and reliability in action research is measured by the willingness of local stakeholders to act on the results of the action research thereby risking their welfare on the 'validity' of their ideas and the degree to which the outcomes meet their expectations. Thus, cogenerated knowledge is deemed valid if it generates warrants for action."

The criteria, *Fairness*, *Satisfaction* and *Commitment* are generally accepted as subjective and measured using surveys in which the participants express their own subjective views [Christensen and Fjermestad, 1997; Hacker, 1997; Brower, 2000]. Hacker [1997] points out that self-reported satisfaction measures may involve significant biases. She refers to an observation by Muckler and Seven [1992] that self-report data are abhorred by objective measurement advocates because misrepresentation and misperception are possible. Hacker [1997], however, states that most scholars still include self-report data regarding satisfaction. This is not surprising as bias can only be defined in terms of an objective measure and as

satisfaction is clearly a subjective perception for which no objective measure exists, satisfaction can only be known if the perceiver reports on it.

While the subjective nature of these three criteria is accepted, the situation is more complicated when decision-making *Quality* is considered. In the case of decision-making *Quality*, attempts are made in the literature to set up *objective* means of measuring and comparing Quality. Watson et al [1988] argue that in Group Decision Support Systems (GDSS) research the performance of the group is often compared to an objective measure of decision quality but that many organizational GDM occur without prior or post knowledge of the *'correct'* outcome. Only if quality is deemed objectively measurable does the possibility of identifying and eliminating biased results exist. Kruglanski and Ajzen, [1983] review the literature on bias and error in human judgment and list three ways in which such biases or errors are identified, i.e. *Normative Models*, *Direct Measurement* and *Investigator (or Expert) Judgment*. They point out that all three approaches implicitly assume the existence of an objectively verifiable correct result against which actual decisions can be compared to determine the existence of bias or error. The differences are only in the different *ways* in which this 'correct' result is uncovered [Kruglanski and Ajzen, 1983]. Funder [1990] criticizes the Kruglanskian *Lay Epistemic Theory* for focusing on *process* rather than on *content* (i.e. accuracy) of judgments. In his view the accuracy of a judgment cannot be evaluated by investigating the process by which judgments are made. Funder's [1990] view is that far too negative a picture of human judgment competence is presented by the error and biases researchers hampering progress in social psychological research. He argues that Accuracy research should be located in a *Realist* paradigm of an objectively knowable reality to which judgments can be compared in determining their accuracy. Funder and West [1993] clearly state their realist position when they say that inaccuracy can be inferred only if it is assumed that the property that the subjects have been asked to judge 'exists' and that one value of the property excludes the possibility of other values of the property. Funder [1990] emphasizes accuracy in social judgment to be fashioned on the objectivist Brunswikian approach to Psychology, i.e. that of *correspondence* of judgment to objective reality as defined by expert opinion. Whereas the work of Funder and his colleagues represent the *Investigator (Expert) Judgment* category of Kruglanski and Ajzen's [1983] three-category scheme, the work of Kahneman and Tversky [1979, 2000], on the other hand, represent the *Normative Models* category. This research paradigm is realist too, in that the decisions used are choices between simple monetary gambles with *objectively* specified probabilities and at most two non-zero outcomes. The use of expert judgment as the benchmark of a true outcome is widely advocated and used [Muckler and Seven, 1992; Christensen and Fjermestad, 1997; Brower, 2000]. The latter two papers, however, argue that although using expert judgment in this way is widespread, it is, in fact, *wrong* as real decision-making situations are rarely (if ever) of the type where the true result is known to experts and the groups are only to discover this true result. Brower [2000] criticizes the realist judgment research design as being based on *simple* or *contrived* situations in which either an objectively correct result exists or an expert panel can provide an outcome that is deemed to be the correct result. Muckler and Seven's [1992] review indicates that even within this paradigm, the search for objective measures are not always successful. This is a further indication that an attempt to cling to objectivity and objective measures is not a fruitful paradigm to adopt in GDM research. Multi-stakeholder GDM, particularly on environmental issues, cannot be seen as 'objective' and hence no right answer exists against which decision outcomes can be measured.

