DRAFT1/22/90
A flexible six-step program for defining and handling bias in knowledge elicitation

MARY A. MEYER, JANE M. BOOKER, JEFFREY M. BRADSHAW*
MS F600, Statistics Group, Los Alamos National Laboratory
Los Alamos, New Mexico 87545, USA

**Advanced Technology Center, Boeing Computer Services*
P.O. Box 24346, M/S 7L-64, Seattle, Washington 98124 USA

While concern with the quality of expert knowledge is sensitizing knowledge engineers to the issue of bias, little guidance yet exists on how to recognize or handle its occurrence in knowledge acquisition. Bias is regarded here as a skewing of the data from some reference point. If the reference point is mathematical or logical rules, then biased data violates these rules, such as occurs with cognitive bias. If the reference point is the *?*expert's expressions of his thinking, biased data are distortions of these expressions, such as in the case of motivational biases. In this paper, we provide information on how to select a definition of bias (cognitive or motivational) and how to choose a combination of steps for handling bias in a particular application. The steps are (1) anticipate which biases are likely to occur in the planned elicitation; (2) redesign the elicitation to make it less prone to these biases; (3) make the experts aware of the potential intrusion of these biases and familiarize them with the elicitation procedures; (4) monitor the elicitation for the occurrence of biases; (5) adjust, in real time, to correct for their occurrence; and (6) analyze the data to determine if they occurred. The program includes detailed descriptions of how to perform each of these steps using selected biases as examples.

1. Introduction

The presence of bias in expert judgment and in knowledge bases has recently been recognized, but few guide-lines yet have been developed for handling its occurrence. Bias in knowledge acquisition is of concern because it can effect the quality and credibility of the knowledge. By its definition, biased data is viewed as undesirable, as skewed in comparison to unbiased data. Thus the presence of bias can cause discrepancies that confuse and slow knowledge acquisition. Bias can also enter into the knowledge base and impair the advice given by the knowledge-based system.

Programs are needed to help knowledge engineers deal with the potential problems posed by bias. By programs, we mean a set of steps that the knowledge engineers can perform to achieve certain ends with respect to bias. Such programs would include information on when bias was likely to be of concern and when particular steps would be beneficial We envision such programs as serving several purposes:

1. bringing the possible occurrence of bias to knowledge engineers' attention and stimulating them to consciously consider whether they need to be concerned about bias in each new project.

2. causing them to consciously examine what they consider desirable in the data and to explicitly set, rather than assume, the reference point from which bias will be defined. This reference point, be it mathematical/logical rules or the expert's unaltered thinking, can then serve as a base for selecting the elicitation methods (e.g., those which are less likely to create the bias) and the later means for validating or verifying the knowledge base.
3. training knowledge engineers in bias and hopefully lessening their tendencies to introduce bias in their elicitation, representation and analyses of the knowledge.
4. offering a series of steps on what to do about bias. The knowledge engineers would be guided to follow different steps depending on whether they wished to simply become aware of bias, counter its occurrence, or analyze its effects.

In this paper, expert knowledge will include *JB-rev*delete the word “only” here*JB-rev** elicited knowledge, such as the experts' solutions to problems and their descriptions of their mental processes (definitions, assumptions, and algorithms) in arriving at solutions. In addition, knowledge will refer to information that *JB-rev*experts have*JB-rev** given in numeric form, such as probability estimates, and nonnumeric form, such as *JB-rev*their descriptions of their*JB-rev** assumptions. We view expert knowledge as dynamic, *JB-rev*delete “as”*JB-rev** changing as the expert receives new information (Ortiz, Wheeler, Meyer, and Keeney, 1988, p. 2) *JB-rev* and, indeed, even as the expert. is engaged in describing what they know (Shafer & Tversky, 1985). Gaines (1989) has provided a foundation for understanding how an expert's knowledge and performance evolve through positive feedback processes that work on both the individual and the societal level. He proposes that expertise is continuously developed through the expert's exposure to interesting problems to solve and to the data of others in the professional community. Random differences in early performance may be amplified as the most interesting problems are given to those with a previously strong track record. Invitations to scientific congresses, scholarships, etc. may further magnify these effects. The role of experts in society can thus be seen as trading their advice for access to problem-solving, and hence learning, opportunities.*JB-rev**

The evolving nature of expertise has implications for knowledge acquisition. The knowledge acquisition process must reflect the changing nature of knowledge and allow for continual updating of the knowledge base or system (Gaines, 1989). Benbasat and Dhaliwal (1989) describe knowledge acquisition as occurring in phases where the knowledge base is repeatedly refined and where different kinds of validation are performed at the different phases. The approach to handling bias being proposed in this paper is envisioned as fitting into this dynamic environment of evolving knowledge bases or knowledge-based systems. Particularly, the approach to bias could be used during any phase of knowledge acquisition when expert knowledge is being elicited. Then too, it could be used as part of the validation process proposed by Benbasat and Dhaliwal (1989), particularly when the validation check calls for comparing the knowledge in the knowledge base or system to that of the expert source.

1.1. THE MEANING OF BIAS

Bias cannot be understood without a context because it is identified as *bias* in comparison to some reference point. Bias is generally defined as a *skewing* so its definition includes the concept of its counterpart--that which is considered *unskewed* or *not biased.* Implicit in this general definition of bias is the view that what is *not* biased is *reality* or *truth*. For this reason, a brief discussion of the nature of reality and of human perception is necessary to clarify the meaning of *bias*.

In this paper, we hold the *JB-rev* assumption of "constructive alternativism" (Kelly, 1955). According to this view, 'reality' does not reveal itself to us directly, but rather is subject to as many different constructions as we are able to invent (Adams-Webber, 1989; Agnew & Brown, 1989; Ford, 1989). Bateson (1979, p. 30) gives the example of someone stepping on his toe. He argues that what he experiences is not the stepping on of his toe but the image of this event from neural reports reaching his brain after the event has happened. He concludes that there is always " a transformation, a coding between the report, and the thing reported..." (1979, p. 30). Not only the quality of our emotions but also the actual degree to which we experience pain is determined by our attributions about the context of why and how and by whom the act was performed. Our memory for the event, our perceptions of the present, and our anticipations of the future can never be made "objective" because they arise from an invevitably subjective and personal model of events (Bradshaw & Boose, 1990). * As Shaw and McIntyre (1974) obeserve: "Ideas are not in the mind, nor objects in the world, but… both are in the meeting of mind and matter".

Similarly, we cannot decide objectively what is biased and what is not *JB-rev**. We can only make value judgments as to what we consider desirable (e.g., more objective) in the data. *JB-rev*Langer (1989, p. 200), argues that all our decision making is ultimately based on individual values (cf. Bradshaw & Boose, 1990). To illustrate that decisions are not made solely*JB-rev**on the basis of data, as we typically believe, she takes the example of those who must decide whether to prolong life when a patient is in intolerable pain. Receiving additional data does not enable the persons to make a decision because no amount of information will make one answer absolutely right. Then too, there is no logical stopping point to cue the persons on when to end gathering data. Eventually, the discussion of whether to prolong life returns to the basis of individual values. "The doctor, judge, and patient must decide between the principle of prolonging life at all costs and the 'right' to determine the quality of life." (p. 201)."

Because our decision making is based on value judgments, it is appropriate that the term *bias* have a negative connotation. Bias is what we label as undesirable or less good. Our view of bias can be likened to dirt in *JB-rev*Douglas'*JB-rev** anthropological portrayal of ritual pollution. "As we know it, dirt is essentially disorder. There is no such thing as absolute dirt: it exists in the eye of the beholder...Eliminating it is not a negative movement, but a positive effort to organise the environment.... In chasing dirt, in papering, decorating, tidying we are not governed by anxiety but are ...positively re-ordering our environment, making it conform to an idea. (*JB-rev**JB-rev**1970, p. 2).

---

*The fact that we usually take our models of reality for reality does not invalidate the above argument. As Bateson has stated (1979, p. 31) "Our civilization is deeply based in this illusion." At the cultural level, meaning and structure are imposed and then taken for reality by members of that culture. For example, members of the Western scientific culture would take the color spectrum, such as in a rainbow, and divide it into four to six colors--violet, blue, green, yellow, orange, and red. In another culture, the people would not see this segmentation of colors. Instead, they may have been conditioned to view the spectrum as consisting of wet and dry colors. The members of both of these cultures have been conditioned to see color in a particular way and to believe that it objectively exists as they perceive it.

JB-rev If we accept that defining bias is a value judgment, then we have the option of explicitly articulating presuppositions about the view we hold, so that the presuppositions themselves can be criticized and improved (Bateson, 1979, p. 26; Weimer, 1979, p.47).[*] One way in which these presuppositions can be improved is by testing them as hypotheses, as described in Feldman, Compton, and Symthe (1989). Thus, the views we present should be seen not as a *picture* of what bias is but a *device for the attainment or formulation of greater knowledge* about bias (cf. Kaplan, 1963). The critical question is not "How do we know our hypotheses about bias are correct?" (every hypothesis is an incorrect simplification) but "How can we expose our hypotheses to maximum criticism, in order to evaluate and refine them as thoroughly as possible?" (cf. Weimer, 1979).

From a constructivist point of view, there are only two currently available criteria for evaluating the adequacy of a definition of bias: 1) its coherence within a system of related ideas, and 2) its predictive utility with respect to our own experience. By coherence is meant "the regularity that binds the ideas together" (Bateson, p. 207). For instance, the fields which have contributed approaches to knowledge acquisition contain their own implicit and internally coherent beliefs as to what good data is, what bias is, and what methods can be used to achieve better data. Knowledge engineers can examine whether the definition of bias which they are considering would be consistent with these ideas. A definition of bias can also be evaluated according to*JB-rev** "what is the most useful way of construing the problem from the point of future prediction and control" (Adams-Webber, 1981, p. 3). For example, if a project's aim was the validation of the expert's estimates against logic, a view of bias that defined bias as a violation of logical rules would be most useful.

*JB-rev*Bias need not be expressed verbally, but will also be evident in other aspects of an expert's behavior. In this paper we will follow the trend of most current literature, and speak as if bias were mainly a matter of propositional content.*JB-rev**

1.2. TWO VIEWS OF BIAS AND THEIR PLACES WITHIN KNOWLEDGE ACQUISITION

Artificial intelligence, decision analysis, and social science literature *JB-rev*discuss two somewhat*JB-rev** different views of bias. These two views differ in terms of the reference point from which bias is judged. The reference point for one definition of bias is *JB-rev* what the "actually" knows or does; the reference point for the other definition is some set of "objective" criteria for rationality*JB-rev**. In the descriptions below we have exaggerated the distinctions between the definitions for the purpose of illustration.

*JB-rev*Those who are looking from the first reference point have concerned themselves largely with kinds of biases we will term *motivational*. For example, the conscious or unconscious accommodations an expert makes in order to please or annoy the knowledge engineer constitute forms of motivational bias. The concern with motivational bias comes is prevalent in the soft sciences, particularly psychology and cultural anthropology. These disciplines generally consider descriptions of the expert's performance to be the *gold standard*

[*] In this paper, we have chosen to handle the masculine gender bias of indefinite pronouns by designating the knowledge engineer as female and the expert as male.

that they wish to emulate during knowledge-based system development (Henrion and Cooley, 1987).

