Chapter 3: THREE PARADIGMS FOR VIEWING DECISION BIASES

Marvin S. Cohen
Cognitive Technologies, Inc.

I. EVALUATING DECISIONS

Decisions can, and do, go wrong: a doctor misdiagnoses a patient's illness; a new product fails in the marketplace; a military commander mistakenly engages a civilian aircraft. Undesirable outcomes, however, do not necessarily imply faulty decision making; consider General Hod's decision to shoot down a Libyan airliner (described at the end of the last chapter). Even though the aircraft turned out *not* to be on a hostile mission, his conclusion that it was hostile might have been justified by the information he had at the time, by his efforts to gather further relevant data, and by the costs of a terrorist incident. It can also happen, of course, that a bad decision works out well. It is natural, then, for psychologists to look for a way of evaluating the decision itself, or the process that led to the decision, as distinct from its outcomes: to point, for example, at false prior beliefs, inappropriate priorities, shaky inferences from data, or even logical inconsistencies, rather than simply a bad outcome.

A widely accepted research paradigm in psychology (e.g., Kahneman, Slovic, and Tversky, 1982) has taken Bayesian decision theory as the standard by which reasoning is to be judged, and has identified pervasive *patterns* of error, called biases, in laboratory performance. According to these researchers, unaided decision processes employ rules of thumb (or "heuristics") that under many (but not all) conditions lead to "severe and systematic errors" (Tversky and Kahneman, 1974). The bottom line of this research has been a rather pessimistic view of human reasoning. To illustrate, let us extrapolate some of the laboratory results on biases and their interpretation to a hypothetical physician:

- In assessing the probability that she will encounter cases of diseases A, B, and C among her patients, the physician may rely on the ease with which she can recall or imagine instances of each disease. This is the so-called "availability heuristic", postulated by Tversky and Kahneman (1973). Availability may be influenced by factors like the recency or salience of the physician's own experiences, which do not reflect the true relative frequencies of the diseases in the relevant population.

- In estimating a quantity, such as the required length of treatment for disease B, the doctor may first generate her best guess, then adjust it upwards and downwards to allow for uncertainty, e.g., 10 days plus or minus 2. This is the "anchoring and adjustment" heuristic. According to Tversky and Kahneman (1974), adjustments are typically insufficient. The result is an "overconfidence" bias (Lichtenstein, Fischhoff, and Phillips, 1982): e.g., 95% confidence intervals that contain fewer than 95% of the actual cases.

- If the description of a patient's symptoms resembles the stereotypical picture of disease C, the physician may assign C a high probability, even if it is in fact extremely rare in comparison to diseases A and B. This is called "base rate neglect," and may result from the "representativeness heuristic," a tendency to judge probabilities by the similarity of a sample or instance to a prototype of its parent population (Kahneman and Tversky, 1972).

- Once she has arrived at an opinion about the illness that is causing the patient's symptoms, the doctor may fail to revise her opinion in the light of new symptoms or test results that conflict with it; she may even find ways to explain away the apparently conflicting data. This is the so-called "belief bias," or "confirmation bias," extensively reviewed in Nisbett and Ross (1980). Tversky and Kahneman (1980) attribute it to the compelling nature of causal beliefs.

- In evaluating different treatment alternatives, the doctor may take her most important objective first (e.g., reducing the size of a tumor by x%) and eliminate options that fail to achieve it; she may then compare the surviving options to her next most important goal (e.g., avoiding certain side-effects), and so on until only one option is left. This is the "elimination-by-aspects" strategy (Tversky, 1972). It may lead the doctor to overlook important compensatory relationships, e.g., she may reject an option that just misses one goal but which is outstanding in other respects.

- The doctor may regard a treatment more favorably if she happens to think of the outcomes in terms of potential gains, such as chance of survival, and less favorably if she happens to think of those same outcomes in terms of potential losses, such as chance of death. This is the "reference effect," discussed in Tversky and Kahneman (1981). They attribute it to pre-decisional processes that select a neutral reference point for the representation of outcomes, and to judgmental processes that weight losses with respect to the reference point more heavily than comparable gains.

On what grounds do psychologists claim that the doctor's assessments, inferences, and choices in these examples are mistaken? In some cases (e.g., availability, anchoring and adjustment) the doctor's probability assessments may be compared to actual frequencies of the relevant events (i.e., disease types and treatment durations, respectively). In other cases (such as the examples of base-rate neglect and confirmation bias) the relevant events are more complex and one-of-a-kind, and empirical frequencies will often be unavailable. Nevertheless, whether frequencies are available or not, the doctor's judgments or inferences can be evaluated in terms of their *internal* coherence. Bayesian probability theory provides a normative standard that is accepted by many researchers, and which specifies how a person's beliefs *should* be related to one another. Similarly, Bayesian decision theory (which includes probability theory as a part) provides a standard for evaluating the rationality of the doctor's *choices* among treatment options, even when a large sample of actual choice outcomes is not available, in terms of the internal coherence among her beliefs, preferences, and actions.

According to this research, a decision bias is not a lack of knowledge, a false belief about the facts, or an inappropriate goal; nor does it necessarily involve lapses of attention, motivation, or memory. Rather, a decision bias is a systematic flaw in the internal relationships among a person's judgments, desires, and/or choices. Human reasoning depends, under most conditions, on heuristic procedures and representations that predictably lead to such inconsistencies. It follows that human reasoning processes are error-prone *by their very nature*.

Few recent areas of psychological research have had as much impact on psychology as the work on "heuristics and biases," or have been as widely cited in the literature of other fields and in the popular press (as noted by Berkeley and Humphreys, 1982; and by Lopes, 1988). Perhaps more importantly, this research has motivated efforts to help people make better decisions, by automating or supporting the "normatively correct" methods for processing information (e.g., Edwards, 1968). Nevertheless, there is a growing chorus of dissent (e.g., Jungerman, 1983; Berkeley and Humphreys, 1982; Anderson, 1986; Lopes, 1988; Shanteau, 1989; Einhorn and Hogarth, 1981; and many others). Some of the original investigators have begun to emphasize methodological and conceptual problems in the research on biases, and have concluded that, at the very least, human shortcomings have been exaggerated at the expense of human capabilities (e.g., Kahneman and Tversky, 1982; Von Winterfeldt and Edwards, 1986).

A focus on naturalistic decision making adds a new perspective to this debate, opening to question some of the basic assumptions of the decision bias research. For example, what is the actual impact (or lack of impact) of each bias in real-world domains? Can we predict when errors will be serious and when not? Does the dynamic and open-ended quality of real tasks, as opposed to laboratory tasks, help reduce the effects of biases? Are biases sometimes mitigated by task-specific knowledge? Do they sometimes occur as side effects of using knowledge effectively? Do biases sometimes reflect a decision maker's capacity limitations or adaptations to the cost of information processing? How are such costs measured, and how is such adaptation achieved? Finally, are we really sure what a decision-making "error" is? What conclusions follow if we look more closely at how people actually reason before fitting a normative model that says how they "ought" to reason?

These questions are by no means settled. Nor is there any assurance that the answers, when they come, will support a more optimistic view of decision making; e.g., errors may be worse rather than better in dynamic, open-ended environments. Nevertheless, these questions establish the need for research that is both carefully controlled and representative of real-world decisions. It is surely worthwhile to consider seriously an alternative to the standard view, which, although not proven, is consistent with all the evidence we now have. According to this alternative picture, *people tend to use decision-making strategies that make effective use of their substantive knowledge and processing capacity; such strategies are generally subject to incremental revision and improvement in dynamic environments; and the net result is performance that is usually adequate, though subject to improvement in specific respects.* Such a picture, by exposing and challenging assumptions underlying the standard view, may be a fruitful stimulus to research that is both more valid and ultimately, more useful. Moreover, if this picture is true, even in part, *decision aiding and training should be targeted at strengthening the decision maker's preferred approach to a problem rather than replacing it altogether*.

Our discussion of these topics is organized as follows. Section II of this chapter will describe three alternative paradigms for viewing decision biases, comparing the paradigms with respect to the types of normative and explanatory models they utilize, and the style of empirical research they encourage. The next chapter will explore the "naturalistic paradigm" in more depth, discussing six challenges to the prevailing "rationalist" view of decision biases. These challenges emphasize the information-rich and dynamic character of the decision environment, the capacity limitations and knowledge of the decision maker, and the importance of cognitive and behavioral criteria in the selection of an appropriate normative benchmark. Taken together, these challenges lead to a more convincing and more useful, naturalistic concept of decision bias. In Section D we turn finally to the implications of the naturalistic paradigm for decision aiding and training.

II. A TALE OF THREE PARADIGMS

Recent demonstrations of decision errors have been dramatic, but not because anyone had really thought that humans were perfectly rational. Discrepancies between behavior and apparent normative constraints had been well-known to an earlier generation of researchers. By the same token, recent researchers on biases have not painted an unremittingly gloomy picture. What happened may be best illuminated by the metaphor of a "paradigm shift" (cf., Lopes, 1988): research on decision biases has changed the way conflict between behavior and normative models is *interpreted*; it has also changed the character of psychological models, and the style of empirical research. New work may now be causing all of these to change once again.

There are three basic paradigms, or filters, through which this subject may be viewed: the "formal-empiricist" paradigm, which preceded the research on biases as the standard approach to decision making; the "rationalist" paradigm, which has spawned most of the present controversy; and the naturalistic paradigm, which is now emerging and, we argue, offers the most fruitful perspective.

A. The Formal-Empiricist Paradigm

Up to the late 1960's, researchers on decision making wanted their theories to do double-duty: both to fit empirically observed behavior and to have normative plausibility. If behavior failed to fit a model, they did not condemn the behavior as "irrational"; instead, they regarded the *model* as inadequate - both to describe behavior and to evaluate it (Lee, 1971; Beach, Christensen-Szalanski, and Barnes, 1987; Barclay, Beach, and Braithwaite, 1971). The experimenter's task was to devise a new formal description of the anomalous behavior that brought out its good features - that provided a *rationale*.

Paradoxes, in which carefully considered judgments or decisions clashed with a model, were occasions to question, and possibly to improve, the model. According to a proposed normative rule for choice under uncertainty, for example, one should select the option that has the highest "expected value." The expected value of an option is an average obtained by adding the payoffs associated with each possible outcome, while weighing each outcome by its probability of occurrence. (If a lottery ticket pays $1000 to the winner, and there is a 1 in 10,000 chance of winning, the expected value of the ticket is: (.0001)($1000) + (.9999)($0) = $0.10; a rational decision maker, according to the expected-value rule, should be willing to pay no more than ten cents for such a ticket.)

Many features of ordinary behavior - such as purchasing insurance and gambling - are inconsistent with maximization of expected value. Instead of heralding these as "decision biases," Bernoulli tried to make sense of them (focussing in particular on a famous problem called the St. Petersburg paradox). In 1738 he proposed replacing the *objective* measure of preference (e.g., money) with a *subjective* one (utility), and assumed that the utility of each additional dollar is smaller as the number of accumulated dollars increases. The same rule, with more elaborate assumptions about the way utility is related to dollars or other objective payoffs, is still used today to reconcile the normative model with ordinary intuitions and behavior.

Although utility is subjective, probability might be defined objectively, in terms of the relative frequency of an event (e.g., heads) in some sample space (e.g., a series of coin tosses). Yet people also find it meaningful to talk about, and make decisions that depend on, probabilities of unique events, e.g., "Bush will probably be re-elected President." (Even the "objective" notion of probability seems to depend on judgment in selecting an appropriate sample space.) The next major step in the evolution of formal-empiricist models replaced frequency-based probabilities with "subjective probabilities," or personal degrees of belief. As a price for accommodating unique events and individual differences in decision-making, normative models could no longer dictate the *content* of a person's beliefs or preferences.

What *did* normative models do? De Finetti (1937/1964) and Savage (1954) developed formal systems for merging subjective preferences and subjective probabilities in a new normative rule, maximization of subjectively expected utility (SEU). This rule looks the same as maximization of expected value, except that subjective probabilities and utilities are substituted for objective probabilities and payoffs, respectively. The surface similarity disguises an important difference, however. Unlike maximization of expected value, the SEU "rule" does not imply a *procedure* for decision making: probabilities and utilities are defined by reference to a decision maker's choices among gambles; they do not *guide* such choices. There is no longer a rationale for starting with probabilities and preferences as inputs and deriving choices, or measures of the desirability of options, as outputs; the decision maker could just as well start with the desirability of an option and assess probabilities and utilities afterwards. What the SEU rule and the associated laws of probability do is specify *consistency relationships* that probabilities and utilities (and the choices that define them) should satisfy.

Savage and De Finetti successfully showed that behavior satisfying these consistency relationships had certain very general attractive features, which they described in postulates or axioms: for example, judgments about the *probability* of an event are the same regardless of changes in *preference* for that event (Savage's independence postulate). Conversely, de Finetti and Savage showed (by derivation of the SEU rule from such postulates) that if you want your decisions to have the attractive features, then your judgments and decisions *must* satisfy the SEU rule and the laws of probability.

Psychologists tested formal-empiricist models by asking subjects to make choices in sets of interrelated gambles that varied in their uncertain events and payoffs (e.g., Davidson, Suppes, and Siegel, 1957). If a subject's choices were consistent with the SEU axioms, then utilities of the payoffs, and subjective probabilities of the events, could be said to exist and could be numerically defined for that subject. (For example, if a subject were indifferent between $.40 for sure and a gamble with a 50-50 chance of winning $1.00, the utility of $.40 would be one-half the utility of $1.00 for that subject.) Experimental tests, however, typically showed deviations from the formal-empiricist predictions (e.g., Mosteller and Nogee, 1951; Marks, 1951; Swets, 1961; Tversky, 1967). One conclusion, according to a review by Peterson and Beach (1967), was that humans were pretty good, but not perfect, "intuitive statisticians." Another response was to continue to loosen the constraints imposed by models of decision making: for example, to introduce a notion of probabilistic choice (obeying a weaker set of axioms) in place of deterministic choice (Luce, 1959, 1977), or to include the variance among outcomes as an attribute affecting the desirability of a gamble (Allais, 1953).