Thus, decision *quality* must also be measured in terms of how the participants experience and view the decision outcome.

5. Conclusion

The AHP/ANP is an important method in all types of decision-making situations, particularly in multi-stakeholder GDM, which is becoming more and more important in the world today [Petkov et al, 1998; Saaty and Peniwati, 2008]. An important line of research, discussed here, is the validation of the AHP/ANP. This is predominantly done using the *judgment* paradigm with known outcomes against which to compare the AHP/ANP outcomes. We conclude that within this paradigm the validity examples show the AHP/ANP to be a valid method. Further, the rebuttals of criticisms of the formal aspects of the AHP/ANP prove (to our satisfaction) its mathematical foundation to be valid. But these efforts need to be complimented with a *new* direction in research - Research to test the validity of the AHP/ANP in terms of the world of the multi-stakeholder group, where no known outcomes exist and it is paramount that the subjective preferences and values of the participants be acknowledged and incorporated in a fair and transparent process. We have argued for the need of new validity criteria to show the AHP/ANP as capable of assisting real decision-makers in addressing the pressing human problems of - inter alia - conflict, suffering and environmental degradation. We need to test the validity of the AHP/ANP over the whole gambit of decision-making applications. Saaty [2010:432] recommends:

> "…that multicriteria methods put greater emphasis on validation to acquire greater credibility in practice. Validation is much more difficult when all judgments depend on feelings alone without memory from the senses and when the criteria are all intangible. But there are other ways to improve the credibility of the outcome that have been discussed in the literature…"

The literature referenced by Saaty is [Whitaker, 2004] in which the AHP/ANP is evaluated, albeit from a limited perspective.

We have here evaluated and compared the AHP/ANP apple and have found it to be redder than other MCDM apples and we have argued that the oranges of multi-stakeholder GDM have not yet been adequately evaluated. Now we need to take up the challenge and move forward to measuring both apples *and* oranges.

REFERENCES

Anderson, J, Clément, J & Crowder, LV (1998). *Accommodating Conflicting Interests in Forestry Concepts Emerging from Pluralism*; Unasylva 49/3; 3-10.

Aronson, JE, Dennis, AR, Hilmer, KM & Stam, A (1997). *Analytical Hierarchy Process in Group Information Exchange and Decision Making: Much Ado about Nothing*; Proceedings of the 3rd Americas Conference on Information Systems, Indianapolis, IN; August.

Babbie, ER (2007). *The Practice of Social Research*; 11th Edition; Thomson/Wadsworth; Belmont, CA.

Baldridge, DC, Floyd, SW & Markoczy, L (2004). *Are Managers From Mars and Academicians From Venus? Toward an Understanding of the Relationship Between Academic Quality and Practical Relevance*; Strategic Management Journal, 25/11; 1063-1074.

Bana e Costa, CA & Vansnick, J-C (2008). *A Critical Analysis of the Eigenvalue Method Used to Derive Priorities in AHP*; European Journal of Operational Research 187/3; 1422-1428.

Banville, C, Landry, M, Martel**,** J-M & Boulaire, C (1998). *A Stakeholder Approach to MCDA*; Systems Research & Behavioral Science 15/1; 15-32.

Barzilai, J (1999). *On MAUT, AHP and PFM*; Proceedings 5th International Symposium on the Analytic Hierarchy Process; Kobe, Japan; August; 57-60.

Barzilai, J (Forthcoming). *Preference Function Modeling: The Mathematical Foundations of Decision Theory*; in Figueira, J; Greco, S & Ehrgott, M (Eds); *Trends in MCDA*; Springer Verlag; Dordrecht, The Netherlands.