The second reference point defines bias as occurring when judgments obtained from an expert deviate from *objective* rules or standards considered to apply to the situation. We group these deviations under the heading of *cognitive bias*. To illustrate, if an expert tends to remember failures with disproportional frequency to successes, we might infer the possibility of a cognitive bias at work. The view of bias as a departure from principles of rationality has a long tradition in the fields of decision analysis and statistics (Mumpower, Phillips, Renn, and Uppuluri, 1987). Decision analysis practitioners do not deny the importance of research in describing how people actually go about making decisions, but insist that research results be applied not merely to reproduce what people do, but to discover ways to counteract systematic sources of bias in intuitive judgment (Beyth-Marom & Dekel, 1985; Kahneman, Slovic, & Tversky, 1982)*JB-rev**.

These views of bias are often implicit in knowledge acquisition applications or, at least, in particular phases of the application. We have observed that these views coincide with particular project goals, types of elicitation techniques, selections of expert response modes, and validation strategies.

The motivational view of bias, for instance, is more prevalent in knowledge acquisition projects whose goal is the psychological modeling of the source expert. In addition, proponents of this view of bias are likely to *JB-rev*favor *JB-rev**the use of nondirective elicitation techniques in the desire to avoid *leading* the expert. The verbal protocol technique, in which the expert thinks aloud but is not questioned by the knowledge engineer, is one such nondirective method. Also proponents of this view tend to avoid complex numeric response modes and allow the expert to respond spontaneously in his own, often nonnumeric, terms. Additionally, the motivational view of bias can be found in validations that focus on how well the knowledge base or the the knowledge-based system corresponds to the source expert's knowledge.

An emphasis on the cognitive view of bias is likely to coincide with projects whose goal is to improve on the expert's judgment. In projects where the concern with cognitive bias dominates knowledge acquisition, *JB-rev*the expert may be asked to respond using forms that allow greater precision of expression. For instance, it is easier to determine that experts are underestimating uncertainty if they are using numeric probabilities than if they simply state that some event is very likely*JB-rev**. Thus, the formal checks on the expert's knowledge can be done from the very first phases of knowledge acquisition. The focus on validation from the cognitive bias perspective can also occur in later validation phases. For example, Gaines (1989, p. 52) describes one validation criteria of a knowledge-based system as checking *to what extent the knowledge in the system corresponds to actuality* *JB-rev**JB-rev**.

1.3. *JB-REV*SOURCES*JB-REV** OF BIAS

*JB-rev*Different reasons are advanced for the presence of each kind of bias. Motivational bias is assumed to be driven by the self-interest of the expert whereas cognitive bias is generally seen as driven by constraints on human information processing.

1.3.1. Motivational bias

Concerns about the social desirability of a response could influence the reports of experts. For example, if the knowledge engineer asks experts a leading question such as "Did you use subgoal *x* in solving the problem?", they may feel inclined to answer "yes," even if they did not solve the problem in this way. Nonverbal cues from the knowledge engineer can influence the expert's reporting. For instance, if the knowledge engineer leans forward, displaying intense interest in something that the expert is saying, the expert may unconsciously respond by giving undue emphasis to statements on this topic, and ignoring relevant data that evoked a lukewarm response. The reactions of other experts present in the room can have similar effects on the what the expert says. Furthermore, experts' thinking can be affected by their imagination of how those not physically present, such as clients or supervisors, might view their responses. For example, in a study examining whether magnetic fusion projects would meet their technical milestones, some of the experts said that they considered their organization's judgments in giving their estimates (Meyer, Booker, Cullingford, and Peaslee, 1982).

There are several reasons why expert responses might be influenced by others. First, most people have a desire to be accepted and to receive approval (Zimbardo, 1983). Thus, in elicitation situations, the expert is likely to be responsive to what he believes the knowledge engineer, his managers, or other experts wish to hear. Secondly, although people are generally unaware of their basis for making decisions and solving problems (Hogarth, 1980, p. ix), they dislike admitting ignorance (Denning, 1986, p. 345). Experts are likely to fabricate or to acquiesce to suggestions of an acceptable answer or to means by which they might have solved the problem.

Additional errors may be introduced through misinterpretations on the part of the knowledge engineer or expert. However, since these are more problems of communication and representation mismatch than of bias, we will say no more about them here.*JB-rev**

1.3.2. Cognitive bias

*JB-rev*The mind is limited in how much information that it can process and in how much it can remember (Hogarth, 1980, p. 9). In order to reduce the cognitive burden, people tend to take short cuts when solving a complex problem. Thus they start with a first impression and integrate the information in a sequential manner, making only a few minor adjustments. Later, if additional information is received, they probably will not adjust their initial impression to give a more accurate judgment. In other words, if an individual who has already reached an initial solution is given contradictory data, he will probably not take this data sufficiently into account when generating a final answer. In particular, this sequential means of integrating information handicaps us in making predictions where large or sudden changes are involved. This limiting effect is called *anchoring* or *anchoring bias*.

The human mind has limited memory capacity for information processing. As Miller (1956) has noted, most individuals can not discriminate between more than seven things at a time. This limitation in information processing causes people to be inconsistent in working through the problem. For instance, people commonly forget an assumption made earlier and contradict it, thus causing *inconsistency bias*. (??)

Then too some data is more easily recalled than others. For instance, data involving catastrophic, familiar, concrete, or recent events may be easier to recall (Cleaves, 1986; Gold, 1987). This effect, termed *availability bias*, can lead to the overestimation of the frequency of some events.

Discuss briefly other biases as well, e.g.,
Failure to account for regression to the mean

Basing decisions on “representativeness”

Ignoring the differential effects of “framing”

Planning biases. See also list in Von Winterfeldt and Edwards book. Should some discussion of these be included in the program?
*JB-rev**

1.4. IMPORTANCE OF BIAS TO KNOWLEDGE ACQUISITION

The problem of bias is guaranteed importance to knowledge acquisition because of the definition of bias. In this paper, we have argued that bias is data judged to be undesirable in terms of the individual's values, whether what he values is data matching the expert's thoughts or *objective* standards. Thus, the inclusion of biased data in the knowledge base is to degrade the perceived quality of this knowledge. For example, if motivational bias occurs in the form of the knowledge engineer misinterpreting and then misrepresenting the expert's knowledge, the elicited knowledge will not match that of the expert source at the time of the elicitation. If the expert's knowledge is being used as the reference point for validating this information, it would be judged invalid. The same problem occurs with cognitive bias. For example, a plant pathologist gave a higher expected value to fungicide treatment than the sensitivity analysis performed on the knowledge base yielded (Henrion and Cooley, 1987, p. 475). The analysis showed that the fungicide treatment had a lower value than the expert judged because the infection is difficult to cure and the fungicide not highly effective. If this cognitive bias was not checked, the implemented knowledge-base system could have provided advice which was not mathematically optimal.

In addition, bias needs to be treated because of its negative impact on other's perception of expert knowledge and knowledge-based systems. Several articles have recently questioned the accuracy of expert clinical judgments in psychology and psychiatry (Faust and Ziskin, 1988; Dawes, Faust, and Meehl, 1989). The credibility of expert knowledge is open to this same sort of questioning if no attempts have been made to examine and, if necessary, counter bias.

Section 2 of this paper describes a six-step program for handling bias. The first part of the program leads the reader through the process of selecting a combination of steps for his or her application. Guidance is given, then, on how to follow each step using examples from both motivational and cognitive bias. A summary of the work is given in Section 3.

2. Program for handling bias

Approaches to handling biases are rare (Cleaves, 1986) and in their early stages. They are perhaps as much art as they are science. The program proposed here is no exception. One reason that there are not more programs for handling bias is that it is a difficult topic to study.

Studying, much less trying to counter bias, is complicated by not having a readily available baseline from which to determine the direction and magnitude of the bias. For the questions that expert knowledge is elicited, there are frequently no known single *right* answers or empirical data. Thus, the expert's data cannot be simply compared to an answer looked up in a reference or to the result of modeling data. Measurement of motivational bias is further complicated by the absence of *objective* standards for comparison. At least with cognitive bias there are traits of the expert's answers that can be compared to standards. For example, do the expert's exclusive but dependent probabilities sum to one as they should? With motivational bias, the baseline of the expert's knowledge, has its description not been altered, is difficult to determine, especially because expertise is not static. As Rosenfield (1988, p. 194) argues in his book on memory, higher mental functions are not fixed procedures but subject to constant reconstruction. While we recognize that both biases are difficult to detect, we believe that for progress to occur, programs like this one must be proposed and applied to knowledge acquisition.

Our approach differs from the one presented by Cleaves (1986). First, Cleaves' proposal focuses on cognitive bias; second, Cleaves tries to anticipate biases by the judgment processes in which they are likely to occur, namely hypothesis and solution generation, decision rule articulation, uncertainty assessment, and hypothesis evaluation. While we agree that biases occur during these processes, we try to anticipate the biases by the elicitation components that are likely to exhibit them. We assumed that many knowledge engineers were lay persons in the areas of human cognition and that this approach might be easier for them, at least as a starting point.

Another major difference in our program is its real-time emphasis. Given the evolving nature of expertise, bias is best detected when it is being elicited. This program stresses monitoring and adjusting for bias, particularly motivational bias, in real time rather than mathematically compensating for it afterwards. In particular, it is much more difficult to determine that motivational bias has occurred after the elicitation because the baseline--the expert's knowledge--is likely to have changed. For this reason, we consider each elicited datum to be a snapshot that can be compared to a snapshot of the expert's state of knowledge at the time of the elicitation.

Our proposed program consists of these general steps:

Step 1. Anticipate the potential sources of bias in the elicitation and redesign the elicitation, if necessary.

Selected biases are listed according to the situations in which they are likely to occur (See section 2.2.1.2. *Index of Selected Biases* and section 2.2.1.3. *Definitions of Selected Biases*). As a general rule, this step should always be performed. However, step 1 becomes more important if the knowledge engineer is in danger of introducing bias. Reading about the biases makes the knowledge engineer aware of her own potential for influencing the expert, misinterpreting the expert's words, or misrepresenting the knowledge, and hopefully lessens tendencies.

Step 2. Redesign the planned elicitation to make it less prone to the anticipated biases.

If the exercise of anticipating the biases leads the knowledge engineer to wish to redesign the elicitation at this point, she may use the *Definitions*. The *Definitions* include information on why the biases occur and thus can be used to design an elicitation which is less prone to these biases. Additionally, some of the *Suggestions for Countering Selected Biases* (step 5) could be implemented in planning the elicitation. (Another strategy for reducing bias in the elicitation methods is to use many different methods in the hopes that a more representative picture will emerge.)

Step 3. Make the expert aware of the potential for particular biases and familiarize him with the elicitation procedures.

The expert needs to be informed about the biases he is likely to exhibit given the elicitation situation. For example, the expert could be told about the tendency to underestimate uncertainty in making predictions. The expert then has a chance of giving answers that more fully encompass the actual amount of uncertainty. The expert should be given a review of the elicitation procedures, as described in section 2.2.3.1. If the experts are confused about how and when they are to respond, the data gathered as well as the expert's cooperativeness will be negatively affected.