The formal-empiricist paradigm focussed on behavioral testing of formal models, not on the cognitive processes actually underlying decisions. Little effort, for example, was made to collect concurrent think-aloud protocols from subjects as they made decisions, to interview them afterwards about the reasons for their choices, or even to vary parameters that might affect performance but which were not in the formal model. The models themselves, as already noted, impose mathematical consistency constraints on a subjects' judgments and preferences, but make no reference to actual psychological steps or representations. Not surprisingly, then, some psychologists have questioned the cognitive plausibility of SEU even in cases where it fits behavior. Lopes (1983; Schneider and Lopes, 1985), for example, has argued that real decision makers are less concerned with an option's *average* outcome, than with the outcomes that are most *likely* to occur.

In sum (as shown in the first column of Table 1), the formal-empiricist paradigm: (1) allowed human intuition and performance to drive normative theorizing, along with more formal, axiomatic considerations; (2) used the resulting normative theories as *descriptive* accounts of decision-making performance; and (3) tested and refined the descriptive/normative models by means of systematic variation of model parameters in artificial tasks.

B. The Rationalist Paradigm

Since Plato and earlier, "rationalist" philosophers found ordinary reasoning riddled with flaws, and have held out the promise of a more rigorous method for establishing the truth. At least since Descartes, the preferred method has involved replacing intuitive leaps of thought by short, logically self-evident steps. Both of these elements - the disparagement of ordinary reasoning and the promotion of more valid methods - have flourished, somewhat independently, in the last 20 years.

Decision analysis has "come of age" as a body of techniques for applying decision theory in management consulting (Ulvila and Brown, 1982). In contrast to the purely formal constraints of decision *theory*, decision *analysis* specifies procedures: e.g., Bayesian inference (for drawing conclusions or making forecasts based on incomplete or unreliable evidence), decision tree analysis (for choices with uncertain outcomes), and multiattribute utility analysis (for choices with multiple competing criteria of evaluation) (see Raiffa, 1968; Brown, Kahr, and Peterson, 1974; Keeney and Raiffa, 1976). The prescribed problem-solving strategy is to decompose a problem into elements, to have appropriate experts or decision makers subjectively assess probabilities and/or utilities for the components, and then to recombine them by the appropriate mathematical rule.

The prevailing concept of decision biases emphasizes the other side of rationalism: errors in *unaided* decision making. Errors have now taken on a more dramatic and important role than in previous research. Underlying this change was a "paradigm shift" in the relationship between normative and descriptive theorizing. The rationalist paradigm takes decision theory as a norm that is fully justified by its formal properties (i.e., the postulates that entail it), not by its fit to the way people in fact make decisions; rationalism attributes discrepancies between behavior and a model to the irrationality of decision makers, not to flaws in the model.

What motivated the change in attitude toward normative theories? According to Kahneman and Tversky (1982), the goal of their work was to make the psychology of decision making more cognitive. The formal-empiricist paradigm had combined normative and descriptive functions in the same formal models; the rationalist paradigm separates the functions of (cognitively) describing or explaining behavior and (formally) evaluating it. To make their case for a cognitive approach to explanation, however, Kahneman, Tversky, and other researchers had to do more than show divergence between the normative model and actual decisions: after all, in the formal-empiricist paradigm a model could always be revised. Decision bias researchers foreclosed this possibility by promoting a picture of normative theory as a *fixed* benchmark, immune to descriptive influence. The emphasis on human irrationality was a by-product.

Part of the appeal of rationalist research has been its use of easily understood demonstrations of errors. Everyday problems replaced artificial choices about lotteries; systematic variation of model parameters gave way to the less tedious presentation of a few simple variants of the same problem, sufficient to demonstrate inconsistency, though not to fit a model (Lopes, 1988). Readers can thus confirm conclusions about "biases" by checking their own intuitions. Nevertheless, the realism of these experiments is limited. Stimuli, although ostensibly drawn from real life, typically involve unfamiliar situations briefly described to college or high school students; moreover, they are usually pre-structured and pre-quantified, and involve a single response to a static rather than an unfolding situation: i.e., the problems specify numerical frequencies, probabilities, and/or payoffs, and subjects are asked to make one-time decisions about explicitly identified hypotheses or options.

Consider the following problem, used to demonstrate "base rate neglect" by Tversky and Kahneman (1980):

> A cab was involved in a hit-and-run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:
> (i) 85% of the cabs in the city are Green and 15% are Blue;
> (ii) A witness identified the cab as a Blue cab. The court tested his ability to identify cabs under the appropriate visibility conditions. When presented with a sample of cabs (half of which were Blue and half of which were Green) the witness made correct identifications in 80% of the cases and erred in 20% of the cases.
> Question: What is the probability that the cab involved in the accident was Blue rather than Green?

According to Tversky and Kahneman (1982), the probability that the guilty cab is Blue is computed from Bayes' Rule as follows:

$$(.15)(.80) \,/\, ((.15)(.80) + (.85)(.20)) \quad = \quad .41$$

where .15 is the base rate of Blue cabs, and .85 is the base rate of Green cabs. The base rates in this case strongly favor Green, and should outweigh the witness's testimony that the cab was Blue. Most subjects, however, inferred the probability of Blue to be at or near .80, apparently ignoring base rates. In a different variant of the problem, however, when the witness (item ii) was omitted from the problem, subjects did use base rates. Subjects also paid more attention to base rates when (i) was replaced by a more "causally relevant" base rate, the frequency of accidents attributed to Blue and Green cabs.

It is perhaps no surprise that in such experiments unaided human decision making has been found wanting. Biases have been observed in virtually every context that a statistician could imagine, including:

Assessment of probabilities -

- overconfidence in estimating probabilities of simple events,
- overconfidence in estimating probability distributions for quantities,
- reliance on ease of recall or generation (i.e., availability) in estimating frequencies of events in a class,

- overestimating, after an event has actually occurred, the probability that would have been assessed for the event before it occurred (hindsight bias),

- relying on theoretical preconceptions rather than data in estimating correlations between events (illusory correlation);

Inference -

- disregarding prior statistical information in responding to a single piece of evidence (base rate neglect),

- disregarding or discounting evidence that conflicts with a prior hypothesis (belief bias),

- confirmation bias in selecting observations to test a hypothesis;

- failing to update belief sufficiently in the light of new evidence (the conservatism bias),

- disregarding sources of uncertainty: acting "as if" earlier conclusions were known with certainty when reasoning proceeds in stages; adopting the most likely hypothesis as a "best guess,"

- ignoring sample size in assessing the accuracy of estimates or the probability of a sample,

- overestimating the probabilities of compound events (the conjunction fallacy),

- mistaken conceptions of randomness in estimating the probabilities of chance sequences of events: overestimating the probability of sequences with many alternations and representative proportions,

- overextreme predictions (neglecting regression to the mean) when predicting one quantity based on its correlation with another quantity,

Choice -

- effect on decisions of changes in the reference point for describing simple outcomes: e.g., risk aversion if outcomes are described as gains, risk seeking if the same outcomes are described as losses,

- effect on decisions of how multiple events are grouped together and associated with options as their outcomes ("psychological accounts"),

- effect on decisions of ignorance regarding true probabilities; resort to "worst-case" or "best-case" strategies in defining outcomes (Ellsberg's paradox),

- a greater effect on preference of reducing the probability of an outcome by a given ratio when the outcome was certain, than when it was not certain (the certainty effect, or common ratio effect),

- effect on decisions of how outcomes are sequenced in time;

evaluating outcomes as if earlier, uncertain contingencies were known to occur (the pseudo-certainty effect),

- effect on decisions of changes in the payoff for an outcome that is the same regardless of which option is chosen (Allais' paradox, or the common consequence effect).

We will touch on some of these errors in the discussions that follow; reviews can be found in Einhorn and Hogarth (1981); Slovic, Fischhoff, and Lichtenstein (1977); Tversky and Kahneman (1974); Smithson (1988); Hogarth and Makridakis (1981).

At the deepest level, biases are violations of consistency constraints imposed by probability theory or decision theory. (Even in the case of assessment biases, such as overconfidence and availability, agreement with empirical frequencies is relevant only if it is expected according to the formal theory.) For example, in one condition of the cab study (when the witness's testimony was included), subjects regarded base rates as irrelevant; while in another condition (when the witness was omitted), they regarded the same base rate data as relevant. In this example, as in many others, the formal equivalence between conditions is implicit, resting on the experimenter's assumptions about the subjects' other beliefs and/or preferences (e.g., that frequency data do not become less relevant due to the presence or absence of a witness).

From an explanatory point of view, however, biases have been attributed to any of a rather large number of cognitive processes. "Base rate neglect," for example, is explained in the cab problem by preference for information that can be interpreted causally (Tversky and Kahneman, 1982). In other cases the same formal error (base rate neglect) is explained by the tendency to assess probabilities in terms of "representativeness," or the similarity of a sample to its population (Tversky and Kahneman, 1982). Other explanatory hypotheses (which we have already alluded to) include availability (assessing the probability of a class by the ease of recalling or generating instances), anchoring and adjustment (beginning the estimation process with a salient anchor point and insufficiently adjusting it to allow for other factors), and representation processes that distort the decision maker's use of probabilities and payoffs in choice problems (Kahneman and Tversky's "Prospect Theory"; 1979).

Rationalist experiments, unlike formal-empiricist ones, often study factors not contained in the normative model: one example is the comparison of a "causally relevant" base rate with a "statistical" one in the base rate neglect study. Such manipulations expose additional reasoning errors (when the manipulated variable has an effect even though it should not), and also help test hypotheses about the cognitive processes underlying performance (e.g., the reliance on causal problem representations). Nevertheless, such manipulations fall far short of a systematic investigation of cognitive mechanisms. The primary goal of the experiments is simply to compare rival hypotheses of normatively "correct" versus "incorrect" behavior. As in formal-empiricist experimentation, there has been virtually no effort to explore cognitive processes more directly - by means of verbal protocols, interviews, or other process-tracing techniques (e.g., eye-movements or information requests).

The result, ironically, has been a failure by the rationalist paradigm to successfully integrate decision making research with the rest of cognitive psychology. Attention has focussed on classification of an ever-growing array of biases, defined negatively as deviations from "the" normative theory (Anderson, 1986); there has been insufficient effort to test alternative psychological explanations (Shanteau, 1989), to systematically study how and when the postulated heuristics and representation processes occur (Fischhoff, 1981), or to develop underlying theoretical principles and links with other areas of psychology such as problem-solving and learning (Wallsten, 1983).

The rationalist paradigm promoted a desirable transition to cognitively oriented theories of performance by adopting a less desirable tactic: creating a rigid normative concept as a strawman, and designing experiments that often do little more than discredit the strawman. In sum (as shown in the second column of Table 1), the rationalist paradigm (1) adopts a static and purely formal view of normative standards; (2) gives an explanatory account of reasoning in terms of a diverse set of unrelated cognitive mechanisms; and (3) experimentally demonstrates errors with pre-structured and pre-quantified "real life" stimuli.

C. The Naturalistic Paradigm.

The argument of this book is that a third paradigm is now emerging, distinguished from both the formal-empiricist and the rationalist paradigms by a more pronounced concern for decision making in realistic, dynamic, and complex environments (see Chapter 1 above), and the adoption of research methodologies that focus more directly on decision processes, as well as their real-world outcomes (see Woods, this volume). In this section and in the next chapter, I will explore the implications of that new paradigm for the notion of decision-making "error".

The naturalistic point of view involves more than simply looking for the same "biases and heuristics" in realistic settings. From the naturalistic perspective, an unquestioning acceptance of the relevance of classical normative standards is untenable, because real-world decision makers appear to use qualitatively different types of cognitive processes and representations. If these decisions are to be evaluated, other standards may often be appropriate. The new paradigm thus breaks the spell of classical probability/decision theory. Research is not tethered to it either as an explanatory model (the formal-empiricist paradigm) or as a strawman (the rationalist paradigm).

The naturalistic paradigm agrees with the rationalist approach (and differs from the formal-empiricist approach) in its explanatory emphasis on cognitive representations and processes. But it gets there without reliance on the tactic of looking for human "irrationality" under every behavioral stone. Formal models fail not because people irrationally violate them (as the rationalists argue), but because the models themselves do not capture the adaptive characteristics of real-world behavior. By focussing on the way people actually handle complex environments, the naturalistic paradigm illuminates the *functions* that cognitive processes serve. As a result, it stands a better chance of developing a successful and coherent set of explanatory models. One by-product is a decreased emphasis on relatively *ad hoc* cognitive procedures, like "heuristics," and more focus on an integrated picture of how knowledge structures are created and adjusted in dynamic environments.

The naturalistic point of view does not wholly banish the idea that errors occur when people make decisions -- or even the idea that those errors are systematic: everything "natural" is not good. In many respects, decision making in naturalistic settings is surely more difficult than in laboratory tasks (e.g., options, hypotheses, goals, and uncertainties may all be unspecified), and there is still a need to both evaluate and improve performance. The notion of a "decision bias" may yet prove useful in both respects. From the naturalistic perspective, however, evaluation of reasoning is more subtle and demanding: no longer a cookie-cutter comparison between performance and an unquestioned normative template. In the naturalistic framework, the reciprocity between normative and descriptive concerns that characterized the formal-empiricist approach can be retained - and even expanded- *if cognitive as well as behavioral criteria are incorporated into normative modeling*. Normative theories are intellectual tools whose justification depends in part on how well they fit a particular decision maker's goals, knowledge, and capabilities in the task at hand (cf., Shafer and Tversky, 1988); they are products of a negotiation between competing sets of intuitions about specific problems, general principles, and cognitively plausible methods.