Barzilai, J & Golany, B (1994). *AHP Rank Reversal, Normalization and Aggregation Rules*; INFOR 32/2; 57-63.

Belton, V & Gear, T (1983). *On a Shortcoming of Saaty's Method of Analytic Hierarchies*; Omega 11/3; 228-230.

Bogdan, R & Biklen, S (1992). *Qualitative Research for Education*; 2nd Edition; Allyn and Bacon; Boston, MA.

Bowden, B & Swartz, N (2004). *Truth*; The Internet Encyclopedia of Philosophy; On-line at http://www.iep.utm.edu/t/truth.htm

Brower, HH (2000). *Group Decision Making in the Boardroom: A Model for Effectiveness*; Proceedings 2000 Annual Conference; Midwest Academy of Management; Chicago, IL; March.

Burrell, G & Morgan, G (1979). *Sociological Paradigms and Organisational Analysis*; Gower; Aldershot, UK.

Checkland, PB (1981). *Systems Thinking, Systems Practice*; John Wiley & Sons; Chichester, UK.

Christensen, EW & Fjermestad, J (1997). *Challenging Group Support Systems Research: The Case for Strategic Decision Making*; Group Decision and Negotiation 6/1; 351-372.

Clarke, MA, Anand, V & Roberson, L (2000). *Resolving Meaning Interpretation in Diverse Decision-Making Groups*; Group Dynamics: Theory, Research and Practice 4/3; 211-221.

Cook, TD & Campbell, DT (1979). *Quasi-Experimentation: Design and Analysis Issues for Field Settings*; Rand McNally College Publication Company; Chicago, IL.

Cooksey, RW (2000). *Managerial Judgment and Decision Making*; in Dahiya, SB (Ed); *The Current State of Business Disciplines*: *Volume 5 – Management II*; Spellbound Publications; Rohtak, India.

Cosier, RA (1982). *Methods for Improving the Strategic Decision: Dialectic Versus the Devil's Advocate*; Strategic Management Journal 3/4; 373-374.

Cosier, RA, Ruble, TL & Aplin, JC (1978). *An Evaluation of the Effectiveness of Dialectical Inquiry Systems*; Management Science 24/14; 1483-1490.

Dick, B & Swepson, P (1994). *Appropriate Validity and its Attainment within Action Research: An Illustration Using Soft Systems Methodology*; Available On-line at http://www.scu.edu.au/schools/gcm/ar/arp/sofsys2.html

Dyer, JS (1990). *Remarks on the Analytic Hierarchy Process*; Management Science 36/3; 249-258.

Dyer, RF & Forman, EH (1992). *Group Decision Support with the Analytic Hierarchy Process*; Decision Support Systems 8; 99-124.

Eden, C & Huxham, C (1996). *Action Research for the Study of Organizations*; in Clegg, SR; Hardy, C & Nord, WR (Eds); *Handbook of Organizational Studies*; SAGE Publications; London, UK.

Edmunds, D & Wollenberg, E (2001). *A Strategic Approach to Multistakeholder Negotiations*; Development and Change 32/2; 231-253.

Ettling, JT & Jago, AG (1988). *Participation under Conditions of Conflict: More on the Validity of the Vroom-Yetton Model*; Journal of Management Studies 25/1; 1-22.

Feyerabend, PK (1978). *Against Method: Outline of an Anarchistic Theory of Knowledge*; Verso; London, UK.

Finan, JS & Hurley, WJ (2002). *The Analytic Hierarchy Process: Can Wash Criteria be Ignored?*; Computers & Operations Research 29/8; 1025-1030.

Flood, RL & Jackson, MC (1991). *Creative Problem-Solving: Total Systems Intervention*; John Wiley & Sons; Chichester, UK.

Forman, EH, & Gass, SI (2001). *The Analytic Hierarchy Process – An Exposition*; Operations Research 49/4; 469-486.

Forman, EH & Selly MA (2001). *Decision by Objectives: How to Convince Others that You are Right*; World Scientific Publications; Singapore.