Step 4. Monitor the elicitation for the occurrence of bias.

For each of the selected biases, there are signs that they may be occurring. (See *Signs* in section 2.2.3.1) For instance, if experts in a group situation are deferring to another individual, this could signal that their judgments are being influenced by that individual.

Step 5. Adjust, in real time, to counter the occurrence of bias.

Some techniques are given for modifying the elicitation process to prevent particular biases from intruding. (See section 2.2.5.1. *Suggestions for Countering Selected Biases*). Often prevention of bias involves controlling those factors which contribute to the bias, such as fatigue leading to increased inconsistencies. Other times, it requires the encouragement of an opposite bias to counter the occurrence of the first bias. For example, the tendency to be unconsciously influenced by others in the group can be countered by encouraging the expert to anchor to his first impression.

Step 6. Analyze the data for the occurrence of particular biases.

Examples are given of biases that we have tested for. (See section 2.2.6.) If step 6 is the only step of the program being followed, the analysis will necessarily be simpler than if steps 4 and 5 were also followed. If these other steps were followed, they would provide the additional data needed for performing more complex analyses. In general, adequately testing for one of the motivational biases requires this more complex testing. Occurrence of a cognitive bias, such as the underestimation of uncertainty, can often be determined by simple mathematical tests.

2.1. DETERMINING WHICH STEPS TO APPLY

The steps of the program above can be applied in sequence or singly, depending on the needs of the project. For example, if information on bias were not important to the project, none of the steps or only step 6, analyzing for bias, would be necessary. If on the other hand, the knowledge engineer wished to follow some but not all of the steps, she could do so. We suggest that step 6 always be done regardless of the other steps because it provides a general check on the expert data.

To pick steps for use in a project, consider the following: (1) your reason for addressing the problem of bias; (2) which view of bias, motivational or cognitive, will be employed; and (3) which particular sources of bias are of special interest.

2.1.1. The reason for focusing on bias

Your reason for focusing on bias can provide a criteria for determining which steps of the program to implement. For example, if the project personnel's interest in bias stems from a desire to avoid having reviewers criticize aspects of the project, their selection of steps would probably differ from those whose goal is to analyze the knowledge base for the presence of bias. It is also possible that, on some projects, there would be no reason for delving into the problem of bias. If this is the case, we would suggest that you record your reasons for not investigating bias and proceed as planned because examining bias demands extra time and resources,.

We have identified several reasons for focusing on bias:

to become aware of bias.
to prevent or inhibit the occurrence of bias.
to avoid criticism of the quality of the knowledge base.
to analyze the data for the presence of bias.

The above-mentioned aims can be accomplished by different combinations of the six steps. The flow chart below in Fig. 1 summarizes which steps achieve these aims.

2.1.1.1. To become aware of bias. Wishing to learn about bias is similar to wanting to discover and articulate one's assumptions. In both cases, the awareness allows the individual to consciously examine the phenomena and chose his beliefs or actions as opposed to blindly exhibiting them. To become aware of bias, use step 1, *anticipate bias,* and 3, *make the expert aware of the potential for introducing particular biases*. If the knowledge engineer only wishes to learn more about biases for a particular application, only step 1 is necessary. If, however, the knowledge engineer also wants the experts to be aware of bias, both steps 1 and 3 will be needed. This is because the knowledge engineer is the one who informs the experts about bias and the elicitation procedures. We consider familiarizing the expert with the elicitation procedures part of the process of informing him about bias. First, the elicitation methods form a backdrop against which bias can occur. Second, acquainting the expert with the elicitation procedure may prevent some of the expert's confusion, a contributor to the occurrence of many biases.

Step 4, *monitor the elicitation for the occurrence of bias*, can also be used to help sensitize the knowledge engineer to bias. Occasionally, the experts can assist the knowledge engineer in looking for the signs of selected biases.

2.1.1.2. To prevent or inhibit the occurrence of bias. If the reason for focusing on bias is to prevent or inhibit its occurrence, we recommend using at least steps 1, 2, 3, 5 and 6. Step 1 enables the knowledge engineer to anticipate the likely biases; step 2 allows her to use this information to redesign the elicitation in making it less susceptible to these biases. Step 3, *make the expert aware*, can also be implemented to reduce the likelihood of the expert introducing bias. Step 5, *adjust in real time to counter the occurrence of bias*, should be executed very carefully because the countering techniques could add bias or complicate later analyses. Step 6, *analyze the data for the occurrence of particular biases*, is performed as a check. Without step 6, there is only the knowledge engineer's claim that steps 1, 2, 3, and 5 have prevented bias. In our experience, conducting step 6 has usually allowed project personnel to state that no evidence was found of particular biases.

Although not necessary, step 4, *monitor the elicitation*, can also be performed as part of the goal of preventing the occurrence of bias.

2.1.1.3. To avoid criticism of the quality of the data. Sometimes the reason for focusing on bias comes from forces outside the project, such as other's concerns with the quality of the data, rather than from within the knowledge engineer. We have noticed that those who have had the unpleasant experience of having their project's data criticized by outside reviewers are highly motivated to avoid this situation in the future. This motivation is different from the other three mentioned in this section in that those who are motivated by this desire will probably not be interested in performing more than the minimum of steps, in this case, step 6. However, to analyze the data for bias assumes some knowledge of bias, so step 1, *anticipate the biases*, might also be necessary. In addition, step 4, *monitor the elicitation*, can provide data for the analysis.

In our experience, the reviewers have most commonly questioned data regarding the presence of these three sources of bias: social pressure from the knowledge engineer, social pressure from group think, and wishful thinking by the expert.

2.1.1.4. To analyze the data for the presence of bias. Analyzing the data for the presence of bias may be done for a variety of reasons. The knowledge engineer may be interested in analyzing for the presence of bias as a means to an end, such as to study the causes and effects of bias, or as an end in itself. If the knowledge engineer wishes to study bias, the techniques and situations that affect its appearance, any of the steps (1, 2, 3, 4, and 5) could be used in combination with step 6. For example, if the knowledge engineer was investigating the effect of using a particular step or elicitation technique, she could use it in one study and not in the control study, and then analyse and compare the results. One advantage of studying bias in this way would be the fresh look that analysts could take at bias. For instance, the analysts could determine if there was a difference in applying the step and whether they judged that difference as positive or negative. In other words, the researchers would not automatically assume that bias always leads to poorer results but would evaluate for themselves the peculiar properties of bias or its absence. (For this idea, we are indebted to Maurice Sharp, University of Calgary, Canada).

If analyzing the data for bias is the aim, such as for validating a knowledge base, steps 1 and 6 can be used. Step 1 allows the knowledge engineer to learn about likely biases and step 6 enables her to check for their presence. Generally, step 6 is more effective in investigating sources of cognitive bias. For example, the analysts could check for underestimation of uncertainty, as described in section 2.3.5.1. If, however, the knowledge engineer wants to analyze the data for how well it matches the expert's thinking (as would be the case with the motivational definition), we suggest adding step 4. Step 4, *monitor the elicitation*, enables the knowledge engineer to watch for some of the motivational biases that interfere with obtaining the expert's thoughts.

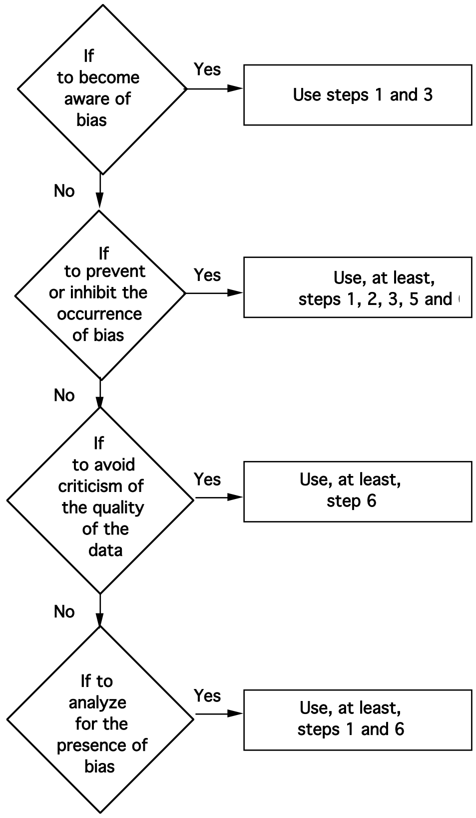

FIG. 1. Steps suggested according to the knowledge engineer's reason for focusing on bias.

2.1.2. The selection of the view of bias, motivational or cognitive

The selection of the view of bias, motivational or cognitive can also identify which set of steps will be most effective.

While the view of bias, motivational or cognitive, is often implicit in a knowledge acquisition project, we recommend that the reader consciously select which view or views of bias to apply. By consciously selecting the view of bias, the reader can consider which view will best serve the application in terms of coherence and usefulness. (The application could be a project, phase of knowledge acquisition, or validation phase.) For example, if the goal of the project is learning the expert's problem solving in order to emulate it, the motivational definition of bias would be more appropriate. If the project involves estimating the likelihood of future events, the cognitive definition would be a natural choice. People are inaccurate in making predictions and the cognitive view of bias would help to combat this weakness.

We suggest that the reader select and use only one view of bias at a time to avoid being contradictory. For example, use of the cognitive definition would propose that a mathematically incorrect judgment be modified. This act would cause a misrepresentation of the expert's data, a bias, according to the motivational definition of bias. Thus, we recommend that the reader formulate rules for when each view of bias will be applied, if more than one view is to be used. To illustrate, the definition of motivation bias could be used in acquiring knowledge for building the conceptual knowledge base and the cognitive definition used thereafter.

2.1.2.1. Motivational definition. As a general rule, if the motivational definition of bias is selected, steps 1, 2, and 3 will be most helpful. In step 1, the knowledge engineer is given information on what biases to expect. In step 2, she referred to the *Definitions* (section 2.2.1.3.) to receive ideas on how to minimize bias in planning or redesigning the elicitation. For example, if the index led the knowledge engineer to expect group think, the definition of group think would provide the knowledge engineer with ideas on how to design an elicitation that was less prone to it. Group think is more likely when behavioral aggregation is used so the knowledge engineer could plan to aggregate the responses mathematically. Use of step 3 *make the experts aware of the potential for bias and familiarize them with the elicitation procedures* would inform the experts about bias, and hopefully make them less prone to it. In this book, we have focused on presenting methods of elicitation and analysis which we believe minimize influencing the experts and force fitting their data. Thus, just using the methods suggested in this book, regardless of any program for handling bias, should provide some protection from motivational bias. In addition, step 4 *monitor the elicitation* is often helpful in detecting motivational biases such as group think. The *Signs of Selected Biases* (section 2.2.3.1.) tell the knowledge engineer which behaviors might be indicators of particular biases.