We suspect that decision biases have not been satisfactorily identified, described, or explained within the prevailing rationalist approach. If errors are perceived where they do not exist and if other, perhaps more important types of error are overlooked, then decision aiding and training cannot hope to be effective or accepted. The naturalistic paradigm may cause us to see decision-making errors in a new light.

Chapter 4: THE NATURALISTIC BASIS OF DECISION BIASES

Marvin S. Cohen
Cognitive Technologies, Inc.

Research within the rationalist paradigm has thrown a spotlight on decision-making errors. But if the rationalist understanding of decision errors is inadequate, as I claimed in Chapter 3, what can we put in its place? A convenient strategy is to break down the rationalist argument and to examine the assumptions that seem most questionable from the naturalistic point of view. Out of this discussion, a new, naturalistic notion of decision bias will emerge.

The rationalist paradigm starts with the claim that unaided decisions are often formally inconsistent, and draws two conclusions. The rationalist argues, first, that the decision processes that predictably lead to such decisions are flawed. From the same premise, the rationalist also concludes (at least implicitly) that the outcomes to which such decisions lead will be undesirable. Both of these conclusions, as well as the original premise of formal inconsistency, can and have been challenged. Each of them, I will argue, disregards important characteristics of real-world environments, real-world decision makers, and real-world tasks.

> Rationalist claim: Inconsistent decisions lead to undesirable outcomes.
>
> > Challenge 1: A desirable *overall* outcome can be achieved in *real-world problem domains* even though some individual decisions have undesirable outcomes.
> >
> > Challenge 2: *Real-world task environments* facilitate desirable outcomes from *individual* decisions.
>
> Rationalist claim: Decision processes are flawed because they lead to inconsistent decisions.
>
> > Challenge 3: The inconsistency of decisions is mitigated by the benefits of using the decision maker's *real-world knowledge*.
> >
> > Challenge 4: Constraints on the decision maker's *information-processing capacity* justify use of non-Bayesian procedures.
>
> Rationalist claim: Decisions are often formally inconsistent.
>
> > Challenge 5: There are *alternative Bayesian models* that better capture the subject's understanding of specific tasks.
> >
> > Challenge 6: There are *alternative, non-Bayesian normative concepts* that justify the decision maker's way of approaching specific tasks.

Such challenges are not mutually exclusive. They vary, however, in where they draw the line against rationalist pessimism: i.e., at outcomes (challenges 1 and 2); more aggressively, at the decision processes that lead to the outcomes (challenges 3 and 4); or, more aggresssively still, at the very notion of an inconsistent decision (challenges 5 and 6).

Challenges (1) and (2) focus on the failure of the rationalist research to take account of the effects of decisions in real-world environments. By themselves, (1) and (2) do not challenge the rationalist paradigm very profoundly: they agree that decisions are often inconsistent, and that decision processes are therefore biased, but ask only how much it really matters in real task domains or in specific tasks. Challenges (3) and (4) go further; they focus on the failure of rationalist research to take full account of the "internal environment" with which the decision maker must deal. Decision processes may be justified *even though* they sometimes produce inconsistent decisions because they reflect effective use of the decision maker's knowledge or efficient rationing of her cognitive effort. Challenges (5) and (6) criticize the rationalist approach for disregarding or misunderstanding important elements of the decision task, from the point of view of the decision maker. Appropriate modeling of the decision maker's beliefs and preferences concerning a task, whether within a Bayesian or non-Bayesian normative framework, shows that decisions are not flawed even in the narrowest sense. Challenge (6) questions the underlying relationship between descriptive and normative concerns that has been central to the rationalist paradigm.

The naturalistic paradigm criticizes the rationalist concept of decision error, but it also proposes a replacement. Each challenge to rationalist pessimism is associated with a new concept of how decision making *can* go wrong, which is more plausible and ultimately more relevant to decision aiding and training than the rationalistic emphasis on formal inconsistency. Successful decision making does not require a logically complete formal model of every problem; what it does require is the ability to adapt: to focus attention where one's knowledge will have the most impact, and to adjust to new information, environmental changes, and to shortcomings that may appear in one's problem-solving approach. I will try to pull together the threads of these concepts into the outline of an alternative, naturalistic synthesis: at the descriptive/explanatory level, a set of recognition processes and metacognitive strategies that manage the deployment of knowledge and capacity in evolving situations (challenges 1, 2, 3, and 4); at the normative level (challenges 5 and 6), a process of assumption-based reasoning which accommodates the effort by decision makers to extend their knowledge into ill-understood, novel, and changing task environments.

I. DO INCONSISTENT DECISIONS LEAD TO BAD OUTCOMES?

Challenge 1: A desirable overall outcome can be achieved in real-world problem domains even though some individual decisions have undesirable outcomes.

Challenge (1) looks at the overall level of achievement of goals in a task domain, and concludes that the rationalist paradigm has overstated the frequency and the importance of biases: they are rare and, on average, inconsequential (Christensen-Szalanski, 1986). There are both (a) methodological and (b) formal arguments for this conclusion.

Variant (a): Methodological. In most research on biases, stimuli are not selected randomly, but are designed to maximize the chance of detecting suboptimal processes (Lopes, 1988). The answer given by an experimental subject to a problem nearly always points unambiguously to either the normatively correct decision process *or* to a heuristic. Such studies are efficient but biased: i.e., they will be systematically non-representative of domains in which heuristics and normative methods generally give the *same* answers. Subjects are also selected non-randomly (i.e., they are typically students) and do not represent the range of experience ordinarily found in a domain. In addition, the use of between-subjects designs makes it unclear whether any individual subject actually shows the bias under study (Fischhoff Slovic, and Lichtenstein, 1979; Scholz, 1987).

Another consequence of the rationalist effort to demonstrate errors has been stress on the statistical significance of effects rather than on their size, *measured on some meaningful scale*. Christensen-Szalanski (1986) has argued that researchers should provide domain-specific measures of the importance of a bias, and estimates of its prevalence in a domain. He cites the example of an impressive cluster of biases discovered in medical diagnosis, whose opportunities for occurrence turned out to be "embarrassingly small" and whose effects on treatment choices when they did occur were found to be negligible. Yates (1982; Centor, Dalton, and Yates, 1984) concluded empirically that overconfidence errors in estimating the probabilities of events were of little practical consequence. Even if biases occasionally cause large errors, they may reflect a reasonable tradeoff among goals in a task domain. For example, Klein (1989) speculates that the "belief bias" may be a by-product of efficient expectancy-based processing, that increases speed of response at the cost of errors in a minority of tasks.

Biases may also seem more frequent than they are because of the way in which findings are cited. Christensen-Szalanski and Beach (1984) found that studies reporting poor performance were cited preferentially in the social sciences literature over studies reporting good performance. Moreover, researchers in other fields (e.g., behavioral auditing; Shanteau, 1989) have tended to accept the "heuristics and biases" framework even when it provides only a poor fit to their own data.

Variant (b): Formal. In fact, there are theoretical reasons to expect that suboptimal strategies and normative models often agree in their *outcomes*. Von Winterfeldt and Edwards (1973) showed that it is unnecessary to be very precise or accurate in the estimation of parameters (e.g., importance weights) for many normative models; large errors will have a negligible effect on the decision maker's expected payoffs. Similarly, Dawes and Corrigan (1974) showed that simply counting the attributes or variables in a linear model is virtually as good, on average, as using the normatively correct weights. Moreover, simple linear models are robust enough to accurately describe many non-linear processes (Dawes, 1979; Goldberg, 1968). When there is random error in the assessment of parameters, simple but incorrect linear models can actually outperform complex models that correctly describe interdependencies among variables (Makridakis and Hibon, 1979; unpublished research by Paul Lehner). Thorngate (1980) showed with a Monte Carlo simulation that "biased" choice strategies can often select the same alternative as the Bayesian model, even when the biased strategies omit significant amounts of information. Large errors may occur, on the other hand, when important variables are omitted (Dawes and Corrigan, 1974) or if information is not utilized to eliminate grossly inadequate alternatives (von Winterfeldt and Edwards, 1975). These results suggest that using "optimal" procedures and accurately assessing parameters for them may be less important in successful performance than gross knowledge of what factors are relevant.

Even more critically, it is possible to anticipate the conditions under which suboptimal strategies will lead to bad results. For example, Payne, Bettman, and Johnson (1989) found that simply counting favorable aspects or outcomes of an option is not a bad strategy when aspects or outcomes do not differ much in importance or probability: e.g., house A is better located and has more rooms, house B is less expensive and prettier and available sooner; since house B has more advantages, choose house B. But elimination-by-aspects (which screens options by the more important aspects or more probable outcomes first) is a better approximation to normative methods when aspects or outcomes do differ significantly. Elimination-by-aspects rejects options that do not achieve cutoffs or aspiration levels on various dimensions (e.g., not enough rooms, not pretty enough, too expensive, etc.); it does not consider how much the decision maker would be prepared to give up on one dimension (e.g., cost) in order to gain a certain amount on another dimension (e.g., more rooms). Elimination-by-aspects in turn works well unless tradeoffs are crucial, i.e., it is significantly suboptimal *only* when there are options that just miss achieving a goal on one dimension, but which are outstanding in other respects.

This ability to pinpoint the undesirable outcomes to be expected and their conditions of occurrence for different strategies has an important implication for decision aiding and training: it opens the possibility of helping decision makers avoid the *specific* pitfalls that are associated with their preferred problem-solving method, rather than forcing them to radically alter the method itself. The very fact of using a non-normative decision strategy can no longer be regarded as an "error"; the failure to compensate for its known shortcomings can be.

Challenge 2: Real-world task environments facilitate desirable outcomes from suboptimal decisions.

Still more optimistically, challenge (2) argues that apparently suboptimal decision processes lead to desirable outcomes even in a single task, if we adopt a bigger picture of the task environment. Real-life tasks are not the "snapshot" decisions studied in the laboratory (Hogarth, 1981). Rather, "decisions" are typically (a) made in "information rich" environments, e.g., they are stretched out in time, with redundant cues, incremental stages of commitment, feedback from earlier actions, and shared responsibility; (b) the underlying circumstances of the task may themselves be changing; and (c) important consequences of the decision maker's actions may only be apparent over long time periods. The overall process may be unbiased, even though small time-windows are not.

Variant (a): Information richness. Many biases appear to reduce the amount of information utilized by the decision maker, or to reduce the impact of the information that the decision maker does use. Such biases would be exacerbated in a spare laboratory environment (where each cue is essential) and attenuated in an information-rich, highly redundant real-world environment (Einhorn, Kleinmuntz, and Kleinmuntz, 1979; Hoch and Tschirgi, 1983: Wright and Murphy, 1984). These considerations apply to biases in inference, multiattribute choice, and planning.

In the case of inference, two biases, conservatism (Edwards, 1968) and the belief bias (e.g., Tolcott et al., 1987), both involve a failure to update beliefs based on the full "normative" impact of new evidence. But in dynamic environments they may reflect an approach to belief revision that relies on feedback from initial guesses, additional redundant clues, and opportunities for subsequent correction. Tversky and Kahneman (1974) regard the "anchoring and adjustment" heuristic as a source of bias due to insufficient adjustment. But in a continuous environment, *iterated* adjustments may move assessments progressively closer to the "normative" target (Lindblom, 1959). Similarly, the availability heuristic may involve taking initial direction from the cues that first come to mind (i.e., instances of a class that are easily recalled) and adjusting later when other cues are encountered (Hogarth, 1981). In organizations, the effect of multiple players and multiple constituencies may offset the effects of individual errors in a similar manner (Lindblom, 1959).

In choice, elimination-by-aspects can be thought of as a failure to evaluate options based on the full "normative" implications of one's preferences. The decision analytic technique called multiattribute utility theory (Keeney and Raiffa, 1976) requires the up-front development of a full numerical model, incorporating precise tradeoffs among different criteria, and its application in one fell swoop to all options. Elimination-by-aspects, however, evaluates options by their performance on individual goals, without reference to tradeoffs. Similarly, Simon (1955) described another heuristic strategy called "satisficing," in which decision makers set goals on a few relevant dimensions and accept the first option they find that is satisfactory on them all. The apparent disadvantages of elimination-by-aspects and satisficing are: (a) no option may survive on all criteria, or (b) too many options may survive. In a dynamic environment, however, decision makers can adjust their goals as they encounter options, raising aspirations if they find it easy to discover satisfactory alternatives and lowering aspirations if they find it difficult (Simon, 1955). Such dynamic adjustments constitute a form of *learning* about preferences that is analogous to the learning about evidence discussed in the previous paragraph. From a rationalist point of view, it may violate Savage's postulate on the independence of beliefs and utilities, but it may nonetheless produce highly adaptive results.

Finally, decision makers seldom stop to plan out all their options and all the contingencies that might befall them; in short, they fail to make full "normative" use of all the available information about options and future outcomes. Fortunately, as Connolly and Wagner (1988) point out, only a few decisions (e.g., having a child, waging nuclear war, committing suicide) require once-and-for-all non-reversible commitment; more typically, tentative actions are possible, errors can be corrected, choices do not entirely close off other options, and what has been accomplished along one path might even be useful if one changes direction. In these latter cases (e.g., choosing a career, courting a potential mate, hiring a new employee, adopting a foreign policy), a full-scale analysis of all options, including the probability and desirability of each possible outcome, may be less successful than a more exploratory approach.