Funder, DC (1990). *Process versus Content in the Study of Judgmental Accuracy*; Psychological Inquiry 1/3; 207-209.

Funder, DC & West, SG (1993). *Consensus, Self-Other Agreement and Accuracy of Personality Judgments: An Introduction*; Journal of Personality 61/4; 457-476.

Gergen, KJ (2002). *Beyond the Empiricist/Constructionist Divide in Social Psychology*; Personality and Social Psychology Review 6/3; 188-191.

Gigone, D & Hastie, R (1997). *Proper Analysis of the Accuracy of Group Judgments*; Psychological Bulletin 1; 149-167.

Greenwood, DJ & Levin, M (2005). *Reform of the Social Sciences and Universities Through Action Research*; in Denzin, NK & Lincoln, YS (Eds); *SAGE Handbook of Qualitative Research*; 3rd Edition; SAGE Publications; Thousand Oaks, CA.

Greenwood, DJ & Levin, M (2007). *Introduction to Action Research: Social Research for Social Change*; 2nd Edition; SAGE Publications; Thousand Oaks, CA.

Guba, EG (1981). *Criteria for Assessing the Trustworthiness of Naturalistic Inquiries*; Educational Communication and Technology Journal 29/2; 75-92.

Guba, EG & Lincoln, YS (1994). *Competing Paradigms in Qualitative Research*; in Denzin, NK & Lincoln, YS (Eds); *Handbook of Qualitative Research*; SAGE Publications; Thousand Oaks, CA.

Habermas, J (1984). *The Theory of Communicative Action: Reason and Rationalisation of Society*; Volume 1; Polity Press; Cambridge, MA.

Hacker, ME (1997). *The Effect of Decision Aids on Work Group Performance*; Unpublished PhD Thesis; Virginia Polytechnic Institute and State University; Blackburg, VA.

Hammond, KR (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*; Oxford University Press; New York, NY.

Harker, PT & Vargas, LG (1990). *Reply to 'Remarks on the Analytic Hierarchy Process' by JS Dyer*; Management Science 36/3; 269-275.

Harper, RM, Apostolou, NG & Hartman, BP (1992). *The Analytic Hierarchy Process: An Empirical Examination of Aggregation and Hierarchical Structuring*; Behavioral Research in Accounting; 4/1; 96-112.

Hastie, R (2001). *Problems for Judgment and Decision Making*; Annual Review of Psychology 52/1; 653-683.

Holder, RD (1990). *Some Comments on the Analytic Hierarchy Process*; Journal of the Operational Research Society 41/11; 1073-1076.

Ishizaka, A & Labib, A (2009). *Analytic Hierarchy Process and Expert Choice: Benefits and Limitations*; OR Insight 22/4; 201-220.

Kahneman, D & Tversky, A (1979). *Prospect Theory: An Analysis of Decision under Risk*; Econometrica 47/2; 263-291.

Kahneman, D & Tversky, A (Eds) (2000). *Choice, Values and Frames*; Cambridge University Press; New York, NY.

Kerlinger, FN (1977). *The Influence of Research on Education Practice*; Educational Researcher 6/8; 5-12.

Kerlinger, FN (1979). *Behavioral Research: A Conceptual Approach*; Holt, Rinehart and Winston; New York. NY.

Kim, WC & Mauborgne, RA (1995). *A Procedural Justice Model of Strategic Decision Making: Strategy Content Implications in the Multinational*; Organization Science 6/1; 44-61.

Kim, WC & Mauborgne, RA (1997). *Fair Process: Managing in the Knowledge Economy*; Harvard Business Review 75/4; 65-75.

Kirchler, E & Davis, JH (1986). *The Influence of Member Status Differences and Task Type on Group Consensus and Member Position Change*; Journal of Personality and Social Psychology 51/1; 83-91.

Kruglanski, AW & Ajzen, I (1983). *Bias and Error in Human Judgment*; European Journal of Social Psychology 13/1; 1-44.