2.1.2.2. Cognitive definition. If the cognitive definition of bias is selected, step 6 *analysis* is particularly effective. Analysis is generally more definitive with cognitive than motivational sources of bias because cognitive biases can be determined mathematically or statistically. Cognitive bias can usually be measured because it is defined as a violation of logical or statistical standards. Thus, three of the four cognitive biases mentioned in this paper should be included in the analyses. For example, *underestimation of uncertainty* can be analyzed using the experts' ranges on their estimates (as described in 2.2.6.1.). Inconsistency is analyzable through logic checks. Anchoring can be investigated by comparing experts' first impressions to their later ones, given that the problem environment has changed such as through the receipt of new information. Analysis of cognitive bias presumes that the knowledge engineer knows which biases might pose problems. Thus, step 1 is useful in dealing with cognitive biases. In addition, step 3, *make the expert aware of the potential for bias and familiarize them with the elicitation procedures,* is often used with cognitive biases to lessen the expert's tendencies to introduce these.

The flow chart (Fig. 2) below summarizes the steps most suited to investigating cognitive or motivational bias.

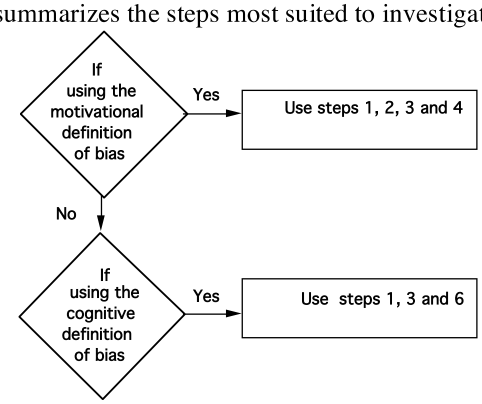

FIG. 2. Steps suggested for use with the two definitions of bias.

2.1.3. Interest in particular sources of bias

Frequently, project personnel are more interested in some causes of bias than others. This interest in particular sources of bias can be used to select specific steps because some steps are more effective with certain biases. The knowledge engineer will have to identify which biases she is most interested in and then select those steps that address these biases. We describe why we picked particular steps for three of the selected biases below and simply list the steps for the rest (Fig. 3).

2.1.3.1. Social pressure from the knowledge engineer. For instance, steps 1, 2, and 3 would be most helpful in dealing with social pressure from the knowledge engineer. In doing step 1, the knowledge engineer would become aware of her tendency to introduce this bias and perhaps be able to curb it. As a result of step 2, she might decide to select elicitation methods that were nondirective, like the verbal protocol, to minimize her tendency to influence the expert. In step 3, the expert would be made aware of this bias so he would be able to guard against it.

2.1.3.2. Social pressure from group think. For dealing with social pressure from group think, all the steps may be necessary. In step 1, the knowledge engineer is able to assess whether her application is susceptible to group think. Assuming that it is, she can redesign the elicitation using information from step 2. Step 3 allows the knowledge engineer to alert the experts to the potential occurrence of this bias, and step 4 allows both parties to monitor the sessions for its occurrence. In step 5, the knowledge engineer can apply techniques for countering this bias. In step 6, the data may be analyzed for the presence of group think.

2.1.3.3. Wishful thinking. Step 1 informs the knowledge engineer that wishful thinking can occur in many elicitation methods. Step 2, allows her to redesign the elicitation to select those experts who were less likely to exhibit wishful thinking, those who had less at stake in the judgments. In one study whose aim was to obtain probabilities of whether project deadlines will be met on time, we focused on interviewing those physically working on the project rather than their managers (Meyer, Booker, Cullingford, and Peaslee, 1982). Additionally, we required that the experts disaggregate the question into its parts, give estimates for each part, and explain their reasoning for each estimates. This practice was to make it more difficult for the interviewees to give highly optimistic estimates.

Source of Bias	Steps
Social pressure from KE	1, 2, and 3
Social pressure from group think	1, 2, 3, 4, 5, and 6
Social pressure, impression management	1, 2, 5, and 6
Misinterpretation	1 and 2
Wishful thinking	1 and 2
Inconsistency	1, 2, 3, 4, 5, and 6
Anchoring	1, 2, 3, 4, 5, and 6
Availability	1, 2, 3, 4, and 5
Underestimation of uncertainty	1, 2, 3, 5, and 6

FIG. 3 List of steps necessary to addressing selected biases.

2.2. HOW TO EXECUTE SPECIFIC STEPS

2.2.1. Step 1--Anticipate potential sources of bias in the elicitation

Anticipation of potential sources of bias is the first step in the proposed program. This process is illustrated using nine frequently encountered biases. These were selected because they represent a range of bias within the two definitions. Certainly these are not the only sources of bias. However, they are ones that have frequently been encountered and for this reason are used as a starting point .

To anticipate bias, the knowledge engineer determines the parts of her planned elicitation using the *Components of Elicitation* below as an aide. She then searches the *Index of Selected Biases* (section 2.2.1.2.) for the biases to which they are prone. For instance, if the knowledge engineer planned to use the interactive group method, she would see that it was prone to the following biases: social pressure from group think, wishful thinking, and inconsistency. She then turns to the section after the table (*Definitions of Selected Biases*) to look up the definitions and causes of the selected biases. Through the process of looking up these biases in the *Index* and *Definitions*, the knowledge engineers will become aware of the bias's existence and of their own tendencies to introduce or acerbate them.

2.2.1.1. Components of elicitation

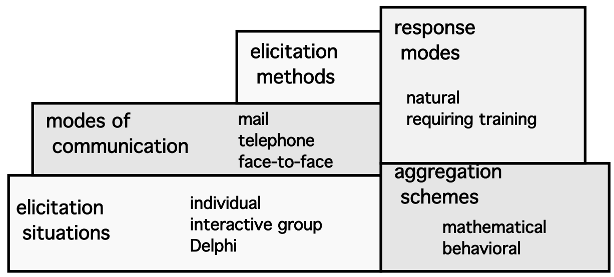

FIG. 4. Building blocks of elicitation.

1. The elicitation situation is where the knowledge engineer communicates with the expert. There are three distinct situations which can overlap each other in practice.

The individual meeting is the situation where the knowledge engineer meets with an expert face to face. The individual meeting could be focused, structured, or unstructured in its questioning and include problem discussion, description, or analysis, as described by Boose and Gaines (1988, p. 38-39). Most of the elicitation methods, even those for eliciting detailed problem-solving data (e.g., protocol analysis, the verbal probe or the ethnographic method), can be used in this situation.

The interactive group meeting is defined as where the knowledge engineer meets with the expert in the presence of other experts. Usually the knowledge engineer moderates these sessions and leads the questioning. This elicitation situation can range from traditional meetings, to negotiations, to decision-analysis sessions, and to interactions of users on terminals in a computer network (McGraw and Harbison-Briggs, 1989, p. 263). The interactions of the group members can be structured to differing degrees. An example of an unstructured interactive group would be the traditional business meeting; a highly structured interactive situation would be one in which the experts were requested to provide their responses to each other in specific forms at particular times. For instance, experts in army export control were required to discuss the problem and then give their data in the form of estimates with a few sentences explaining their reasoning (Meyer and Johnson, 1985). Elicitation techniques that provide broad coverage (e.g., the verbal probe), as opposed to in-depth probing, are best used in the interactive group situation.

The Delphi situation occurs when the knowledge engineer keeps the experts physically separate but acts as a channel for distributing each expert's information to all of the experts. (This technique was developed by Rand Corporation to avoid having members of the group influence each other's responses.) In strict Delphi, the knowledge engineer obtains the experts' responses through the mail, makes these responses anonymous, redistributes these anonymous responses to the individual experts, and allows the individual experts to revise their previous responses. This process can be repeated as many times as desired. In general, the Delphi has not been used to elicit in-depth problem-solving data because this type of data is best elicited when the knowledge engineer and expert are face to face. Instead, structured survey types of questions are asked in the Delphi situation. The Delphi has been used for classifying problems and for group negotiations (Boose and Gaines, 1988, p. 40).

2. The modes of communication are the means--face to face, telephone, or mail (electronic or postal)--by which the knowledge engineer and expert communicate during the above-mentioned elicitation situations. Several modes can be combined in one elicitation process. For example, the knowledge engineer can send the expert the problem, answer the expert's questions about it over the telephone, and then elicit the expert's problem solving in person. In addition, modes can sometimes be substituted for one another. For example, in the Delphi situation, the information could, if it were simple and brief, be communicated by telephone rather than by mail.

3. The Elicitation techniques is the means by which the expert is led to verbalize aspects of his knowledge, such as his definitions for terms, and the processes and procedures by which he solves problems. (The expert's verbalizations of his answers are covered under a separate component--the response mode.) One elicitation technique is interviewing, which itself is composed of ways of asking questions, such as the verbal probe and the ethnographic technique (Meyer, Mniszewski, and Peaslee, 1989). Another elicitation technique is the verbal protocol in which the knowledge engineer observes the expert and listens to him as he solves problems. Repertory grids are another elicitation technique whereby the key elements of the expert's thinking are elicited and then related to the distinctions that the expert makes regarding these elements. The techniques vary according to: the part of the knowledge acquisition cycle in which they are used (e.g., to extract the structure of the domain or more detailed production rules); whether they are applied while the expert is solving the problem or afterwards; whether they include a whole sequence of techniques to be applied at different points in the acquisition, such as KADS (Wielinga, Akkermans, Schrieber, and Balder, 1989) or whether they are only one technique; the degree to which they are automated (e.g., the manual versus automated elicitation of repertory grids), and the amount of structuring (e.g., spontaneous questioning versus interviewing by topic area). The elicitation techniques are not listed in the index below because some, like general interviewing, can be performed in such varying ways that they would be subject to different biases. For a more comprehensive description of elicitation techniques and of the larger set of methods used for knowledge acquisition, see Shaw and Woodward (1989) and Dhaliwal and Benbaset (1989).

4. The response mode is the form in which the expert is asked to give his answers or solutions to problems. Frequently in knowledge acquisition the expert is allowed to respond naturally rather than encode his thoughts into a form selected by the knowledge engineer. Other times, a numeric response mode, such as probabilities, continuous linear scales, Saaty's paired comparisons, ratings, uncertainty measures, and confidence ranges, is requested.

5. The aggregation scheme is the means by which multiple and differing expert's responses are made to form one response, such as one line of reasoning or a numerical solution. The two major techniques are mathematical aggregation and behavioral aggregation.

> **Mathematical aggregation** is the use of mathematical means to combine multiple expert's quantitative answers into one answer. Some mathematical methods weight the expert's answers equally (e.g. the mean), others weight the experts' data differently in attempts to give higher weights to the *more expert* data. Mathematical aggregation will not be addressed below because it is outside the scope of this paper.
>
> **Behavioral aggregation** relies on the experts reaching a consensus (Seaver, 1976). The aggregation occurs during, rather than after, the elicitation session. When behavioral aggregation is used in interactive group situations, the experts are informed that they must reach a consensus and often employ persuasion and compromise to do so. When behavioral aggregation is used with the Delphi technique, the experts give their responses and receive feedback until they reach one response.