There is no guarantee, however, that an incremental approach will always be successful - in inference, choice, or planning. Incremental commitment has its own dangers, e.g., that irreversibility will creep in without being noticed (Brown, 1989a), that "good money will be thrown after bad," or that feedback will be ineffective (Einhorn, 1980). Nevertheless, the availability of successful incremental strategies in dynamic environments once again forces a revision in the notion of a decision-making "error". The failure to fully model a problem beforehand is not *per se* an error; what *is* an "error," though, is failing to incrementally improve one's understanding of relevant beliefs, preferences, and options as the problem evolves.

Variant (b): Change in the world. Variant (a) emphasizes the opportunity, in continuous problem environments, to learn about a world that is assumed to be fixed while we learn about it. Another important aspect of continuous environments is the possibility of change in the underlying processes that we are trying to learn about. The possibility of change increases the adaptiveness of strategies that do not immediately make full "normative" use of evidence, current goals, or available options and contingencies. Biases that appear to involve insufficient reaction to new evidence may serve the decision maker well in the face of possible real-world changes that affect the reliability of evidence and its significance. In choice, shiftable reference points and aspiration levels may help decision makers cope with changes in the real-world conditions that determine what can be achieved by an action (Hogarth, 1981). Contingency plans that try to anticipate every possible change tend to be either unmanageably complex or unrealistically over-simplified (Brown, 1989a); overplanning can suppress the variability that is necessary for learning and the ability to innovate if the unexpected occurs.

In all three cases, a strategy of tentative, incremental commitment, improvising in the face of the unexpected, may work better.

Variant (c): Long-term Consequences. Static laboratory studies preclude the observation of long-term consequences that may arise from real-world decisions. Anderson (1986) and others, for example, have noted the social advantages of overconfidence in one's ability to control events. Tribe (1971) has argued that explicit quantification of the probability of a defendant's guilt or innocence during a trial may eventually undermine society's confidence in the judicial process. In still other contexts, formal inconsistencies in judgments and decisions may involve processes of trial and error that help decision makers determine what cues are important and what strategies will work (Hogarth, 1981). Inconsistency may have other adaptive consequences, too: e.g., to promote unpredictability of one's own behavior in competitive situations.

II. ARE DECISION PROCESSES FLAWED BECAUSE THEY PRODUCE INCONSISTENT DECISIONS?

Challenge 3: The inconsistency of decisions is mitigated by the benefits of using the decision maker's real-world knowledge.

This challenge is more optimistic still: real-world task environments need no longer come to the rescue of flawed decision processes. Decision processes lead to adaptive decisions because they draw on decision-maker knowledge. There are (at least) four variants of this challenge: (a) People are more likely to be normatively consistent in tasks with which they are familiar or expert; (b) people are more likely to be normatively consistent when they have been explicitly trained in the use of appropriate general-purpose intellectual tools, such as decision analysis; (c) applying knowledge or expertise to a task is frequently associated with *non*-normative behavior, but the contribution of domain-specific knowledge to the quality of decisions offsets the effects of normative inconsistency; and (d) people do not use normative procedures because such procedures demand types of knowledge that people often do not have. According to (a) and (b), special-purpose or general-purpose knowledge, respectively, causes people to adopt normatively correct procedures; according to (c) and (d), the appropriate handling of knowledge and/or ignorance is what makes people adopt *non*-normative (but justified) procedures. All four variants, however, reject the idea of universal, pure decision processes, which operate in the same way in the real world and the laboratory, independent of what people know.

Variant (a): Domain-specific knowledge reduces biases. A number of researchers have supported the idea that decision-making knowledge is embodied in special-purpose packages, such as schemas, frames, or scripts (e.g., Minsky, 1975; Schank and Abelson, 1977). On this view, there is no general cognitive machinery that ensures consistency with respect to probability theory (or logic or any other normative standard). Rather, there are a large number of specialized structures which, most of the time, happen to produce normatively consistent performance. An experimental task that uses artificial materials, even if it seems formally identical to a corresponding real-life task, may fail to elicit an appropriate knowledge package (Simon and Hayes, 1976); hence, performance may be qualitatively different and, perhaps, defective. Even more strikingly, Ebbesen and Konecni (1980) showed how performance by experienced decision makers could be dramatically different in a simulated laboratory version of a task and in the real world.

Some support for the idea that normative consistency depends on domain-specific knowledge has come in the area of logical reasoning. Wason (1968) showed subjects four cards, which (they were told) had a letter on one side and a numeral on the other. Subjects were asked to test rules of the form, "If a card has a vowel on one side, then it has an even number on the other." For example, if the four cards showed an A, B, 4, and 7, logic, according to Wason, dictates that subjects select the cards with A and 7 - since rules of the form *If-p-then-q* are false only if *p* is true and *q* is false. Nevertheless, most subjects turned over the cards with A and 4. Wason interpreted this as a confirmation bias: choosing to collect information that can only confirm rather than disconfirm a hypothesis. The card showing 4 cannot disconfirm the rule regardless of what is on the other side; the card showing 7 (which the subjects neglected) could disconfirm the rule if its other side had an A. This bias, however, seemed to disappear when concrete materials were substituted for the meaningless letters and numbers; in a study by Johnson-Laird, Legrenzi, and Legrenzi (1972) subjects correctly tested rules such as, "If an envelop is addressed on one side, it must be stamped on the other." Cheng and Holyoak (1985) argued that valid performance does not depend on the familiarity of the stimuli as such, but on the ability of the subjects to apply learned relationships, called "pragmatic reasoning schemas," such as those associated with obligation, permission, or causality.

The notion that people are less "biased" in familiar tasks has been tested elsewhere, with mixed results. For example, an effect of using between-subjects designs is that individual subjects are never able to *become* familiar with a task. When studies on base rate neglect and on the effects of sample size were replicated using a within-subjects design instead of a between-subjects design, the bias was reduced, presumably because the salience of relevant variables was heightened for individual subjects who experienced all conditions (Fischhoff, Slovic and Lichtenstein, 1979; Birnbaum and Mellers, 1983; Leon and Anderson, 1974). Base rate neglect is also reduced when subjects experience actual events instead of being given statistical summaries (Nisbett, Borgida, Crandall, and Reed, 1976). May (1986) provides an empirical and theoretical analysis that attributes overconfidence in probability estimates to substantive ignorance regarding specific items, rather than to an abstract shortcoming in probabilistic reasoning. Shanteau (1989) summarizes studies with auditors that show non-existent or reduced effects of biases associated with representativeness, availability, and anchoring and adjustment.

There is other evidence, however, supporting the importance of knowledge about uncertainty handling itself, in addition to knowledge of the problem domain. For example, weather forecasters have been found to be well-calibrated in their probability estimates; but bankers, clinical psychologists, executives, and civil engineers did show overconfidence, despite their experience in their respective domains, presumably because of lack of training in probability judgment *per se* (cited in Fischhoff, 1982). Finally, as Evans (1989) points out, schema-theory simply does not account for some people's ability, at least some of the time, to reason correctly about abstract or unfamiliar material.

Variant (b): General-purpose knowledge reduces biases. As we have seen, domain-specific knowledge does not guarantee normatively correct performance. In any case, the real world contains unfamiliar as well as familiar tasks; Fischhoff (1982) argues that artificial laboratory studies in which biases occur are not necessarily unrepresentative of real-world novelty. Real-world decision makers, however, have another sort of knowledge available to them that laboratory subjects typically do not: the use of general-purpose "intellectual tools" (von Winterfeldt and Edwards, 1986). This view rescues human rationality by emphasizing human malleability: the heuristics and biases literature overlooks the ability to bring one's thinking into line with an appropriate tool.

Von Winterfeldt and Edwards regard the attempt to study pure "statistical intuitions," unassisted by the customary use of technical knowledge, books, calculators, notes, or other "external" aids, as misguided. They argue that a cognitive process is not a fixed method, but a "learned intellectual or judgmental skill executed with whatever tools seem necessary" (p. 554). The boundary of the skin is arbitrary: education moves information and processing from outside to inside; technology (paper and pencil, calculators, computers) moves them outside again. L. J. Cohen (1981) makes a related point: people cannot be condemned as "irrational" or "biased" when they fail to utilize principles that they have not been taught and whose discovery required mathematical sophistication and even genius. Shafer (1988; Shafer and Tversky, 1988) has argued that unaided human intuitions are not precise or definite enough to be regarded as either coherent or incoherent with respect to normative theories; rather, people learn how to *construct* precise and definite judgments by using normative theories . The metaphor of "intellectual tools" thus shifts the emphasis from human shortcomings to the natural ability to improve - culturally by inventing new tools, individually by learning how to use them, and on a particular occasion by using the tools to build a model of one's preferences and beliefs.

The tool metaphor is important, because it rescues the idea of decision aiding from paradoxes that are implicit in both the formal-empiricist point of view and in the rationalist point of view. If people's beliefs, preferences, and choices are already (pretty nearly) consistent with respect to normative standards, as the formal-empiricists supposed, then there is no need for decision aiding. On the other hand, if beliefs, preferences, and choices are inconsistent with respect to the normative theory, as the rationalists suppose, there is still no good rationale for decision aiding! Decision aids themselves depend on subjective inputs regarding probabilities and utilities; but if these inputs are subject to bias, then how can the conclusions of an aid be trusted? As far as normative decision theory is concerned, modifying the inputs is just as good a way of achieving consistency as adopting the conclusion.

The tool metaphor breaks out of this dilemma by rejecting the premise (accepted by both formal-empiricists and rationalists) that decision makers have pre-existing definite and precise beliefs and preferences about every relevant issue. Decision aiding is useful, then, simply because it helps the decision maker generate beliefs and preferences that she *isn't* sure about from beliefs and preferences that she *is* sure about. A successful normative model matches up with the pattern of a decision maker's knowledge and ignorance: it demands as inputs things the decision does know, and produces as outputs things the decision maker wants to know.

The tool metaphor, interestingly enough, undermines the *formal* justifications of decision *theory* offered by Savage (1954/1972), De Finetti (1937/1964), Lindley (1982), and others - since these all depend on the assumption of definite and precise beliefs and preferences about everything. But the tool metaphor substitutes something that might be better: a *cognitive* justification for *decision analysis.*

Unfortunately, the case for the cognitive plausibility of decision analytic procedures is less than overwhelming. Rationalist research on biases suggests that people do have strong intuitions about the solutions to problems that conflict with their own inputs to standard decision analytic models; moreover, people often fail to agree with the normative rule itself when it is explicitly presented to them (Kahneman and Tversky, 1982), and are unpersuaded by arguments in support of the normative rule (Slovic and Tversky, 1974). Finally, as we will see, decision makers do not utilize, and do not have precise knowledge about, many of the *inputs* required in decision analytic models. In the face of these difficulties, one can persist in hoping that intuition can be "educated" to become decision analytic (von Winterfeldt and

Edwards, 1986), or one can look for *other* tools in addition to decision analysis that may, at least for some users in some tasks, provide a better fit to their own particular patterns of knowledge and ignorance.

Some researchers have argued, ironically, that it is difficulty in using decision analysis, by contrast to more natural methods for handling uncertainty, that causes biases. Artificiality in the presentation of information about uncertainty, by means of numerical probabilities, may degrade performance. Zimmer (1983) argues that people ordinarily describe uncertainty verbally, in terms of such expressions as "highly probable," "likely," "quite possible," and so on, rather than with precise numerical probabilities. Zimmer applied a measurement technique for matching these expressions to ranges of numerical probabilities. His subjects turned out to be reasonably well-calibrated (not overconfident or underconfident) when allowed to use the verbal labels instead of numbers. Zimmer also claims that the "conservatism" bias, i.e., a failure to adequately update beliefs in the light of new evidence (Edwards, 1968), was reduced when subjects used verbal labels instead of numerical probabilities. In a study of the "regression fallacy," i.e., overextreme predictions of one quantity based on another, Zimmer asked subjects not only for a precise prediction (as in Kahneman and Tversky, 1973), but also for a verbal assessment of degree of confidence and the likely direction of error. Subjects showed considerable awareness of the pitfalls in the precise estimate: confidence was generally low, and almost all subjects were aware that the true values would probably be less extreme. Subjects' descriptions of their own reasoning (in a different experiment) suggested that verbal and numerical response modes prompted quite different problem-solving processes. Subjects using verbal labels took into account a wider range of qualitative variables than subjects using numbers (cf., Hammond, 1988).

In sum, general-purpose knowledge about decision making may sometimes reduce biases - whether because a decision maker has subjected his thinking to a technical discipline, or because the problem lends itself to problem-solving techniques that are already embedded in the language and the culture.

Variant (c): Effective handling of domain-specific knowledge causes biases. On another view, biases, instead of being eliminated by knowledge, are the by-products of domain-specific knowledge; they are caused by the knowledge structures (such as schemas) or cognitive processes (such as pattern matching) that people use to solve problems. An important implication of this view is that people might be unable to apply their knowledge effectively if they were to change their ways of thinking; performance might *suffer* on the whole if people were to adopt standard normative procedures in place of natural methods.