Laughlin, PR (1996). *Group Decision Making and Collective Induction*; in Witte, E & Davis, JH (Eds); *Group Behavior: Consensual Behavior by Small Groups*; *Volume 1*; Lawrence Erlbaum Associates; Mahwah, NJ.

Leskinen, P (2000). *Measurement Scales and Scale Independence in the Analytic Hierarchy Process*; Journal of Multi-Criteria Decision Analysis 9/4; 163-174.

Lind, EA & Tyler, TR (1988). *The Social Psychology of Procedural Justice*; Plenum Press, New York, NY.

MacCoun, RJ (1998). *Biases in the Interpretation and Use of Research Results*; Annual Review of Psychology 49/1; 259-287.

Ma, D & Zheng, X (1991). *9/9-9/1 Scale Method of AHP*; Proceedings 2nd International Symposium on AHP; Pittsburgh, PA; August.

McEwan, T (2001). *Managing Values and Beliefs in Organisations*; Financial Times/Prentiss Hall; Harlow, UK.

McTaggart, R (1998). *Is Validity Really an Issue for Action Research?*; Studies in Cultures, Organizations and Societies 4/2; 211-236.

Millet, I (1997). *The Effectiveness of Alternative Preference Elicitation Methods in the Analytic Hierarchy Process*; Journal of Multi-Criteria Decision Analysis 6/1; 41-51.

Mitroff, II (1982). *Talking Past One's Colleagues in Matters of Policy*; Strategic Management Journal 3/4; 374-376.

Mitroff, II (1994). *The Cruel Science of World Mismanagement: An Essay in Honor of C. West Churchman*; Interfaces 24/4; 94-98.

Mittroff, II & Mason, RO (1981). *The Metaphysics of Policy and Planning: A Reply to Cosier*; Academy of Management Review 6/4; 649-651.

Mouton, J & Marais, HC (1990). *Basic Concepts in the Methodology of the Social Sciences*; Human Sciences Research Council; Pretoria, RSA.

Muckler, FA & Seven, SA (1992). *Selecting Performance Measures: 'Objective' versus 'Subjective' Measurement*; Human Factors 34/4; 441-455.

Pavitt, C (1993). *Does Communication Matter in Social Influence During Small Group Discussion?: Five Positions*; Communication Studies 44/3-4; 216-227.

Peniwati, K (1996). *The Possibility Theorem for Group Decision Making: The Analytic Hierarchy Process*; Unpublished PhD Thesis; University of Pittsburgh; Pittsburgh, PA.

Peniwati, K (1999). *A Unifying Theory for Social Choice and Synergetic Group Decision Making: The Analytic Hierarchy Process*; Proceedings ISAHP 1999; Kobe, Japan; August; 153-158.

Pérez, J (1995). *Some Comments on Saaty's AHP*; Management Science 41/6; 1091-1095.

Pérez, J, Jimeno, JL & Mokotoff, E (2006). *Another Potential Strong Shortcoming of AHP*; TOP: Journal of the Spanish Society of Statistics and Operations Research 14/1; 99-111.

Petkov, D, McEwan, T, von Solms, SH & Vezjak M (1998). *Moving Towards Compliance with Standards for Environmental Protection - An Example of Mixing Approaches to 'Messy' Problems*; Proceedings 14th European Meeting on Cybernetics & Systems Research; Vienna, Austria; 252-257.

Petkov, D & Mihova-Petkova, O (1996). *Problem Structuring in the Analytic Hierarchy Process and Soft Systems Methodology*; Proceedings 4th International Symposium on the Analytic Hierarchy Process; Vancouver, Canada; July.

Petkov, D & Mihova-Petkova, O (1997). *The Analytic Hierarchy Process and Systems Thinking*; Proceedings 13th International MCDM Conference; Cape Town, RSA.