2.2.1.2. Index of selected biases

Components	View of bias--Source
Elicitation situations	
Individual meeting	Motivational--Social pressure from KE
	Motivational--Wishful thinking
	Cognitive--Inconsistency
Delphi	Cognitive-Inconsistency
	Motivational--Wishful thinking
	Cognitive--Anchoring
Interactive group meeting	Motivational--Social pressure from group think
	Motivational--Wishful thinking
	Cognitive--Inconsistency
Modes of communication	
Face-to-face	Motivational--Social pressure from KE
	Motivational--Wishful thinking
	Cognitive--Underestimation of uncertainty
Telephone	Motivational--Social pressure from KE
	Motivational--Wishful thinking
	Cognitive--Availability
	Cognitive--Anchoring
	Cognitive--Underestimation of uncertainty

	Mail	Motivational--Social pressure, impression management Motivational--Misinterpretation by KE Motivational--Wishful thinking Cognitive-Inconsistency Cognitive--Availability Cognitive--Anchoring Cognitive--Underestimation of uncertainty
Response modes		
	Complex, numeric response modes, such as probability, distributions, especially if the expert is inexperienced in their use.	Motivational--Misinterpretation by expert Motivational--Wishful thinking Cognitive--Inconsistency Cognitive--Underestimation of uncertainty
Aggregation		
	Behavioral aggregation	Motivational--Social pressure, group think

2.2.1.3. Definitions of selected biases

Motivational Bias is the altering of the expert's responses through social pressure (from the knowledge engineer, group think, or impression management), and misinterpretation. In addition, the expert's motivation, such as in wishful thinking, can influence his responses. Wishful thinking bias differs from the other selected motivational biases in that the expert's expressions are not necessarily changed by the expert himself or by the knowledge engineer. However, we have included wishful thinking in the motivational biases because it originates from the expert's motivation.

Social pressure is the altering of the expert's descriptions of his thoughts arising from the desire to be accepted and to see himself in the most positive light possible. This altering can take place consciously or unconsciously. The social pressure can come from those physically present, such as the knowledge engineer or the other experts, or from the expert's own internal evaluation of others' reactions.

Social pressure from the knowledge engineer is most likely to occur in those methods where she is meeting in a face-to-face situation with the experts, such as in the individual and the interactive group meeting. In face-to-face situations, the knowledge engineer can intentionally or unintentionally influence the expert through her body language, facial expression, intonations, and choice of words. We expect this source of bias to be weaker in telephone conversations and weaker still in communications by mail. These last two communication modes do not allow some of the above-mentioned means of expression that the face-to-face mode does. In addition, social pressure bias is more pronounced when the knowledge engineer is asking leading questions. Thus, it is weaker when the knowledge engineer is using one of these elicitation techniques--the verbal protocol, verbal probe, or ethnographic phrasing (Meyer, Mniszewski, and Peaslee, 1989). The verbal protocol mentioned here requires the expert to think aloud as he solves a problem and thus avoids leading the expert because it does not involve questioning him; the verbal probe uses general, nonleading phrases; and the ethnographic techniques uses the expert's own words in formulating questions. We would also expect other knowledge acquisition techniques, such as KADS (Wielinga, Akkermans, Schrieber, and Balder, 1989) and SORTAL (Regoczei and Hirst, 1989), that are driven by the expert's terms and meaning to be less prone to social pressure from the knowledge engineer.

Group think is another source of social pressure. Social pressure from others in the group induces individuals to slant their responses or to silently acquiesce to what they believe will be acceptable to their group (Meyer 1986, p. 89). Zimbardo, a psychologist, explains that it is due to the basic needs of people to be loved, respected, and recognized that they can choose or be induced to behave in a manner which will bring them affirmation (1983). There is abundant sociological evidence of conformity within groups (Weissenberg, 1971). Generally, individuals in groups conform to a greater degree if they have a strong desire to remain a member, if they are satisfied with the group, if the group is cohesive, and if they are not a natural leader in the group. Furthermore, the individuals are generally unaware that they have modified their judgments to be in agreement with the group.*

One mechanism for this unconscious modification of opinion is explained by the theory of cognitive dissonance. Cognitive dissonance occurs when an individual finds a discrepancy between thoughts that he holds or between his beliefs and actions (Festinger, 1957). For example, if an individual holds an opinion that conflicts with that of the other group members and he has a high opinion of the other's intelligence, cognitive dissonance will result. Often the individual's means of resolving the discrepancy is by unconsciously changing his judgment and/or the reporting of that judgment to be in agreement with that of the group (Baron and Byrne, 1981). For example, Janis' study (1972) of fiascoes in American foreign policy illustrates how presidential advisers often silently acquiesce rather than critically examine what they believe to be the group judgment. This phenomena has also been called group think bias, the follow-the-leader or bandwagon effect.

Group think is only likely to be a concern in an interactive group situation. It is further likely to occur in situations where behavioral aggregation is used because this type of aggregation requires that pressures toward conformity be encouraged.

Impression management is another type of social pressure that can occur as the expert imagines the reactions of those not physically present. This effect can occur in any elicitation situation. However, it seems to be more noticeable when it is not covered by other effects, such as social pressure caused by the knowledge engineer. For this reason, its occurrence is most noted in mail surveys or Delphi situations. The individual may try to answer in such a way as to bring the most approbation, such as from the person who has written the questions. Then, too, he may try to respond in such a way as would be acceptable to his employer or to society in the abstract. For this reason, this source of social pressure has been termed impression management (Goffman, 1959). Payne (1951) has found evidence of individuals giving the responses that they perceived to be the most socially acceptable rather than those which accurately portrayed their thoughts or actions. For example, on surveys, people claim that their educations, salaries, and job titles are better than they are. Often there is a 10% difference between what is claimed for reasons of prestige and what objectively *is* (Meyer 1986, p. 90).

Misinterpretation is the altering of the expert's responses as a result of the methods of elicitation and documentation. While this effect is prevalent, it has not received much attention. Frequently misinterpretation occurs as a result of the response mode. If the expert can not adequately translate his judgment into the response mode, misinterpretation will result. We have noticed that experts seem to have more difficulty with the following response modes: probability distributions, ranks, and percentiles.

Misinterpretation is also more likely with elicitation and documentation methods that are written from the knowledge engineer's, rather than the expert's, viewpoint. For example, we have all had the frustrating experience of trying to force fit our views into the limited response options of a mail survey.

Wishful thinking occurs when an individual's hopes influence his judgment (Hogarth 1980). What the subject thinks should happen will influence what he thinks will happen. To illustrate, presidential election surveys show that people predict the winner to be the candidate that they expect to vote for (Armstrong 1981: 79). The above instance is one where the subjects stand to gain very little personally from their answer. The wishful thinking effect is stronger where the subjects are personally involved or would gain from their answers. Hence, this bias is also called *conflict of interest*. In general, people exhibit wishful thinking about what they can accomplish in a given amount of time: they overestimate their productivity (Hayes-Roth 1980).

---

* Note that we are making a distinction between unthinking conformity and and an expert changing his mind as a result of new information received. The former is group think, the later an illustration of knowledge in action.

Wishful thinking is not particular to any elicitation method. Instead it relates to selection of experts and the assignment of them to specific questions or problems. If they have a special interest in the answer, wishful thinking is likely to occur whether the individual interview, interactive group or Delphi elicitation method is used or whether the communication is face-to-face, by telephone, or mail. For this reason, we suggest that you select those experts who have the least to gain from the response that they give. Often, however, those most qualified will also be those with the most at stake. *JB*For this reason , Boose and Shaw (1989, p. 73) advise obtaining data from experts holding divergent views. *JB** One approach to obtaining diverse expertise is to select experts that represent different organizations (e.g., government, academia, and industry), and the various theoretical stances on the subject.

In general, the effects of wishful thinking will be most pronounced when the expert does not have to explain his reasoning. The expert's highly optimistic responses are checked by having him disaggregate the problem and explain his problem solving. For example, Hayes-Roth (1980) found that having people break down the tasks that they had earlier thought they could accomplish in a given time led to more realistic estimates.

Cognitive Bias occurs when data fails to follow mathematical and logical standards because of inconsistency, anchoring, availability, or underestimation of uncertainty.

Inconsistency is the inability to be consistent in one's solving of problems, especially through time. Of all of the biases mentioned here, this is the most common. Individuals often unintentionally change definitions, assumptions, or algorithms that they meant to hold constant throughout the problems. Inconsistency in an individual's judgment can stem from his remembering or forgetting information during the elicitation session. For example, the individual may remember some of the less spectacular pieces of information and consider these in making judgments later in the session, or the individual may forget that particular ratings were only to be given in extreme cases and begin to assign them more freely toward the end of the session.

As Dawes, Faust and Meehl (1989, p. 1671) have noted, such factors as fatigue, recent experience, or seemingly minor changes in the ordering of the information or in the conceptualization of the "...task can produce random fluctuations in judgment. Random fluctuation decreases judgmental reliability and hence accuracy." These inconsistencies may result in answers that do not make logical or Bayesian sense. For instance, a series of answers proposing that factor A was more critical than B, B more than C, and C more than A would not make logical sense. Similarly, if an expert gave the same probability of A for two situations: one of which involved an influential factor C and one which did not, his answers would not be coherent from a Bayesian viewpoint.

The natural tendency toward inconsistency is acerbated by several conditions such as memory problems, confusion, and fatigue. During elicitation sessions of more than 30 minutes, people often forget the instructions, definitions, or assumptions that they were requested to follow. For example, the experts may forget that a rating of nine meant a near certainty and assign it more easily than the definition specified. Thus, unstructured elicitations, which do not have periodic reviews of the problem information, are more likely to have high inconsistency. This inconsistency can be between experts' answers (e.g., the experts meant different things by the same numerical answer) or within an expert's answer (e.g., sometimes the expert gave a specific rating more easily than at other times). Also, situations where the expert's understanding through time cannot easily be monitored are more prone to inconsistency. These situations include the Delphi or mail survey.

Confusion can also lead to inconsistency. Thus, any of the more complicated response modes, such as probability distributions and percentiles, are more prone to this problem. This confusion is why training the expert in the use of these modes (step 3) is recommended. In addition, if the experts must mentally juggle more than five to seven things, such as in rating them, they are likely to become confused and inconsistent. It is for this reason that the Saaty paired-comparison mode is used even though it is more time consuming than some of the other response modes.

Anchoring is the failure to adjust sufficiently from one's first impression in solving a problem. We would rate it next to inconsistency in terms of frequency of occurrence. Sometimes this tendency is explained in terms of Bayes Theorem as the failure to adjust a judgment in light of new information as much as it would be in terms of Bayes mathematical formula (Meyer, 1986, p. 88). Spetzler and Stael von Holstein (1975) and Armstrong (1981) describe how people tend to anchor to their initial response, using it as the basis for later responses. Ascher (1978) has found this problem to exist in forecasting where panel members tend to anchor to past or present trends in their projection of future trends. Ascher determined that one of the major sources of inaccuracy in forecasting future possibilities, such as markets for utilities, was the extrapolation from old patterns that no longer represented the emerging or future patterns. Another example of anchoring occurs when a member of a group's last estimate is closer to his initial impression than it would be had he fully taken earlier group discussions into account .