One process that is key to human problem solving (but neglected in standard normative theories) is pattern-matching or recognition. There is evidence that expertise in a variety of fields depends on the ability to recognize important features of the problem and to directly retrieve appropriate actions or solution techniques; in contrast, the more analytical approach of sophisticated novices requires explicitly generating and evaluating alternative methods for reaching a goal (e.g., Larkin et al., 1980). Chess masters, for example, may be distinguished from novices at least in part by their ability to recognize a very large number of familiar patterns, reducing the need to search through a tree of possible moves and countermoves (De Groot, 1965; Chase and Simon, 1973). Polya (1945), Newell (1981), Klein (1980), and Noble et al. (1989) have emphasized how new problems may be solved by recognizing their similarity to older, better-understood problems and by appropriately transforming the old solution to take account of differences. Experts, unlike novices, perceive similarities in terms of the fundamental laws or principles in a domain rather than in terms of superficial features (Chi et al., 1981). According to Lopes and Oden (in press) the virtues of pattern-based reasoning include robustness under conditions of noise or error, general applicability without stringent preconditions, and practicability with brain-like hardware (e.g., massive parallel processing versus step-by-step serial processing).

Pattern-based reasoning may provide an explanation of a broad range of biases. In the "conjunction fallacy" (Tversky and Kahneman, 1983), for example, people estimate the probability of two statements' both being true (e.g., "she is a feminist lawyer") as higher than one of the individual statements ("she is a lawyer"). Lopes and Oden (in press) speculate that the tendency to overestimate the probability of conjunctive classifications may be due to the improved match between conjunctive labels and experimentally provided descriptions of people (e.g, "feminist lawyer" is a better match than simply "lawyer" to the description of a woman who is "single, outspoken, and very bright," and who is "deeply concerned with issues of discrimination and social justice..."). Leddo, Abelson, and Gross (1984) accounted for overestimation of the probability of conjunctive explanations in terms of improved matches with schemas that specify the components expected in a good explanation (e.g., a statement of both the crime and the motive is judged more probable than a statement of the crime alone).

Pennington and Hastie (1988) have argued that jurors evaluate evidence by fitting it into a story that is constructed in accordance with explanatory schemas. The "belief bias," in which new evidence is interpreted to fit the currently held hypothesis, may reflect such a process. More generally, research in cognition suggests that prior expectations play a normal and important role in interpreting data. Expectations fill gaps and help organize experiences in perception (Bruner, 1957), recall (Bransford and Franks, 1971), everyday reasoning (Minsky, 1975), and science (Kuhn, 1962). Schemas, frames, scripts, and other knowledge structures permit successful action under conditions of incomplete and noisy data and limited time. The cost of these benefits is an occasional error when the schema is inappropriate.

Expectancy-based processes may also account for "distorted" concepts of randomness. When asked to produce random sequences, e.g., of heads and tails, subjects provide too many alternations and too few runs of all heads or all tails, in comparison to "truly random" (binomial) processes like tossing a coin (Wagenaar, 1972). When asked to estimate the probability of a particular sequence of "random" events, such as the birth order of girls and boys in a family, subjects overestimate the probabilities of sequences that contain equal numbers of each kind, and the probabilities of sequences which contain many alternations (Kahneman and Tversky, 1972). Lopes (1982) attributes such errors to powerful top-down processes that account for our ability to detect patterns against a background of noise.

An alleged bias in choice involves the assignment of outcomes to "psychological accounts" (Tversky and Kahneman, 1981). In one problem, some of the subjects were asked to imagine that they purchased a $10 ticket to a play, and on entering the theater discovered they have lost it; other subjects were asked to imagine that they decided to see the play, and on entering the theater to purchase a ticket found that they have lost a $10 bill. Subjects who lost the $10 bill were much more likely to purchase a ticket than subjects who lost their original ticket (worth $10). Nevertheless, the "total asset positions" that would result from buying the ticket is the same for the two sets of subjects, and that is what should, according to normative theory, determine the decision. According to Tversky and Kahneman, however, people do not lump all the consequences of their actions into a single pool. When they lost a ticket, subjects included the cost of the original ticket in the "account" containing the new ticket price, raising the psychologically perceived price of the show to $20; on the other hand, when $10 was lost, they regarded it as irrelevant to the cost of the ticket. Psychological accounts appear to group outcomes that belong together according to causal schemas or goal-oriented "scripts." Such groupings may facilitate learning significant correlations between one's actions and the events that they cause, and also help people keep track of different attitudes toward risk in different psychological accounts, e.g., avoiding risks in accounts that pertain to major investments, while seeking risks within accounts allocated for risky investments, vacations, luxury items, etc. (von Winterfeldt and Edwards, 1986).

The reference effect involves an effect on choice due merely to changes in the way actions or outcomes are described; e.g., description in terms of "chance of survival" may lead to risk aversion, whereas description in terms of "chance of death" may lead to risk seeking. Such effects might occur because different action or outcome descriptions match different internal knowledge structures, activating different arguments for or against an action (Shafer, 1988).

More generally, and perhaps more importantly, the "heuristics" cited by Kahneman and Tversky to account for decision biases may be better understood as integral parts of schema-based processing, rather than as isolated and somewhat *ad hoc* explanatory mechanisms. The availability heuristic corresponds to the retrievability of a schema; representativeness corresponds to the degree of similarity between the current situation and a schema; and anchoring and adjustment involve transformations of the solution associated with a schema to accommodate a mismatch with the current situation.

Variant (d): Effective handling of domain-specific ignorance causes biases. According to variant (c), normative models might interfere with the effective use of knowledge. The other side of the coin is that normative models may interfere with effective handling of *ignorance*; in effect, they force the decision maker to pretend that she knows things that she does not know. Presupposing the existence of precise probabilities and preferences, as required in standard normative theories, may prematurely close off the possibility of *learning* about one's beliefs and preferences in dynamic environments, through subsequent experience or reflection (March, 1988; Levi, 1986). But even when decisions must be made on the spot, decision makers may be more successful when they do not adopt a false precision.

Decision makers may have little or no knowledge from which to assess the scores, weights, and probabilities required by decision analytic models for the integration of different evaluative dimensions and uncertain outcomes. Frequently they adopt strategies that require much less precise information (Svenson, 1979; Tyszka, 1981): for example, satisficing requires only yes-or-no judgments about alternatives on each dimension (e.g., does the alternative achieve a goal or not?) and no comparisons at all among different dimensions; elimination-by-aspects merely adds a requirement for rank ordering dimensions by importance; the lexicographic decision rule (i.e., pick the best candidate on the most important dimension; if there is a tie, go to the next most important dimension, etc.) adds a requirement for rank ordering alternatives on a given dimension, but remains far less demanding than decision analytic modeling. Similarly, in the regression fallacy, when decision makers provide overly extreme predictions of one quantity based on another, they may simply not know enough to estimate the degree of correlation between the known variable and the variable to be predicted (especially if there is a possibility of change in the underlying process). In the belief bias, decision makers may feel unsure of the reliability of a source of apparently disconfirming evidence, and thus discount it. In base rate neglect, decision makers may be unsure of the reliability of frequency data, and thus disregard it.

The inputs required by a decision analysis often do not correspond to what a decision maker knows with confidence. But is the *output* of a decision analysis worth the effort? Sometimes, at least, it seems not. The end result of a decision analysis is usually an "average" hypothesis, outcome, or preference, which has little meaning in terms of what the decision maker needs to do or know (Cohen at al., 1988). He can avoid, plan for, and/or react to specific situations or outcomes, not an unrealizable average. In planning a defense, for example, it is useful for a general to know that the main enemy force might be planning to attack at point A or else at point B; he may try to predict how well various defensive options will fare in each of those cases. But he will have little use for a prediction of enemy intent in terms of the probability-weighted average future enemy force at each location, or for an evaluative score reflecting a defensive plan's success averaged across the two situations. It is plausible to speculate that there is a "basic level" of description, neither too detailed nor too general, that is most usefully linked to the rest of a decision maker's knowledge in a particular task (Rosch et al., 1976). Simon (1972) observed that normative models of chess, which attempt to summarize all the implications of a move with a single abstract evaluative measure, are less successful than non-normative models, which ignore some possible outcomes and explore a small number of well-chosen paths to an appropriate depth.

In sum, variant (d) implies that decision analysis is, on at least some occasions, not a "good intellectual tool": it fails to match the pattern of what a decision maker knows and needs to know. Variant (c) - that non-normative behavior flows from schema-based processing - underscores this conclusion, implying that decision analysis conflicts with the way people ordinarily use what they know in order to solve problems and make decisions. All four variants of this challenge imply a redefinition of the notion of a decision making "error": not in terms of logical inconsistency, but in terms of the failure to effectively exploit one's knowledge in the service of one's needs.

We may, nevertheless, be somewhat uncomfortable with a picture of "natural" reasoning that is exclusively focussed on knowledge. Such a view may fall short in accounting for the flexibility that decision makers sometimes display in novel situations. For example, while experts may "recognize" familiar problems, recognition itself is not simple: it may incorporate a series of transformations and retransformations of the problem until the expert finally "knows" how to solve it. Physics experts, according to Larkin (1977), first sketch the superficial objects and relations in a problem; if the depicted system is still not familiar, they may transform it into an idealized, free-body diagram; if recognition still does not occur, they may switch to a more novice-like strategy of means-ends analysis. Once they have solved a problem, physics experts draw on a variety of strategies to verify its correctness, e.g., by checking whether all forces are balanced, whether all entities in the diagram are related to givens in the problem, etc. There is abundant evidence that recognitional processes are not inflexibly automatic, but involve a fairly continuous stream of optional processes that evaluate and guide the application of pre-stored knowledge.

Challenge 4: Constraints on the decision maker's information-processing capacity justify use of non-Bayesian procedures.

Challenge (4) brings flexibility front and center. It adopts an optimistic stance toward human rationality without appeal to factors (such as intellectual tools, knowledge, or incremental commitment and feedback) that operate primarily in real-world settings. It implies that a wider understanding even of the laboratory context helps make normative sense of "biased" performance: Herbert Simon (1955, 1972) argued that strategies such as satisficing that might seem irrational in the absence of information processing constraints are perfectly sensible given the presence of such constraints. According to the effort/accuracy tradeoff hypothesis, adoption of suboptimal strategies may itself be justified at a second-order level, given the cognitive demands of calculating an optimal solution and the "almost as good" quality of simplifying strategies.

Payne (1976) carried Simon's idea one step further: individuals appear to utilize not one, but a variety of simplifying strategies in response to varying task characteristics. Beach and Mitchell (1978) proposed that decision strategies are selected on the basis of a cost-benefit calculation that balances the demand for accuracy in a particular task against the cost of being accurate in that task. Payne, Bettman, and Johnson (1989) have shown (through an analytical model of effort and a Monte Carlo simulation of accuracy) how different choice strategies might in fact trade off in terms of accuracy and effort under different task conditions. No single heuristic does well across all decision environments; but a decision maker can maintain a reasonably high level of accuracy at a low level of effort by selecting from a repertoire of strategies contingent upon situational demands (Payne, Bettman, and Johnson, 1989).

Experimental data suggest that people do in fact adaptively adjust their processing strategies in response to changes in such variables as the number of options (Payne, 1976), or the variance among probabilities and importance weights (Payne, Bettman, and Johnson, 1989). A number of studies have shown that time stress causes selective focussing on negative attributes/outcomes (Leddo, Chinnis, Cohen, and Marvin, 1987; Wright, 1974) and adoption of alternatives that hedge against the worst case (Leddo, Chinnis, Cohen, and Marvin, 1987; Ben Zur and Breznitz, 1981). Display features can also make some strategies harder and others easier (Hammond, 1988; Tyszka, 1980). For example, when information about alternatives is presented numerically, subjects are more likely to compare alternatives directly to one another (as in the lexicographic rule); but when less precise verbal descriptions of alternatives are given, alternatives are compared to a goal, as in elimination-by-aspects (Huber, 1980).

Even in applications where comparable research on effort/accuracy tradeoffs has not been done, it is clear that adaptation to constraints on capacity is necessary. For example, as noted above, some choice biases involve assigning outcomes to different "psychological accounts," based on causal or goal relationships, rather than considering the decision maker's total asset position (Tversky and Kahneman, 1981). If limited capacity is taken into account, we might ask: how could it be otherwise? Changes in one's total asset position may be occurring continuously - e.g., paychecks, retirement accumulations, appreciation of one's home and investments, depreciation of one's car, changes in inheritance prospects, and so forth; it would hardly be worth the effort to try to model the impact of all these events on every decision (Von Winterfeldt and Edwards, 1986).

Biases in inference and probability assessment may also be adaptive responses to capacity limitations. For example, the belief bias could result in principle from the impossibility of questioning every belief in the light of each new experience: some beliefs (e.g., the currently active "schema") must be left unquestioned in order to evaluate others (Quine, 1960). With even a small number of beliefs, examination of all combinations of their truth and falsity rapidly becomes impossible in principle (Cherniak, 1986). The same problem would explain some cases of "base rate neglect." For example, in fitting causal models to correlational data, the number of possible causal arrangements grows very rapidly with the number of variables (Glymour et al., 1987) (e.g., the factors that might cause cancer, such as smoking, diet, geographical location, health care, etc., might also have causal effects on one another). It is impossible to assess base rates for all the possible causal arrangements; and the assessment of a catch-all hypothesis (e.g., "everything else") is hardly satisfactory if we do not know what hypotheses it contains. As a result, decision makers (including scientists) must "satisfice," i.e., look for satisfactory rather than optimal models. In a continuous environment inference may become more like *design*: instead of choosing among a pre-given set of hypotheses, an original hypothesis or small set of hypotheses will be revised and/or elaborated incrementally as shortcomings are discovered.

The effort/accuracy theory introduces still another notion of decision "error": the failure to appropriately balance accuracy and effort in the choice of decision strategies. Formal inconsistency *per se* is not an error in this sense; adopting a highly inaccurate strategy when only a small amount of effort would be required to adopt a more accurate strategy, may sometimes be an error. But then again, in other contexts, with other payoffs, it might be an error to expend more effort in order to improve accuracy.