Petkov, D, Petkova, O, Andrew, T & Nepal, T (2007). *Mixing Multiple Criteria Decision Making with Soft Systems Thinking Techniques for Decision Support in Complex Situations*; Decision Support Systems 43/4; 1615-1629.

Rescher, N (1993). *Pluralism: Against the Demand for Consensus*; Clarendon Press; Oxford, UK.

Rorty, RM (1989). *Contingency, Irony and Solidarity*; Cambridge University Press; Cambridge, UK.

Rosenhead, J (1996). *What's the Problem? An Introduction to Problem Structuring Methods*; Interfaces 26/6; 117-131.

Rubin, JZ (1984). *Introduction*; in Swap, WC & Associates (Eds); *Group Decision Making*; SAGE Publications; Beverly Hills, CA.

Ryan-Nicholls, K & Will, C (2009). *Rigor in Qualitative Research: Mechanisms for Control*; Nurse Researcher 16/3; 70-85.

Saaty, TL (1990). *An Exposition of the AHP in Reply to 'Remarks on the Analytic Hierarchy Process'*; Management Science 36/3; 259-268.

Saaty, TL (1997). *That is Not the Analytic Hierarchy Process: What the AHP Is and What It Is Not*; Journal of Multi-Criteria Decision Analysis 6/6; 320-339.

Saaty, TL (1998). *Reflections and Projections on Creativity in Operations Research and Management Science: A Pressing Need for a Shift in Paradigm*; Operations Research 46/1; 9-16.

Saaty, TL (1999). *The Seven Pillars of the Analytic Hierarchy Process*; Proceedings 5th International Symposium on the Analytic Hierarchy Process; Kobe, Japan; August; 20-33.

Saaty, TL (2001). *Decision-Making with Dependence and Feedback: The Analytic Network Process*; 2nd Edition; RWS Publications; Pittsburgh, PA.

Saaty, TL (2008). *Relative Measurement and Its Generalization in Decision Making: Why Pairwise Comparisons are Central for the Measurement of Intangible Factors – The Analytic Hierarchy/Network Process*; Review of the Royal Spanish Academy of Sciences, Series A, Mathematics 102/2; 251-318.

Saaty, TL (2010). *Principia Mathematica Decernendi: Mathematical Principles of Decision-Making*; RWS Publications; Pittsburgh, PA.

Saaty, TL & Kearns, KP (1985). *Analytical Planning: The Organization of Systems*; Pergamon Press; Oxford, UK.

Saaty, TL & Peniwati, K (2008). *Group Decision Making: Drawing Out and Reconciling Differences*; RWS Publications; Pittsburgh, PA.

Saaty, TL & Vargas, LG (2006). *The Analytic Hierarchy Process: Wash Criteria Should Not be Ignored*; International Journal of Management & Decision Making 7/2-3; 180-188.

Saaty, TL, Vargas, LG & Whitaker, R (2009). *Addressing With Brevity Criticisms of the Analytic Hierarchy Process*; International Journal of the AHP 1/2; On-Line at http://www.ijahp.org/index.php/IJAHP/article/view/53/30

Salo, AA & Hämäläinen, RP (1997). *On the Measurement of Preferences in the Analytic Hierarchy Process*; Journal of Multi-Criteria Decision Analysis 6/6; 309-319.

Schram, A (2005). *Artificiality: The Tension Between Internal and External Validity in Economic Experiment*; Journal of Economic Methodology 12/2; 225-237.

Schwandt, TA (1994). *Constructivist, Interpretivist Approaches to Human Inquiry*; in Denzin, NK & Lincoln, YS (Eds); *Handbook of Qualitative Research*; SAGE Publications; Thousand Oaks, CA.

Sechrest, L (1992). *Roots: Back to Our First Generations*; Evaluation Practice 13/1; 1-7.

Sechrest, L (2005). *Validity of Measures Is No Simple Matter*; Health Services Research 40/5; 1584-1604.