Anchoring is most prevalent in situations where the expert is not likely to experience the opposite bias of being influenced by the knowledge engineer or the group, such as in the <u>Delphi</u> method. In addition, those modes, such as <u>mail</u> or <u>telephone</u> communications, where the expert's thoughts cannot be easily monitored by having the expert think aloud, are prone to this bias. Also, we have noticed that experts are more likely to stick with their anchor if they have either described it orally or in writing and fear losing face for changing their mind.

Availability bias arises from the differing ease with which events can be retrieved from long-term memory. Data involving catastrophic, familiar, concrete, or recent events tend to be easier to recall. *JB* Contributing to this bias is our human tendency to think that "we know more than we do and also that what we don't know must be unimportant" (Kahneman & Tversky, 1978).*JB** Availability bias affects people's ability to accurately estimate frequencies and recall other aspects of the event. For example, the incidence of severe accidents in reactors tends to be overestimated in part because of their catastrophic and newsworthy nature.

Availability bias is more common when the expert does not receive any information from others and, thus, does not have a chance of triggering other, less accessible, memory associations. For this reason, the <u>individual meeting</u> is the most prone to availability bias, and the interactive group, the least. With individual meetings, a series of different scenarios is often used to help the expert enlarge on the sample of things contributing to his final answer.

Availability bias is also more common with <u>telephone</u> and <u>mail</u> modes of communication because the expert is usually not given much background before being asked point blank for the answer. A structured hierarchical presentation of the information, such as from the general to the specific, can alleviate this weakness.

Underestimation of uncertainty occurs when people underestimate the amount of uncertainty in the answers that they give. For example, when people are asked to put a range around an answer such that they are 90% sure that the range encompasses the correct answer, their ranges only cover 30-60% of the total (Capen, 1975). Even when they are given quizzes and feedback on their performance, they cannot break the barrier of covering only 70% (Capen 1975, p. 846). A popular explanation for this effect is that we are uncomfortable with the amount of uncertainty in our lives, and thus, minimize it. In particular, we may avoid confronting the large uncertainties in our judgments. *JB* Also, in planning activities, we may underestimate uncertainty because of the chain-like nature of planning. We may ignore factors which could cause a delay because each factor seems unlikely by itself (Kahneman & Tversky, 1978). *JB**

Although underestimation of uncertainty is very widespread, Martz, Bryson, and Waller (1985, p. 72) have noted that it is more pronounced with <u>probability</u> and <u>chance estimates</u> than with some of the other response modes. Chance estimates, also called odds, are given as 1 chance in a total, such as 1 in 1000.

2.2.2. *Step 2--Redesign the planned elicitation to make it less prone to the anticipated biases*

The definition section mentioned above can be used to redesign the elicitation, if the knowledge engineer as a result of anticipating bias wishes to do so. For instance, suppose that the

knowledge engineer became concerned with the potential for wishful thinking bias as a result of reading the *Index* and *Definitions* in step 1. She could turn to the definition of wishful thinking and replan her strategy for sampling the experts. In the *Definition* section, the knowledge engineer receives the suggestion to select those experts who have the least to gain from their responses. An alternative is to sample experts from different organizations or schools of thought in attempts to obtain diverse views on the subject.
In addition, some of the suggestions given in step 5 for countering selected biases (section 2.2.5.1.) could be integrated in the designing of the elicitation. For example, one of the ways for countering wishful thinking is to ask the expert to explain his thinking in detail. The knowledge engineer could incorporate this requirement into her design for the elicitation (e.g., plan to have the problem disaggregated into its component parts and to have the expert explain his reasoning on each part).

2.2.3 Step 3--Make the expert aware of the potential for bias and familiarize him with the elicitation procedures

The experts need to be (1) informed about the biases, and (2) acquainted with the elicitation procedures. Regarding bias, they need to know the definitions and causes of its sources. Without this information, the experts will not be able to combat their own tendencies toward bias. The knowledge engineer can use the *Index* and *Definitions* provided for step 1 as a base for informing the experts about bias.

It should be noted that making the experts aware of the biases helps but does not completely alleviate the problem. For example, if the expert is aware of how his responses may be misinterpreted or influenced by the knowledge engineer, he can help monitor for these biases. However, in some cases the cause of the bias, such as the expert's underestimation of uncertainty, is too ingrained to be completely overcome. In other cases, the experts will not make the needed effort to counter the natural tendency toward bias. People typically believe that others, not themselves, will suffer from the biases described. With some biases, such as anchoring and underestimation of uncertainty, the experts can participate in tests designed to evoke the bias. Frequently, almanacs are used to construct test questions, such as: *How much rain fell in St. Paul, Minnesota in 1987?* While the experts will not know the answers to such questions, the knowledge engineer can look up the correct answer. The knowledge engineer can read the answers and allow the experts to score their own. Such a demonstration is often necessary to convince the expert that he too is prone to the bias.

The experts also need to be made aware of the elicitation procedures. If they are confused about how and when they are to respond, the data gathered as well as the expert's cooperativeness is negatively affected. One aspect of elicitation that is often confusing is the response mode, if the expert is not accustomed to using it. The use of unfamiliar response modes should be rehearsed by the experts during the training session. [If the response mode is probability distributions, Hogarth (1980, p. 149) provides eight keys to its use.]

2.2.3.1. Familiarize the experts with the elicitation procedures. Information on how to familiarize the expert with the elicitation procedures is given below by elicitation situation.

For an individual meeting

1. Give the expert some sample questions to familiarize him with the use of the response mode, if that mode is likely to be a difficult one for him (i.e., if he does not use it as a part of his normal problem solving).

2. Brief the expert on any biases that were identified as being likely to occur (step 1). Give the expert ideas on how he can strive to counter the tendency towards these biases. (The section Definitions of Selected Biases in section 2.2.1.3. provides examples of this type of information.)
3. Give the expert the set of questions and verbally go over any instructions. Ask the expert if he has any questions.

For an interactive group meeting

1. Review the purpose of the project, its schedule, and, in general the elicitation procedures for the benefit of the experts. Some descriptions of the elicitation procedures are as follows. You will meet together for this week to develop detailed statements or representations of the problems. On the last day, Friday, you will vote on what you think the answers should be. A more detailed overview of elicitation procedures is as follows. You will meet here three times: First, to become familiar with the project and the elicitation procedures; second, to present up-to-date technical information and refine the rough-drafted questions; and third, to give your expert judgment in private meetings with a knowledge engineer.
2. Give the experts sample questions to work so that they can practice using the response mode. If there are any techniques to properly using the response mode, they can be introduced and practiced here.
3. Brief the experts on those biases that were identified in step 1 as being likely to occur. This briefing should include an explanation of why the selected biases occur and of how the expert can reduce his tendency to introduce them. (See the Definitions section.) In addition, the briefing on bias should include exercises that are designed to evoke the selected biases. After the experts have completed the exercises, the moderator/knowledge engineer can read the answers and allow the experts to correct their own. These exercises can convince the expert that he, like everyone else, is prone to these biases.
4. Ask if there are any questions. Afterwards, state that the introduction is concluded and the elicitation sessions will now begin.

For a Delphi situation

- If the entire Delphi will be conducted by mail, the expert will not be introduced to the elicitation by the knowledge engineer/moderator in person. Instead the expert will receive the cover letter and set of questions.
- If part of the Delphi will be conducted by telephone, call the expert to assist him in understanding the set of questions or just to obtain his answers. Use the items listed for individual meetings above as a basis for introducing the expert to the elicitation process.

2.2.4. Step 4--Monitor the elicitation for the occurrence of bias

Prior to the elicitation sessions, the knowledge engineer looks up the signs that the biases may be occurring in *Signs of Selected Biases* below. For instance, if group think bias was anticipated, the data gatherer would look up this bias in the *Signs* section and read about indications of its presence. One sign of group think is that the experts appear to defer to another member of the group or to each other. The knowledge engineer or a trained observer then watches for this sign of group think during the elicitation. In general, monitoring biases, as described here, requires that the experts verbalize their thoughts and answers. Without this feedback, we have found the monitoring to be much more difficult.

2.2.4.1. Signs of selected biases

Group think, as it develops, may be indicated by several signs. Generally, no difference of opinion is voiced, and the experts appear to defer to another member of the group or to each other (Meyer, 1986, p. 95).

Wishful thinking may be difficult to detect because its signs are subtle. Watch for when an expert responds before he would have been expected to think through the question or for when an expert seems unduly interested in his management's access to his responses.

Inconsistency is indicated by a number of signs. The knowledge engineer can hear many of these, if the experts are verbalizing their thoughts and answers. In particular, she can detect when a response mode or rating is being applied more easily through time (Meyer, 1986, p. 94). Experts tend to apply the extremes of a rating scale more easily as they become fatigued. The knowledge engineer can also hear when the expert is contradicting an assumption that he made earlier. For example, a tank expert chose two very different routes through the mapped terrain because the second time he unconsciously assumed that his company was the main effort and *had to push hard*.

Inconsistency can also be monitored by the use of Bayesian-based scoring and ranking techniques. During the elicitation, the expert's judgments can be entered into a scoring and ranking program, such as that of Saaty's Analytical Hierarchical Process (1980), to obtain a rating of their consistency. Then, if this number is too high, indicating significant inconsistency, the knowledge engineer can ask the experts to redo their judgments.

Availability may be indicated if the expert does not mention more than one or two considerations in giving his answer. If the expert only considers a few things, these were probably the most easily remembered and the answer is likely to be skewed to reflect these few.

Anchoring can be suspected if the experts receive additional information from experts or other sources during the elicitation but never waiver from their first impression. For example, reactor code experts were asked to compare the performance of their computer codes to plots of experimentally generated data. Often they commented on their first impression. When they examined the plots more closely, they typically found places where the computer code did not capture the experimental phenomena. However, the experts usually simply adjusted upward or downward of their initial assessment rather than revising it completely (Meyer and Booker, 1987).

2.2.5. Step 5--Adjust, in real time, to counter the occurrence of bias

In this step, the knowledge engineer looks up the suggestions for preventing a particular bias in *Suggestions for Countering Selected Biases* below. These suggestions vary because we have used two approaches: (1) controlling those factors contributing to a particular bias, or (2) employing the opposite bias. The first approach involves controlling the factors that contribute to the bias. For instance, fatigue is a factor that leads to increased inconsistencies in expert judgment. The interviewer can stop the elicitation sessions or schedule breaks before the experts become fatigued as a means of controlling this contributor to inconsistency. The basis of the second approach, fighting bias with bias, comes from Payne (1951), the grandfather of survey design. Payne believed that all interviewing was biased and that one should therefore aim for equal but opposite biases. An example of this technique is to try to have experts anchor to their own judgments in attempts to counter a group-think situation. Having the experts write their judgments encourages them to form and keep their own opinions even when they hear the opinions of others.

This step is perhaps the most delicate one because if done carelessly it could confuse the results and any later analyses. The knowledge engineer needs to decide in advance on the timing of the adjustment because these adjustments change the conditions under which the data is gathered. If a condition leading to bias is corrected before the analyzable data has been gathered, there is no problem. If, however, the analyzable data has been gathered under two conditions, when bias was occurring and then when it was corrected, the data will be mixed. Unless the situations can be clearly separated, such as before and after correction of the bias the knowledge engineer will not be able to later test for the presence of this bias.