There is a curious disconnection between the effort/accuracy tradeoff hypothesis (challenge 4) and schema-based views of problem solving that emphasize the role of substantive knowledge (challenge 3). As noted above, there is evidence that expertise consists, at least in part, in the ability to recognize a large store of situations and to retrieve appropriate solutions (e.g., Larkin, 1980; Chase and Simon, 1973). By contrast, the effort/accuracy tradeoff model emphasizes rationality at the level of "metacognitive" or higher-order decisions about how to decide. Schema-based approaches seem to emphasize automatic activation of relevant knowledge structures, leaving little room for conscious monitoring and control (e.g., Anderson, 1982); while the effort/accuracy model defines both of its central concepts (effort and accuracy) without reference to knowledge.

We think that each approach needs to be supplemented by concepts from the other. Experts are skilled not only in recognition, but in metacognitive processes that enhance the *likelihood* of recognition and that verify, critique, modify, and/or abandon the results. The primary function of metacognitive processes is to control the application of knowledge, not to choose among knowledge-independent analytical strategies.

A Synthesis: The Interaction of Recognition and Metacognition.

From the naturalistic point of view, several aspects of the effort/accuracy approach to biases bear questioning: its commitment to a measure of effort that ignores how much the person *knows* about a problem; its use of decision analytic strategies as the standard for evaluating accuracy; and its emphasis on conscious higher-order selection of strategies as opposed to local choices about what to do/think next. A Recognitition/Metacognition model revises each of these features:

(a) A notion of effort that incorporates knowledge. Payne et al. (1988) have measured effort abstractly, in terms of the number of generic EIP's (elementary information processes such as *read, compare, add)* required by a strategy; they have validated the model in laboratory studies in which tasks do not involve expertise or experience, and in which all relevant information is pre-digested (e.g., as probabilities and payoffs) rather than inferred or generated from the decision maker's own memory. This model incorporates individual differences in *general-purpose* ability to perform EIP's and in

ability to combine them into strategies (cf., Beach and Mitchell, 1978). In a more naturalistic setting, however, the decision maker's estimate of the difficulty of a strategy will reflect what she believes about her own knowledge of a specific task and of the relevant domain. The decision maker chooses to deploy attention selectively to certain aspects of the problem, to mentally recode or physically transform certain problem materials, and to selectively "rehearse" some of the materials rather than others - because she believes that those aspects or materials are more likely to activate knowledge that will activate other knowledge, and so on, until she arrives at a solution. EIP's for retrieving and transforming knowledge must be incorporated into the theory of mental effort.

The overconfidence bias in estimating the probability of a conclusion might result from metacognitive choices that reflect the effort required to retrieve or generate ways that the conclusion could be false (Pitz, 1974). The overconfidence bias is reduced when subjects are explicitly asked to generate reasons why the conclusion they favor might be wrong (Koriat, Lichtenstein, and Fischhoff, 1980; Hoch, 1985); as new reasons are demanded, subjects exert more effort, selectively activating new knowledge structures, and questioning increasingly fundamental premises of the original conclusion (Cohen, 1990). Similarly, metacognitive choices may underlie the top-down processing that is characteristic of the belief bias, in which apparently conflicting data are perceived as supporting a favored hypothesis. More effort would be required to activate and test alternative explanatory schemas. The optional (or metacognitive) character of the belief bias is suggested by Tolcott and Marvin's (1988) finding that simply briefing subjects on the existence of the bias reduced its effect. More recently, Tolcott and his colleagues found that subjects who were required to actively select evidence bearing on a hypothesis were less likely to interpret conflicting evidence as confirming, than subjects who had evidence passively presented to them; the former subjects may have been induced to attach a higher value to truly testing the hypothesis.

Other biases may involve a similar metacognitive balance between effort and accuracy. Reference effects in choice may reflect the difficulty of accessing alternative ways of describing outcomes when one way of describing them is strongly activated by the wording of the problem. The availability bias, by its very definition, refers to ease of recall of instances as a determinant of probability estimates for a class (Tversky and Kahneman, 1973; Beyth-Marom and Fischhoff, 1977). The hindsight bias might result from the effort that would be required in tracking and undoing the effects of a past event on all the relevant components of a person's knowledge (Fischhoff, 1982).

Different choice strategies may also reflect different metacognitive choices about the most efficient access to relevant knowledge. In one strategy, called dominance structuring (Montgomery, 1983), the decision maker starts by provisionally selecting an option; she then works backward, adding and dropping attributes, revising scores, etc., in an effort to show that the selected candidate is as good as or better than other candidates in all respects; if she fails, she selects another option, and so on. Such a strategy may be quite easy when decision makers have precompiled, intuitive "knowledge of what to do," but have less direct access to knowledge about the attributes that justify such a choice. By the same token, elimination-by-aspects would be easier in domains where knowledge is organized by goals and the means of achieving them, and satisficing would be easier in domains where knowledge is organized by options and what they are good for. Even compensatory strategies, which require comparisons across different dimensions, may be easier when decision makers have readily accessible knowledge upon which they can base judgments of the relative importance of criteria; for example, a military commander might evaluate the cost of losing one of his own units compared to the value of destroying an enemy unit in terms of the relative numbers of the two forces known to be present in a battle area (Cohen, Bromage, Payne, and Ulvila, 1982).

(b) Replacement of decision theory by dynamic adjustment as a benchmark for performance. According to the effort/accuracy model, decision analytic procedures are the ideal from which decision makers deviate under high workload. But decision makers do not necessarily adopt decision analytically correct methods even when workload demands are low (e.g., Cohen, Leddo, and Tolcott, 1988); and decision makers often reject normative rules and arguments for those rules when they are explicitly presented (Kahneman and Tversky, 1982; Slovic and Tversky, 1974). Kahneman and Tversky have cited such results as support for the claim that biases are deeply rooted in our cognitive systems, on the analogy of perceptual illusions. An alternative, naturalistic view is possible, however: under conditions of low workload, decision makers might adopt more effective variants of "non-optimal" strategies. Increased effectiveness may result from iterative improvements in a dynamic environment.

Decision makers might deal with tradeoffs among evaluative dimensions, for example, by adopting a more dynamic and self-critical variant of satisficing or elimination-by-aspects. In these variants, a decision maker starts out with relatively high aspirations on all dimensions; if all goals cannot be achieved, aspirations on particular attributes might be revised downward, in small steps, to accommodate options that just miss a goal but are outstanding in other respects. Such a metacognitive process would accommodate the compensatory relations that are relevant for the problem at hand, without requiring the explicit assessment of a large number of precise weights that are *not* relevant (Cohen, Bromage, Chinnis, Payne, and Ulvila, 1982; Cohen, Laskey, and Tolcott, 1987).

In the same way, more sophisticated variants of non-optimal inference strategies might be adopted in low-stress conditions. In the belief bias, people seem to use an existing hypothesis to interpret new evidence in a top-down manner, producing a more definitive picture of the situation, in contrast to continuously shifting Bayesian probabilities. The natural improvement of this strategy might be to make it dynamic and self-critical: to keep track of the evidence that has been provisionally "explained away," or to maintain a cumulative assessment of the degree of doubt in the current conclusion, and to initiate search for an alternative hypothesis when the amount of explained-away evidence, or the degree of cumulative doubt, reaches a high enough level (Cohen, 1989, 1990). This strategy allows more effective use of decision-maker knowledge in building a coherent explanation/prediction of events; at the same time, it guards against being locked into seriously mistaken conclusions.

Ironically, a Recognition/Metacognition framework has less trouble than the effort/accuracy hypothesis in accounting for cases where people *do* use decision-theoretically optimal procedures. Normatively correct behavior might be *easy* (rather than effortful) when readily activated knowledge structures in a particular domain happen to fit normative rules (Cheng and Holyoake, 1985), or when the decision maker is well-versed in general-purpose normative techniques (Von Winterfeldt and Edwards, 1986).

(c) Local choices rather than deliberative higher-order selection of strategies. The effort/accuracy hypothesis has typically assumed that strategies are selected by a top-down, conscious process, in which normative constraints are satisfied. By contrast, Simon (1972) rejected the notion of a second-order level of decision making that normatively derives heuristics, on the grounds that it would require difficult assessments beyond those needed simply to carry out the heuristic itself; the effort involved would be better devoted to improving the knowledge that is directly exploited in the

heuristic. There is also controversy among researchers in the related area of "metacognition" regarding the degree to which higher-order regulative processes involve conscious awareness: whether they are relatively automatic (Sternberg, 1982), involve awareness of the first-order cognitive events being regulated (Gaveleck and Raphael, 1985), or involve awareness of both first-level and higher-level processes (Kuhn, Amsel, and O'Loughlin, 1988). Think-aloud protocols suggest that top-down selection of decision strategies, based on features of the task, does sometimes occur (Payne, Bettman, and Johnson, 1989). There is also evidence in the problem-solving literature that experts are better than novices at assessing the difficulty of a task (Chi, Glaser, and Rees, 1982). Nevertheless, Payne, Bettman, and Johnson (1989) have themselves recently suggested that in some cases, decision strategies may be "constructed" step-by-step in the course of the decision maker's interaction with a problem, rather than explicitly selected (cf., Connolly and Wagner, 1988). In such an incremental, iterative process, decision makers would utilize feedback from previous cognitive actions to make local decisions about what to do next.

Metacognition does not mean that people are conscious of all the knowledge that makes them effective or expert. It does suggest that people have higher level schemas (which may themselves sometimes be domain-specific and largely automatic) that gauge the familiarity and difficulty of problems or subproblems, and that incorporate responses (a) to enhance the chance of recognition and (b) to control the process of validating a potential problem solution. The metacognitive processes embodied in these schemas are governed by an implicit balancing of effort against expected results, in a way that takes account of such factors as the available time for a decision, the likelihood of errors, the stakes of the decision, the opportunity for feedback and midcourse corrections, and the structure of relevant knowledge representations.

In a naturalistic version of the effort/accuracy hypothesis, "heuristics" may sometimes be both *less* effortful and *more* accurate than "normative" models. If knowledge influences both effort and accuracy, then reasonably efficient decision-making strategies might sometimes emerge simply by "doing what comes to mind". Tradeoffs between effort and accuracy arise in more novel situations, where effective decision makers must be skilled in selecting the parts of their knowledge to be explored, monitoring progress toward a solution, recalling relatively inaccessible parts of their knowledge, and making revisions in beliefs and strategies where necessary (Larkin, 1981; Glaser, 1989; Brown and DeLoache, 1978).

Adaptation: Imperfect, but Important.

Challenges (1), (2), (3), and (4) all depend at bottom on the notion of adaptation - to internal capacity constraints, to the requirements of applying knowledge, to dynamic and rich task environments, and to the performance of a decision strategy across the overall spectrum of tasks in a domain. Deviations from formal consistency may turn out to be adaptive or at least neutral in these contexts, if the benefits associated with them outweigh the harm they do.

The challenges differ in *how* the adaptation is supposed to take place. Consider a simple analogy: both squirrels and humans put away valuable objects for safekeeping. On some level the function, or adaptive consequence, of burying nuts and putting jewelry in a safe (or money in the bank, etc.) is the same. Yet the squirrel's behavior is controlled by a genetically inherited program. He will try to bury nuts even in situations where the behavior does not have any adaptive consequences, e.g., in an enclosure with a concrete floor, or where no competitors for nuts exist. Humans will vary their "squirreling away" behavior (up to a point) to fit their individual circumstances.

Similarly, biases may reflect relatively coarse-grained adaptations: i.e., fixed cognitive traits that do reasonably well across all tasks that humans encounter, and which do not change from task to task. Challenges 1 and 2 require no more than this. For example, if overconfidence is an inherited or culturally conditioned trait with social advantages (Anderson, 1986), we would not expect an individual's overconfidence to be reduced in occasional situations where it is no longer socially advantageous (e.g., in a psychology laboratory). (However, if the environment of the species or the culture were consistently changed, then overconfidence might eventually disappear, over a much longer time period.) Challenges 3 and 4 demand a more fine-grained, flexible adaptation: biases reflect strategies that change in response to features of a task and decision maker, i.e., the familiarity of the task and the effort it demands. Finally, as we shall see in the next section, challenges 5 and 6 address an even more fine-grained and flexible adaptiveness, in which performance changes in response to detailed beliefs and preferences regarding the specific task.

It is obvious that having adaptive consequences, at whatever level, does not mean that a characteristic of decision making is "optimal": there may always be a more fine-grained level and a more refined adaptation. Even *at their own level*, adaptations may not be the optimal solutions. Natural selection, like cultural evolution and like human decision makers, is a satisficer, not an optimizer, and there is always likely to be room for improvement.

The naturalistic challenges, while rejecting rationalist pessimism, have left considerable leeway for error. Indeed, they make *more* sense of the notion of "decision error" than either the formal-empiricist framework (in which it is not clear when performance is in error and when the model is wrong) or the rationalist framework (according to which error is simply formal inconsistency). Challenge (1) introduced the notion of error as a failure to guard against specifically identified pitfalls in a suboptimal strategy. Challenge (2) emphasized the potential failure to respond to feedback or to make midcourse corrections in order to incrementally improve performance in dynamic environments. Challenge (3) stressed the failure to effectively exploit one's own knowledge. And challenge (4) highlighted the failure to appropriately weigh effort against accuracy in the selection of strategies. The Recognition/Metacognition framework incorporates all four kinds of error: biases represent a failure of metacognitive processes that facilitate problem recognition and retrieval of appropriate solutions, that monitor for potential problems in a decision process, and that verify and revise proposed solutions. The point, then, is not to paint unaided decision makers in a rosy glow. The argument is simply this: decision-making errors are better understood against a pattern of generally successful adaptation to real-world contexts, rather than as deviations from a largely irrelevant abstract standard.