Shrader-Frechette, KS (1990). *Scientific Method, Anti-Foundationalism and Public Decisionmaking*; Risk: Issues in Health and Safety 1/4; 341-363.

Slovic, P (1999). *Trust, Emotion, Sex, Politics and Science: Surveying the Risk Assessment Battlefield*; Risk Analysis 19/4; 689-701.

Stasser, G & Titus, W (2003). *Hidden Profiles: A Brief History*; Psychological Enquiry 14/3-4; 304-313.

Susman, GI & Evered, RD (1978). *An Assessment of the Scientific Merits of Action Research*; Administrative Science Quarterly 23/4; 582-603.

Taket, AR & White, LA; (1994). *Doing Community Operational Research with Multicultural Groups*; Omega 22/6; 579-588.

Taket, AR & White, LA; (1997). *Wanted: Dead OR Alive – Ways of Using Problem Structuring Methods in Community OR*; International Transactions in Operational Research 4/2; 99-108.

Tjosvold, D & Field, RHG; (1983). *Effects of Social Context on Consensus and Majority Vote Decision Making*; Academy of Management Journal 26/3; 500-506.

Trochim, WMK (2006). *Qualitative Validity*; in Trochim, WMK; *The Research Methods Knowledge Base*; A Web-based Textbook; This Chapter Available On-line at http://www.socialresearchmethods.net/kb/qualval.php

Tsoukas, H (1994). *Refining Common Sense: Types of Knowledge in Management Studies*; Journal of Management Studies 31/6; 761-780.

Ulrich, W (2002). *Critical Systems Heuristics*; in Daellenbach, HG & Flood, RL (Eds); *The Informed Student Guide to Management Science*; Thomsons Learning; London, UK.

van den Honert, RC & Lootsma, FA (1996). *Group Preference Aggregation in the Multiplicative AHP: The Model of the Group Decision Process and Pareto Optimality*; European Journal of Operational Research 96/2; 363-370.

Vargas, LG (1997). *Comments on Barzilai and Lootsma – Why the Multiplicative AHP is Invalid: A Practical Counterexample*; Journal of Multi-Criteria Decision Analysis 6/3; 169-170.

von Solms, SH (1999). *A Participative Approach to Environmental Impacts Assessment*; TAPPSA Journal; November; 34-36.

von Solms, SH (2003). *Group Polarization, Social Influence and the Analytic Hierarchy Process*; Proceedings 7th International Symposium on the Analytic Hierarchy Process; Bali, Indonesia; August; 475-484.

von Solms, SH (2009). *Homogeneity and Choice Aggregation in the Analytic Hierarchy Process*; Proceedings 10th International Symposium on the Analytic Hierarchy Process; University of Pittsburgh, Pittsburgh, PA; August.

von Solms, SH & Peniwati, K (2001). *To Agree or Not to Agree, That is the Question: Choice Aggregation in the AHP*; Proceedings 6th International Symposium on the Analytic Hierarchy Process; Bern, Switzerland; August.

Watson, RT, DeSanctis, G & Poole, MS (1988). *Using a GDSS to Facilitate Group Consensus: Some Intended and Unintended Consequences*; MIS Quarterly 12/3; 463-478.

Watson, SR & Freeling, ANS (1982). *Assessing Attribute Weights*; Omega 10/6; 582–583.

Whitaker, R (2004). *Validation Examples for the Analytic Hierarchy Process and the Analytic Network Process*; Proceedings 17th International Conference on Multiple Criteria Decision Analysis; Whistler, BC, Canada; August.

Whitaker, R (2007). *Criticisms of the Analytic Hierarchy Process: Why They Often Make No Sense*; Mathematical and Computer Modelling 46/7-8; 948-961.

Wijnmalen, DJD (2001). *Improved Structural Weight Adjustment in Top Down Oriented Conventional AHP Hierarchies*; Proceedings 6th International Symposium on the Analytic Hierarchy Process; Bern, Switzerland; August.