2.2.5.1. Suggestions for countering selected biases

Group think, as a source of social pressure, can be countered using techniques from both the approaches (1) and (2) above. (Also see Meyer, 1986, p. 95-96). Using the first approach, the knowledge engineer can stop the elicitation and warn the group members about group think. If there is an official or even a natural ex officio leader in the group, that individual can be asked to give his responses last, or privately, so as not to influence the other group members. In addition, if someone other than the knowledge engineer has been leading the group meeting, he can be encouraged to be nondirective during the meetings. An explanation of the group-think phenomena usually convinces them that better discussions and data will result from their avoiding *leading*.

The other approach is to try to counter the effects of group think with anchoring. One technique for fostering anchoring is to require the group members to write down their judgments and reasoning. In this way, they are more likely to anchor to their own judgments rather than silently acquiesce to someone else. If the experts are to discuss their judgments, each person can record and report his before the floor is opened to discussion. Once individuals have publicly announced their view, they are unlikely to spontaneously modify it . (They will still modify their view if someone raises a valid point that they had not previously considered.)

Wishful thinking can be countered by making it more difficult for the expert to indulge in it. If the expert must explain his answer in detail, it will become apparent whether there was any basis beyond his motivations for his response. *JB*If the expert is providing an estimate of his own performance, Kahneman and Tversky (1978) recommend focusing the discussion on contingencies that are outside the expert's control but may affect his achievements.*JB**

Inconsistency can be reduced by using several techniques. The first is to address the aspects of the elicitation that are contributing to the inconsistency. As mentioned earlier, fatigue is a contributor to inconsistency. If the knowledge engineer has noted that the experts are becoming more inconsistent with time, she can quickly end the meeting or schedule a break. In general, two hours is the maximum amount of time that experts participate in discussion or problem solving before becoming tired. (Experts often signal their fatigue either by briefer responses or by leaning way forward or back in their chairs.)

Another contributor to inconsistency is faulty memory. If at the beginning of every session the statement of the problem, definitions, assumptions, and response mode are reviewed, the experts will be more consistent in their judgments (Meyer, 1986, p. 96). They will be more consistent between and within themselves. In addition, if there is much time between this first review and when the experts' judgments are requested, the question can be worded to include some of the above information. For example, "What rating would you give to the importance of element *X* over *y* to the reaching of goal *z*?" If they are using a response mode, in this case a Saaty paired comparison, they will need to have the definitions of the scale available in front of them.

Another technique for reducing inconsistency is to have the group members monitor their own consistency (Meyer, 1986, p. 96). This technique was successfully used in a simple interactive group elicitation where the experts were able to watch the knowledge engineer's monitoring of inconsistency and then mimic it. (Meyer, Peaslee, and Booker, 1982). The experts were given copies of a matrix of the elements being judged, the criteria on which they were being judged, and their past judgments. When experts monitor their own consistency they may wish to change an earlier judgment to be in line with their current thinking. If their reasoning does not violate the logic of the model or the definitions, they can be allowed to make the change. Often in this process, the expert may discover that he had forgotten to include some pertinent information. After this addition, some of the judgments may need to be redone.

If Saaty's Analytic Hierarchy Process (1980) had been used and its results indicated high inconsistency, the experts could review and redo the affected judgments.

Availability bias can be countered by stimulating the expert's memory associations. In general, group discussion will cause the expert to think of more than just the first readily accessible information. In addition, free association can be introduced to single experts or those in groups. Free association is having the expert or experts quickly generate any and all elements that might have bearing on the question (Meyer, 1986, p. 94). Free association is similar to brainstorming or the Crawford Slip method (Boose and Gaines, 1988, p. 38). The experts are asked to refrain from being critical in order to generate the widest possible pool of ideas. (This list is later narrowed to those judged to be most pertinent to the problem.) A related technique is to hierarchically structure the presentation of problem information so that it flows from the general to the specific. In this way, the expert is able to consider the pertinent information before having to reach a solution. Again this strategy is to fire as many memory associations as possible so that the maximum number of relevant ones will enter into the expert's final judgment.

*JB*Three additional ideas for countering availability bias are: (1) to ask the expert to describe how other experts might disagree with his responses (2) to ask the the expert to temporarily forget recent events; and (3) to aggregate outcomes with small probabilities into a single larger class to reduce the perceived joint impact, if probability estimates are being elicited (Kahneman & Tversky, 1978).*JB**

Anchoring can be countered by some of the techniques mentioned for availability bias. In particular, giving the expert input from other experts as in a Delphi situation or an interactive group makes it more difficult for the expert to anchor to his first impression. Another technique is to ask the expert for extreme judgments before getting his likely ones (Cleaves, 1986, p. 9-10).

Underestimation of uncertainty may be reduced by asking the expert to further disaggregate the parts of the question and give estimates of the quantities of interest for each small part. In this way, the expert is less likely to overlook some thing and the knowledge engineer can check whether the details of the expert's thinking correctly add up to his answer. *JB* If the underestimation is occurring in a planning problem, a comprehensive list of events that could upset the plans can be elicited from the expert (Kahneman & Tversky, 1978). Creating this list may make the expert realize that he should take some of these possibilities into account in his problem solving. *JB**

2.2.6. Step 6--Analyze the data for the occurrence of bias

This step can be conducted at two different levels: (1) simple analyses and (2) complex analyses.

2.2.6.1. Simple analyses. The first level is often used if step 6 is the only one that the knowledge engineer will be doing. In addition, these simpler tests are frequently used on cognitive biases because they are more amenable to straightforward analyses.

However, even doing this simpler level of analysis, presupposes that sufficient data was gathered. One of the major problems in testing for bias is having a sufficient sample size. The sample can be composed of multiple experts' data on the same problem or a single expert's data on many problems. To ensure that there is enough analyzable data, it is helpful to consult with a statistician before beginning the elicitation. In addition, we would caution the analyst to beware of misrepresenting the data in analyzing it. The use of many statistical techniques require that the analyst make particular assumptions about the data. If these assumptions are not warranted, the data will be misrepresented, biased according to the motivational definition. For example, many statistical techniques require that the analyst assume that the data is normally distributed. The use of these techniques would often misrepresent the data, which in our experience has been multimodally distributed. For this reason, we recommend the use of nonparametric statistical procedures and data-based simulation techniques wherever possible because they do not require distribution assumptions.

Two examples of testing at a simple level are given below.

Group think bias was tested for in a project whose goal was to elicit estimates on the relevance of selected weapon-related components to particular military needs (Meyer, Peaslee, and Booker, 1982). The experts had assigned weights to describe how a weapon would meet a defense need. The weights ranged from 0, meaning not at all related, to 3, meaning 80% to completely related. Group think had earlier been judged a possibility because the experts were from the same work group, had worked together before, and had been selected for this project by their supervisor, who was present to lead some of the sessions. In addition, behavioral aggregation was being used, which further predisposed the elicitation toward group think. Thus, it was possible that the leader's views were influencing the discussions and the assignment of weights. To test for this possibility, the estimates were grouped according to who had led the session and analyzed for correlations to the weights. There was no evidence of this effect, but the conclusion was not absolute because some of the factors overlapped making definitive testing of their effects impossible.

Underestimation of uncertainty was the focus of analysis in another project. The experts had estimated the likelihood of achieving national magnetic fusion milestones within particular time periods. This bias was considered likely because people typically underestimate uncertainty, and therefore, the time and money needed to complete various tasks. In addition, people seem to be more prone to this underestimation when they are estimating on uncertain situations, such as fusion technology. Research by Rand Corporation has shown that cost estimates for a first-of-a-kind project display greater cost errors than for other kinds of projects (Merrow, Chapel, and Worthing, 1979). For these reasons, the expert's ranges on a question were analyzed. They were found to be within one standard deviation of the data. This indicates that the experts thought that they were adequately accounting for uncertainty when they were only accounting for about 60% (Meyer, Booker, Cullingford, and Peaslee, 1982).

2.2.6.2. Complex analyses. The second level of analysis involves using the data gathered from monitoring and adjusting for bias (steps 4 and 5) to do a more complete analysis. Information gathered during these steps, such as who appeared to be exhibiting a particular bias or differed in their problem-solving processes, can offer insights into what biases to test for. In addition, if the later analysis yields confusing results, this data can aid in their interpretation.

For conducting the more complex analyses, we recommend the use of multivariate and simulation techniques. Multivariate techniques allow the simultaneous consideration of two or more variables of interest (Tietjen, 1986). This capability is important because expert data is frequently conditioned on other data. An expert's data can be conditioned on such factors as aspects of the elicitation situation and the expert's interpretation of the question and method of solving the problem. Because of the small sample sizes usually associated with knowledge acquisition, these effects can overlap and cause problems in the analysis. For example, in the early project mentioned above, we were not able to definitively test for group think because we could not separate this variable from others.

We have used several multivariate techniques, such as cluster analysis, discriminant analysis, and general linear models in trying to find the one most applicable to the data in question. If these techniques require distribution assumptions that are not applicable to the data, simulation techniques can be used. Like multivariate techniques, simulation techniques can be used to discover relationships between the variables; unlike them, they do not impose assumptions about the distributions. A statistician can provide guidance in selecting the specific analytic techniques.

Two examples of tests performed at this more complex level are given below.

In two studies we searched for sources of correlation, or dependence, between expert's solutions (Booker and Meyer, 1988; Meyer and Booker, 1987). The experts solved problems in individual meetings. Data was gathered on aspects of the expert's professional backgrounds, the problems, and the expert's means of solving the problem. This data was combined in a large data base, and its multivariate relationships were analyzed using multivariate and simulation techniques.

Anchoring bias, in another project, was examined using an experimental design**.** The experts in this project received limited information from individuals threatening violence. They had to decide if the communication was a ploy or if it should be taken seriously. During the monitoring of the first set of elicitations (step 4), the knowledge engineer discovered that the experts were making implicit assumptions. Follow-up sessions indicated that about half of the experts assumed that the communications were ploys until proven otherwise; the other half assumed that they were credible until proven otherwise. For the problem-solving sessions, experts were assigned to questions so a pair of each assumption type would be represented by a problem. With such a design and the earlier mentioned techniques, it was possible to test for the effect of this assumption on the expert's final conclusion.

3. Summary

An approach for treating bias have been proposed here. The approach includes guidance on when to use particular combinations of steps depending on the knowledge engineer's definition of bias, reasons for focusing on bias, and interest in particular sources of bias. The steps are (1) anticipate which biases are likely to occur in the planned elicitation; (2) redesign the elicitation to make it less prone to the anticipated biases; (3) make the expert aware of the potential for these biases and familiarize him with the elicitation procedures; (4) monitor the elicitation for their occurrence; (5) adjust, in real time, to counter their occurrence; and (6) analyze the data to determine more definitively if they occurred. The steps can be followed in sequence or singly, depending on the needs of the knowledge acquisition task. The approach describes in detail how to perform each step using selected sources of bias as examples.

While this program was developed for manual knowledge acquisition, it could be applied in part to automated, or interactive, acquisition. In automated knowledge acquisition, there will still be a need for the knowledge engineer, according to Boose and Gaines (1988, p. 11). Thus, the knowledge engineer could, in between assisting experts with the automated tools, follow the steps for countering bias. Furthermore, the automated tools could assist the knowledge engineer in detecting and countering bias, especially those covered by the cognitive definition.