It has been argued, however, (e.g., Einhorn and Hogarth, 1981) that claims about adaptation are scientifically unacceptable. These critics argue (a) that such claims are "unfalsifiable," since the *post-hoc* invention of adaptive consequences is all too easy; and (b) that failure to adapt is consistent with natural selection through such phenomena as "genetic drift," persistence after ceasing to be adaptive, and "hitch-hiking" of some traits upon the adaptiveness of others (e.g., Gould and Lewontin, 1979). I think these complaints are misguided.

Adaptive claims would be unscientific if they implied that the function or adaptive purpose of a characteristic is, by definition, whatever consequence it happens to have. (In that case decision strategies would be trivially adaptive, since they have consequences.) The hypothesis of adaptation requires, at a minimum, that a characteristic, such as using a particular decision strategy in certain sorts of tasks, exists *because* of the adaptive consequences it had in the past. This usually has the following testable implication: if conditions are changed so that the consequence no longer occurs, then the characteristic will change (over the appropriate time period, which may involve the species, the culture, or the individual). Empirical or theoretical research can support specific claims of this sort regarding the adaptiveness of specific characteristics. Some work has been done, e.g., changes in decision-making strategies due to task features that measurably affect effort (Payne, 1976), or due to different degrees of familiarity with a task (Larkin, 1977); more work is obviously needed.

The truly insupportable claims (as noted by Dawkins, 1983) are the negative propositions that a decision-making characteristic has *no* important adaptive function, or that there is *no* decision strategy more effective than (the current variant of) decision analysis.

What is most misleading is the suggestion (e.g., in Einhorn and Hogarth, 1981) that the burden of proof is on the adaptationist. On the contrary: the assumption of adaptation has had enormous heuristic value in the life sciences. The relevance of natural selection is not diminished just because evolution *can* produce non-adaptive traits; on the whole, evolution is adaptive, though there is lots of "noise" over short time periods (Dennett, 1983). To my knowledge, no case has *ever* been made in biology for the heuristic value of assuming dysfunctionality.

Tversky and Kahneman (1974), however, argue that the study of errors (in the sense of formal inconsistency) can shed light on information-processing mechanisms, by analogy to the study of perceptual illusions. But the analogy has been criticized on at least two grounds: standards of normative correctness are far less well understood in decision making than in perception, so it is not so clear what counts as an "error" in this sense (Smithson, 1988); and in any case, the "correct" process does not play a causal role in decision making the way the physical stimulus does in perception, where mechanisms can be thought of as transforming (or distorting) the stimulus (Shanteau, 1989; Anderson, 1986). Demonstrations of "error" in the rationalist paradigm show that a particular mathematical model does not fit performance, and (from an explanatory point of view) that is about all. No light is shed on cognitive mechanisms.

Perhaps an alternative is to disregard the question of functionality altogether, and to focus directly on mechanisms. But that would be comparable to studying the eye without any idea of what it was used for. Where would one even begin? What aspects would be worthy of attention? When would an explanation be complete? What would it mean to improve its performance? We think that the study of decision processes without reference to problem domains, task environments, knowledge, or capacity, is a dead end.

From the practical point of view, two kinds of errors are possible: thinking that adaptiveness is there when it isn't, and missing it when it is there. Eliminating an adaptive behavior (through training or "aiding") may be every bit as bad as letting a defective behavior go uncorrected. Moreover, when decision-making behavior does go wrong, corrective steps are more likely to be successful if they take its normal function into account. We turn in the next section to "normative" theories and what they may have to say about critiquing or improving the decision-making process.

III. ARE DECISIONS FORMALLY INCONSISTENT?

Challenge 5: There are alternative Bayesian models that better capture the subject's understanding of the problem.

Rationalist experimenters use decision analytic or logical models to evaluate their subjects' decisions; and thus far, for the sake of argument, we have accepted the validity of these normative standards. The use of a "suboptimal" decision strategy might be justified if it leads to satisfactory outcomes, even if the strategy itself is not inherently correct (challenges 1 and 2); or a "suboptimal" strategy might be justified because it effectively exploits knowledge under conditions of limited capacity (challenges 3 and 4). From this point of view, alternative, naturalistic concepts of decision error *coexist* with the rationalist concept of error as formal inconsistency. The previous challenges claimed that error in the rationalist sense is often outweighed, in the evaluation of performance, by the *absence* of error in the naturalistic sense (i.e, decision makers successfully compensate for specific weaknesses in decision strategies; they incrementally improve their knowledge of the problem; they efficiently use knowledge and capacity). Challenges (5) and (6), however, go farther: they attack the idea that decision processes are biased even in the narrow, rationalistic sense.

Challenge (5) accepts Bayesian decision analysis as a normative standard, but argues that in many cases of alleged biases, it has not in fact been violated. Variants of this challenge are something of a grab bag: they include (a) subjects' misconstruing the instructions of the experimenter and the experimenter's misunderstanding the knowledge and goals of the subjects, as well as (b) more serious errors by the experimenter in modeling the problem. These challenges have the flavor of the formal-empiricist paradigm: deviations of strong intuitions from a decision analytic model are taken as causes for concern about the decision analytic model rather than signs of irrationality by decision makers (e.g., Bell, 1981). To the extent that these arguments are convincing, decision analysis and ordinary decision-making performance may be reconciled. We think, in fact, that the challenges are not always fully convincing: the formal constraints of decision theory are too restrictive; as a result, formally adequate decision analytic models turn out not to be cognitively plausible. It is perhaps more important, however, that these challenges illustrate the dynamic character of "normative" modeling, and point the way to a more interactive process of decision aiding that does not so much dictate to decision makers as negotiate with them.

(a) Subject/experimenter misunderstanding. The normative force of decision theory is not to tell a decision maker what to believe, value, or do. Rather, it indicates when beliefs, preferences, and choices are inconsistent with one another. Consistency itself is relative to a selected model or structure; for example, if a subject (acting as decision maker) perceives a dependence between judgments that the experimenter regards as independent, or if she values attributes of an option that the experimenter has ignored, then apparently inconsistent behavior may in fact be quite rational. It is strange, then, that decision making research in the rationalist tradition has only rarely sought direct evidence of the way subjects/decision-makers represent a problem. Process-tracing methodologies, e.g., in which subjects think out loud as they work a problem, have been used only rarely (e.g., Svenson, 1979; Payne, 1976; Scholz, 1987); subjects are seldom asked at the conclusion of a study why they answered as they did; finally, there is little or no systematic study of the answers in their own right in order to get a better picture of the underlying processes. The only assurance that subjects and experimenters share an understanding of the problem is the instructions, and this is a frail reed. As Berkeley and Humphreys (1982) argue, there is no way that a written description of a problem can remove all ambiguity. The contrast with applied decision analysis is instructive: numerous iterations and extensive interaction between analyst and client are required, before they

can mutually agree on an appropriate model (e.g., Phillips, 1982).

Instructions may in fact cause "biases," as Kahneman and Tversky (1982) acknowledge. In the real world, things are said for a reason; in an experimental context, therefore, subjects may naturally draw inferences from the verbal description of a problem that they would not draw if the "same problem" were actually experienced. One convention that governs ordinary conversation (Grice, 1975), for example, allows the listener to assume that the speaker is trying to be relevant. As noted by Kahneman and Tversky (1982), this makes it difficult to study how subjects handle information that is supposed to be "irrelevant." For example, the finding that subjects interpret supposedly neutral cues in accordance with their favored hypothesis (an effect of the "belief bias") may, in part at least, be due to the assumption that the cue would not have been presented if it did not have some bearing on the question at hand. A similar complaint could be raised about studies in which "irrelevant" individuating evidence causes subjects to ignore base rates (Kahneman and Tversky, 1973). Some instances of "overconfidence," in which subjects are given an anchor, may also reflect reasoning of this kind. One group of subjects were asked the probability that the population of Turkey was greater than 5 million; another group was asked the probability that it was less than 65 million; when both groups were subsequently asked for their best guess as to the population of Turkey, the median estimates were 17 million and 35 million for the low and high anchor groups respectively. This may reflect a legitimate assumption that the phrasing of the question is itself evidence, rather than a bias caused by "insufficient adjustment" of an anchor. If subjects are uncertain of their answers in the first place, they would be foolish not to utilize such information (Macdonald, 1986).

Kahneman and Tversky (1982) appear to defend the allegation of bias even in these examples; they cite an experiment in which an anchor that was randomly chosen (by the spin of a roulette wheel) in the presence of the subjects also influenced estimates, despite the fact that no subject "could reasonably believe that (the anchor) conveys information." But it is far from clear that subjects accept assurances from experimenters regarding "randomness." In some cases, it may be quite reasonable for subjects to assume that the experimenter is lying. Gigerenzer and Murray (1987) cite base rate neglect experiments by Kahneman and Tversky (1973), in which subjects are told that personality sketches are selected from a population of engineers and lawyers at random. In fact, of course, the descriptions were deliberately constructed to match stereotypes of doctors and lawyers. If subjects suspect this (and, of course, they should), they will, as good Bayesians, ignore base rates.

The credibility of many experiments may be undermined by the unrealistic precision with which information is provided. For example, a well-known reference effect experiment (Tversky and Kahneman, 1981) asks subjects to choose between two medical options for fighting a disease that is expected to kill 600 people. In one condition the choice is between program A, in which 200 people will be saved for sure, and program B, which has a 1/3 chance of saving 600 people and a 2/3 chance of saving none. In the other condition, the choice is between program C, in which 400 people will certainly die, and program D, which has a 1/3 probability of no one dying and a 2/3 probability that 600 people will die. The choices in the two conditions are identical, but outcomes are described in programs A and B in terms of gains (saving people), and outcomes are described by C and D in terms of losses (people dying). Despite this identity, people tend to choose program A in the first condition and program D in the second. As Smithson (1989) points out, however, it is highly implausible to predict exact numbers of deaths in this kind of forecast; subjects may thus read unintended ambiguity into the outcome predictions. Berkeley and Humphreys (1982) argue that the description of program A suggests that *at least* 200 people will be saved, and that actions may be discovered that could result in saving more; the description of program C suggests that at least 400 people will die, but possibly more. Under this interpretation, A and C are not identical *to the subjects*. A similar interpretation of ambiguous outcomes may affect experiments on "psychological accounts." Subjects who lose a $10 ticket may be less likely to buy a new ticket than subjects who lose a $10 bill because a variety of subsequent options are relevant (and ethically appropriate) in the case of the lost ticket: e.g., they might instead try to convince a box office clerk to replace the ticket or an usher to seat them (Berkeley and Humphreys, 1982).

In all these examples, the subjects' rationality is rescued by more complex decision analytic structures: taking the wording of instructions as evidence for the correct answer, taking frequency data as imperfect evidence for the true base rate (conditional on random selection), and elaborating a decision tree with subsequent acts. To varying degrees, these are somewhat *ad hoc* adjustments; the problems that they address have the flavor of "experimental artifacts," which might be corrected simply by more carefully controlled experimental procedures (e.g., Fischhoff, 1982). The more important underlying lessons, however, concern the importance of understanding the way subjects represent the problem and of studying decision making in contexts that are sufficiently similar to the real world that subjects know *how* to represent them.

(b) Experimenter mis-modeling. Bias findings may also be inconclusive due to deeper and more general errors in decision analytic modeling. A number of criticisms of the bias literature take on a quite formal and technical character. But they have in common with the issues examined above an emphasis on factors that may be important to subjects but overlooked by experimenters, and which can be incorporated into ever more complex decision analytic models.

Discrepancies between experimenters and subjects in the perceived structure of a problem may occur because of differences in goals. Bell (1981, 1988) has developed a revision of expected utility theory that incorporates the feelings that a decision maker might expect to have after making a choice under uncertainty and discovering the actual outcome. The decision maker feels "regret" if she discovers that a different alternative would have done better than the alternative she chose. She feels "disappointment" if the outcome she achieves does not match the outcome she expected. Avoiding regret or disappointment are goals that may trade off in a multiattribute utility framework with more standard goals such as financial gain. Bell uses these concepts to account for a variety of apparent decision "biases," including Ellsberg's paradox (in which decision makers prefer choices in which the probabilities of outcomes are known, to choices in which the probabilities of outcomes are unknown).

Discrepancies between subjects and experimenter's may also occur in fundamental beliefs or assumptions, e.g., about the possibility of change in model parameters. When people predict one quantity (e.g., next year's economic growth rate) based on another quantity (e.g., this year's growth rate), their predictions typically do not "regress to the mean" as they should given that the two quantities are imperfectly correlated; a very bad year is typically expected to be followed by another very bad year, instead of by a more nearly average year (Kahneman and Tversky, 1974). In a dynamic context, however, the "regression fallacy" may reflect sensitivity to fundamental changes, e.g., the next year could be worse than the present year if the economy is in fact declining. In many cases, the costs of missing such a change would be significantly greater than the costs of a false alarm.

Birnbaum (1983) has criticized the simple Bayesian updating rule that serves as a normative standard in "base rate neglect" studies (e.g., Kahneman and

Tversky, 1973). The simple Bayesian model assumes, in the cab problem, for example, that the witness commits the very fallacy that the experimental subjects are accused of: that the witness did not consider base rates in making his own report! Birnbaum develops a more complex Bayesian inference model that uses signal detection theory to model the witness, and in which the witness balances the costs of different kinds of errors in light of the base rates; such a model leads to the "biased" answer preferred by subjects in this experiment.