In this paper, we hope to make knowledge acquisition practitioners more aware of bias, their options in defining it, and their potential contribution to the problem. We aim to provide a more explicit, disciplined approach to countering bias in elicitation. Such an approach benefits not only manual knowledge acquisition but creates a foundation for automating the countering of bias. Through this paper, we hope to stimulate other work on handling bias in knowledge acquisition and to obtain feedback for developing our proposed program.

Much of the information in this paper is from chapters 1, 3, 8, and 10 of the report, NUREG/CR-5424, Eliciting and Analyzing Expert Judgment: A Practical Guide (Meyer and Booker, 1989). The financial support of the Nuclear Regulatory Commission (FIN #7225-5) for the writing of this report is gratefully acknowledged.

Information was also excerpted from pages 67-73 of the 1989 AAAI knowledge acquisition tutorial presented by John Boose and Mildred Shaw.

We are indebted to the participants of the 4th Knowledge Acquisition Workshop for comments on the first draft of this paper. In particular, Hans Acckermans and Stephen Regoczei generously gave their time and ideas for developing this paper.

References

ADAMS-WEBER, J. (1981). George A. Kelly--An Appreciation, in M. L. G. Shaw (Ed.), Recent Advances in Personal Construct Technology, pp. 1-7, London, England: Academic Press.

ARMSTRONG, J. S. (1981). Long-Range Forecasting: From Crystal Ball to Computer, New York, NY: Wiley-Interscience.

ASCHER, W. (1978). Forecasting: An Appraisal for Policymakers and Planners, Baltimore, MD: John Hopkins University Press.

BARON, R. A. AND BRYNE, D. (1981). Social Psychology: Understanding Human Interaction, Boston, MA: Allyn and Bacon.

BATESON, G. (1979). Mind and Nature: A Necessary Unity, New York, NY: Bantam Books.

BENBASAT, I. AND DHALIWAL, J. S. (1989). A Framework for the Validation of Knowledge Acquisition, *Knowledge Acquisition* , Vol. 1, no. 2, p. 215-233 .

BOOKER, J. M. AND MEYER, M. A. (1988). Sources and Effects of Interexpert Correlation: An Empirical Study, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 18, no. 1, p. 135-142.

BOOSE, J. H. AND GAINES, B. R. (1988). Knowledge Acquisition for Knowledge-based Systems, *Tutorial MP4 of the Seventh National Conference on Artificial Intelligence of the American Association for Artificial Intelligence,* St. Paul, MN, August .

BOOSE, J. H. AND SHAW, M. L. G. (1989). Knowledge Acquistion for Knowledge-based Systems, *AAAI-89 Tutorial Notes of the Eleventh International Conference on Artificial Intelligence*, Detroit, MI, August .

CAPEN, E. C. (1975). The Difficulty of Assessing Uncertainty, presented at the Society of Petroleum Engineers AIME 50th Annual Fall Technical Conference, Dallas, Texas, September .

CLANCY, W. J. (1989). Viewing Knowledge Bases as Qualitative Models, *IEEE Expert*, Vol. 4, no. 2, p. 9-24.

CLEAVES, D. A. (1986). Cognitive Bias and Corrective Techniques: Proposals for Improving Elicitation Procedures for Knowledge-based Systems, in *Proceedings from the 1st Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, p. 9-0 to 9-11.

DAWES, R. M., FAUST, D. AND MEEHL, P. E. (1989). Clinical Versus Acturial Judgment, *Science*, Vol. 243, p. 1668-1673.

DENNING, P. J. (1986). The Science of Computing: Will Machines Ever Think? *American Scientist,* Vol. 74, p. 344-346.

DHALIWAL, J. S. AND BENBASET, I. (1989). A Framework for the Comparative Evaluation of Knowledge Acquisition Tools and Techniques, *Proceedings of the 4th Annual Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Canada, pp. 9-1 to 9-24.

DOUGLAS, M. (1970). <u>Purity and Danger: An Analysis of Concepts of Pollution and Taboo</u>, London, England: Routledge and Kegan Paul.

FAUST, D. AND ZISKIN, J. (1988). The Expert Witness in Psychology and Psychiatry, *Science*, Vol. 241, p. 31-35.

FELDMAN, B. Z., COMPTON, P. AND SMYTHE, G. A. (1989). Hypothesis Testing: An Appropriate Task for Knowledge-Based Systems, *Proceedings of the 4th Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Canada, October, pp. 10-1 to 10-20.

FESTINGER, L. (1957). <u>A Theory of Cognitive Dissonance,</u> Palo Alto, CA: Standford University Press.

GAINES, B. R. (1989). *Knowledge Acquisition,* Vol. 1, no. 1, p. 39-58.

GOFFMAN, E. (1959). <u>The Presenation of Self in Everyday Life</u>, Garden City, NY: Doubleday Anchor Books.

GOLD, P. E. (1987). Sweet memories, *American Scientist*, Vol. 75, 151-155.

HAYES-ROTH, B. (1980). Estimation of Time Requirements During Planning: Interactions Between Motivation and Cognition, Rand Corporation report N-1581-ONR, Santa Monica, CA.

HENRION, M. AND COOLEY, D. R. (1987). An Experimental Comparison of Knowledge Engineering for Expert Systems and for Decision Analysis, in the *Proceedings of the American Association for Artificial Intelligence*, p. 471-476, Los Altos, CA: Morgan Kaufman Publishers.

HOGARTH, R. M. (1980). *Judgement and Choice*, New York: Wiley and Sons.

JANIS, I. C. (1972). <u>Victims of Group Think: A Psychological Study of Foreign Policy Decisions and Fiascos</u>, Boston, MA: Houghton Mifflin.

LANGER, E. J. (1989). <u>Mindfullness</u>, Reading, MA: Addison-Wesley.

MARTZ, H. F., BRYSON, M. C. AND WALLER, R. A. (1985). Eliciting and Aggregating Subjective Judgments--Some Experimental Results, in the *Proceedings of the 1984 Statistical Symposium on National Energy Issues,* (Nuclear Commission report, NUREG/CP-0063, Nuclear Regulatory Commission, Washington, D.C.), p. 63-82, October, Seattle WA.

MCGRAW, K. L. AND HARBISON-BRIGGS, K. (1989). <u>Knowledge Acquisition: Principles and Guidelines</u>, Englewood Cliffs, NJ: Prentice-Hall.

MERROW, E. W., CHAPEL, S. W. AND WORTHING. C. (1979). A Review of Cost Estimation in New Technologies: Implications for Energy Process Plant, Rand report R-2481-DOE, Santa Monica, CA.

MEYER, M. A. (1986). Human Factors Affecting Subjective Judgments, in the *Proceedings of the Thirty-First Conference on the Design of Experiments in Army Research Development and Testing,* p. 85-98, Madison, WI.

MEYER, M. A. AND BOOKER, J.M. (1989). Eliciting and Analyzing Expert Judgment: A Practical Guide, Nuclear Regulatory Commission report NUREG/CR-5424, Washington, D.C. and Los Alamos National Laboratory technical report LA-11667-MS. Los Alamos, NM.

MEYER, M. A. AND BOOKER, J. M. (1987). Sources of Correlation Between Experts: Empirical Results From Two Extremes, Nuclear Regulatory Commission report NUREG/CR-4814, Washington, D.C., and Los Alamos National Laboratory technical report LA-10918-MS, Los Alamos, NM.

MEYER, M. A., BOOKER, J. M., CULLINGFORD, H. S. AND PEASLEE, A. T., JR. (1982). A Data-Gathering Method for Use in Modeling Energy Research, Development and Demonstration Programs, in T. N.

Veziroglu, (Ed.), Energy Programs, Policy, and Economics: Alternative Energy Sources, p. 421-430, Ann Arbor, MI: Ann Arbor Science.

MEYER, M. A. AND JOHNSON, E. R. (1985). Proceedings from the Technology Control Panel Workshop, Los Alamos National Laboratory technical report LA-UR-85-4078, Los Alamos, NM.

MEYER, M. A., MNISZEWSKI, S. M. AND PEASLEE, A. T. JR. (1989). Using Three Minimally Biasing Elicitation Techniques for Knowledge Acquisition, *Knowledge Acquisition*, Vol. 1, no. 1, p. 59-71.

MEYER, M. A., PEASLEE, A. T., JR. AND BOOKER, J. M. (1982). Group Consensus Methods and Results, Energy Systems Group, Los Alamos National Laboratory technical report LA-9584-MS, Los Alamos, NM.

MILLER, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *Psychological Review*, Vol. 63, p. 81-97.

MUMPOWER, J. L., PHILLIPS, L. D., RENN, O. AND UPPULURI, V. R. R. (1987). Expert Judgment and Expert Systems, NATO ASI Series, New York: Springer-Verlag.

ORTIZ, N. R., WHEELER, T. A., MEYER, M. A. AND KEENEY, R. L. (1988). The Use of Expert Judgment in NUREG-1150, (Sandia National Laboratory technical report SAND88-2253C, Albuquerque, NM.), to appear in the *Proceedings of the 16th Water Reactor Safety Information Meetings*, October 1988, Gaithersburg, Maryland.

PAYNE, S. (1951). The Art of Asking Questions, Princeton, NJ: Princeton University Press.

REGOCZEI, S. AND HIRST, G. (1989). Sortal Analysis with SORTAL: A Software Assistant for Knowledge Acquisition, *Proceedings of the 4th Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Canada, October,.pp. 25-1 to 25-18.

ROSENFIELD, I. (1988). The Invention of Memory: A New View of the Brain, New York: Basic Books.

SAATY, T. L. (1980). The Analytic Hierarchy Process: Planning, Priority Setting, and Resource Allocation, New York, NY: McGraw-Hill.

SEAVER, D. A. (1976). Assessments of Group Preferences and Group Uncertainty for Decision Making, Social Science Research Institute, University of Southern California, Los Angeles, CA.

SHAW, M. L. G. AND WOODWARD, J. B. (1989). Mental Models in the Knowledge Acquisition Process, *Proceedings of the 4th Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Canada, October, pp. 29-1 to 29-24.

SPETZLER, C. S. AND STAEL VON HOLSTEIN, C-A. S. (1975). Probability Encoding in Decision Analysis, *Journal of the Institute of Management Science*, Vol. 22, p. 340-358.

TIETJEN, G. L (1986). A Topical Dictionary of Statistics, New York: Chapman and Hall.

WEISSENBERG, P. (1971). Introduction to Organizational Behavior: A Behavioral Science Approach to Understanding Organizations, Scranton, OH: Intext Educational.

WIELINGA, B., AKKERMANS, H., SCHRIEBER, G. AND BALDER, J. (1989). A Knowledge Acquisition Perspective on Knowledge-Level Models, *Proceedings of the 4th Knowledge Acquisition for Knowledge-Based Systems Workshop*, Banff, Canada, October, pp. 36-1 to 36-22.

ZIMBARDO, P. G. (1983). To Control a Mind, *Standford Magazine*, Winter, p. 59-64.