There is a more fundamental problem with the base rate neglect studies. The experimenters assume that the frequency data provided in the instructions are decisive evidence for the "true" base rate (Niiniluoto, 1981; Birnbaum, 1983). Bayes' rule requires that the decision maker estimate the probability (before considering the witness' report) that a Blue cab would be responsible for an accident just like the one that happened, viz., hit-and-run, at night, at this location, etc. But what they are given is quite different: the frequency of Blue cabs in the city. Subjects must decide for themselves whether this frequency is an accurate estimate of the required probability. They are free to discount it or disregard it, and should do so, for example, if they believe the larger company is likely to have more competent drivers than the smaller company, is less likely to have drivers who would leave the scene of an accident, and so forth (Gigerenzer and Murray, 1987).

The above explanation accounts for the finding that subjects do *not* neglect so-called "causal" base rates (the frequency of *accidents* by each cab company). This result is not necessarily a bias in favor of causal reasoning (Tversky and Kahneman, 1980); rather, subjects may have judged quite reasonably that information about the number of accidents is stronger evidence for the required base rate than information about the number of cabs (cf., Bar-Hillel, 1980). Unfortunately, however, this more elaborate decision analytic model still falls short. It does not explain why, if non-causal base rates are deemed irrelevant, subjects *do* use non-causal base rates when the witness is dropped from the story. To accommodate this, the decision analytic model would have to be elaborated further, so that the impact of the frequency data was dependent on the presence or absence of other evidence. While formal consistency might thus be restored, very little insight into the subjects' actual reasoning is provided.

"Overconfidence" is perhaps one of the more intuitively understandable decision biases. Yet Kadane and Lichtenstein (1982) argue that calibration sometimes violates Bayesian normative constraints. In particular, when there is no feedback regarding whether or not predicted events occur, and when the predicted events are non-independent, a consistent decision maker should *not* be calibrated. Events can be non-independent because their occurrence is influenced by some third event. For example, a pilot might believe that 90% of the airports in a given area will be closed if a storm intensifies, but that only 60% will be closed if it does not intensify. If the pilot thinks that the chance of intensification is 50% (and if he is a good Bayesian) he should predict the chance of closing for each airport to be (.5)(.9) + (.5)(.6) = .75. But the true frequency of closings will turn out to be either 60% (in which case the pilot was "overconfident") or 90% (in which case he was "underconfident"). In calibration studies, the same kind of non-independence could occur if a common cognitive model or reasoning method were utilized to assess the probabilities of different events. A Bayesian model that captured this would have to conditionalize probability assessments on the validity of specific aspects of the subject's own reasoning.

Lopes (1982) has argued that an overly rigid normative standard is used in studying intuitions about randomness (e.g., Wagenaar, 1972; Kahneman and Tversky, 1972). The principle function of a concept of randomness is to serve as a baseline against which significant patterns (i.e., *non*-randomness) can be detected. Judgments of randomness, then, will depend on the types of non-randomness that are likely in a particular problem domain. If random and non-random sets of patterns overlap (i.e., certain patterns could be either random or non-random), then the judgment will also depend on the relative costs of missing a significant pattern versus mistakenly identifying a random pattern as significant. So-called misconceptions of chance, then, may be the result of subjects and experimenters using different criteria of randomness. For example, all sequences of heads and tails of the same length are equally likely in a coin toss, yet Kahneman and Tversky (1972) found that subjects regard sequences with representative proportions of heads and tails (e.g., HTHTTH) as more probable than sequences with less representative proportions of heads and tails (e.g., HHHHTH). In most real-world domains, detecting non-random sequences is less important than detecting atypical proportions (e.g., of defective products off an assembly line). The subjects might, therefore, have failed to understand that they were to deal with the sequence as such, instead classifying the coin tosses by the number of heads and tails. In that case, they were correct in regarding the event of 5 heads and 1 tail as less probable, hence more likely to be non-random, than the event of 3 heads and 3 tails. A more comprehensive Bayesian model of randomness judgments might require prior probabilities for different patterns of non-randomness in the appropriate problem domain and a signal detection analysis of the costs of different kinds of errors. It is far from clear that such assessments could be meaningfully provided.

Discussions of the "belief bias" often sound as though classical logic or probability theory dictated the answer: contradictory evidence should prompt rejection of a contradicted hypothesis (cf., Popper, 1959); disconfirming evidence should lower confidence in the hypothesis. Neither logic nor probability theory, however, is so definitive. The basic reason, in both cases, is that prediction from a hypothesis is always implicitly or explicitly dependent on auxiliary beliefs; the failure of the prediction to come true may therefore lead to the rejection of these other beliefs rather than rejection of the target hypothesis. The impossibility of definitive falsification is known in the philosophy of science as the "Quine-Duhem thesis" (Quine, 1952; Duhem, 1914/1962). An Army intelligence officer who has evidence that an attack will occur at a certain location, but who fails to discover in the photographic evidence the expected forward movement of the enemy's artillery, need not change his belief in the location of attack. Instead, he can question the reliability of the negative indicator: perhaps the enemy plans to omit the initial artillery barrage for purposes of surprise, or artillery in some other location has a sufficient range to cover the attack area, or artillery equipment is unavailable or not in working order, and so on; perhaps artillery movement occurred but could not be detected photographically, because weather, foliage, and/or intentional camouflage masked its presence. We quite properly calibrate our trust in one source of information (e.g., a witness, a scientific experiment, an instrument, our own senses) by reference to its agreement or disagreement with other sources of information, and also by reference to its agreement or disagreement with our own beliefs (if a stranger tells me there is an elephant in the next room, I am unlikely to place much credence in anything he says). How could it be otherwise, since there is no directly revealed "ground truth"?

This kind of reasoning can in fact be accommodated within Bayesian models, although at great cost in complexity. The impact of each piece of evidence on the hypothesis may depend on all the other evidence that has been observed, i.e., all potential observations may be combined into a single variable; the set of hypotheses may be expanded to include all combinations of truth and falsity of the original hypothesis and the auxiliary beliefs. In effect, such a model abandons the "divide and conquer" strategy of decision analysis (although Pearl, 1988, has made some efforts to simplify the modeling of some of these effects). Bayesian models may also permit the impact of evidence to depend on prior beliefs, by conditionalizing evidence assessments on the decision maker's own probability judgments (e.g., French, 1978) - but at the price of even greater loss of economy. A final source of complexity is the requirement to explicitly model all the possible temporal orders of the evidence, since two conflicting pieces of evidence may be interpreted differently depending on which is experienced first (Woodcock, et al., 1988). It is extremely implausible to suppose that decision makers have explicit and exact models of this sort in their heads that anticipate their potential reactions to every possible state of affairs. Such models provide no real insight into the reasoning of decision makers: they do not tell us what it is that does or does not make sense about "belief bias" behavior.

A "confirmation bias" has also been found in the process of collecting information to test a rule. As noted above, Wason (1968) used rules of the sort: If there is a vowel on one side of a card, there is an even number on the other. Of four cards, showing an A, B, 4 and 7, respectively, subjects chose to turn over the cards with A and 4. According to Wason's simple logical model, the card showing a 4 cannot falsify the rule, whereas the card showing a 7 could (if there were a vowel on the other side). There is strong reason, however, to doubt the general applicability of such a simple model. A scientist testing the hypothesis that "All ravens are black" will hardly set out collecting non-black things to determine whether they are ravens (this has been called the "paradox of confirmation"; Hempel, 1945). Even if the hypothesis under investigation is really false (hence, there is a non-black raven somewhere), the scientist will find the non-black raven faster by looking at ravens than by looking at non-black things. The reason lies in knowledge about the domain: if the rule is false, the proportion of non-black things that turn out to be ravens will still be much smaller than the proportion of ravens that turn out to be non-black. A fairly complex Bayesian model is required to represent such beliefs (Horwich, 1982, pp. 53-62), but they appear to be quite general across problem domains: useful rules link reasonably specific classes to one another (e.g., ravens and blackness), while the complements (non-ravens, non-black things) will be larger and more diverse. If subjects (mistakenly) carry over these strategies to the card task, turning over the 7 would be rejected as an inefficient method for rule falsification. Klayman and Ha (1987) provide a similar analysis of a rule discovery experiment by Wason (1960).

In all of these examples, experimenters appear to have applied an overly simple decision analytic model to their own problems - perhaps as a sort of "satisficing" in the face of complexity. Potentially important factors are not captured by these models. In particular, subjects do not accept information provided to them at face value; they evaluate instructions, randomness, base rates, conflicting evidence, and potential tests of hypotheses in the light of real-world knowledge (sometimes inappropriately extrapolating from familiar to unfamiliar settings). Subjects' performance can, in principle, be accounted for by more complex decision analytic models - indeed, virtually any decision making performance could be captured by some decision analytic model, given the freedom in choosing structure and inputs (Glymour, 1980). For this very reason, of course, the exercise of inventing a decision analytic model to predict subjects' responses does not prove that their decision processes are "rational."

In the absence of more systematic investigation, however, the very existence of alternative models compels us to suspend judgment about which model is appropriate. It is at least equally arbitrary to assume that subjects' beliefs and goals are best fit by the models adopted by experimenters. Abstractly fitting performance to a convenient normative template is of little help in evaluating a reasoning process. The flexibility of decision analytic modeling is a virtue if one is looking for an intellectual tool, but is a drawback if one is looking for a Platonic test of rationality.

It seems implausible, to say the least, that subjects would be able or willing to provide the many precise assessments demanded by the models that fit their behavior. In other words, the models that make the subjects' behavior appear rational are not very successful as potential "intellectual tools." But it shouldn't be necessary to prove that subjects were actually using (or could use) the hypothesized models. It would be enough to show that they are sensitive in an roughly appropriate way to the variables contained in the model. Developing such models highlights variables to which subjects might in fact be sensitive, such as the credibility of base rates and evidence, or the different types of "non-randomness" that characterize different domains. The question to which we now turn is whether there are simpler and more illuminating (but also normatively plausible) processing strategies in which such variables might be incorporated.

Challenge 6: There are alternative, non-Bayesian normative concepts that justify the decision maker's way of approaching the problem.

Challenge (6) represents the most fundamental criticism of the decision bias paradigm. It rejects the definition of formal consistency that has been used to characterize and define decision biases. It argues, in essence, that *rationality* cannot be equated with decision theory; decision theory is one of a set of competing claims regarding what it means to make rational decisions. An extreme variant is (a) that biases are not possible *in principle*, because the chief criterion for adopting a normative theory is its fit to actual decision making behavior (L. J. Cohen, 1981); if (non-stressed) behavior disagrees with the model, the model must be automatically dropped. A more moderate position is (b) that a variety of alternative normative frameworks and prescriptive concepts now exist for decision making (e.g., Zadeh, 1965; Shafer, 1976; L. J. Cohen, 1977), and some of these may shed more light than decision analysis on the thought processes that decision makers actually utilize, and perhaps provide a more adequate tool for helping them make better decisions.

(a) The possibility of biases. In the previous chapter, we discussed different paradigms for the relationship between normative and descriptive. When decision-making behavior disagrees with a normative model, the rationalist paradigm condemns the behavior; the formal-empiricist paradigm will consider changing the model - as long as the new model has some degree of formal plausibility. L. J. Cohen's position seems to drop the latter qualification: the model *must* be changed if it disagrees with behavior. Before addressing Cohen's challenge, let us back up and ask what we mean by "normatively correct" decisions. What basis is there for choosing among competing "normative" frameworks? It is helpful to distinguish three kinds of arguments, which may be loosely described as formal, cognitive, and behavioral:

- Formal: justifies models like Bayes' rule or maximization of subjectively expected utility by deriving them from "self-evident" axioms. This is the *only* kind of normative justification according to orthodox Bayesians. Not coincidentally, Bayesian theory has a preeminent claim to validity on this basis. De Finetti (1937/1964), for example, showed that unless your beliefs and choices conform to the rules of decision theory, a clever opponent could turn you into a "money pump," i.e., devise a set of gambles in which you would inevitably lose. Savage (1954), Lindley (1982), and others have provided derivations of decision theoretic constraints from other axioms. Axiomatic justifications of other normative approaches have also been developed, however (e.g., fuzzy set theory: Bellman and Giertz, 1973; Nau, 1986).

! Cognitive: justifies a model (e.g., Bayes' rule or multiattribute utility analysis) in terms of its own face validity and practicality. Two kinds of cognitive considerations have been advanced in recent discussions: (1) Does the model require inputs about which the decision maker has confident and precise intuitions (von Winterfeldt and Edwards, 1986)? (2) Are the operations applied by the model to the inputs plausible - i.e., is there a strong analogy between the problem at hand and "canonical examples" of the application of the model (Shafer, 1981; Shafer and Tversky, 1988)? For some theorists, normative justification is exclusively at this cognitive level - although there is disagreement among them on whether decision analytic models will always come out on top (von Winterfeldt and Edwards) or whether alternatives might sometimes be justified (Shafer). For orthodox Bayesians, cognitive concerns are important, too, but merely from an engineering, not a normative, standpoint: Bell, Raiffa, and Tversky (1988) and Brown (1989b) propose a "prescriptive" science to bridge the gap between "normative" constraints (thought of as purely formal) and "descriptive" shortcomings of real decision makers.

! Behavioral: justifies a model in terms of the match between its *outputs* and the actual performance of decision makers (under non-stressed conditions); a normative theory must be revised if it does not describe "human intuitions in concrete, individual cases" (L. J. Cohen, 1981). Systematic flaws in decision-making *competence* (i.e., biases) are ruled out by definition; errors can only be the by-products of ignorance about the correct methods or of *performance* limitations such as fatigue, lapses of memory, inattention, and lack of motivation. L. J. Cohen has championed this position in its most extreme form, but arguments are quite common in the mainstream normative literature for the superiority of one or another framework based on the plausibility of its conclusions in specific examples (e.g., Zadeh, 1984; Bell, 1982; Ellsberg, 1961).

For each of these levels - axioms, models, and outputs - there are theorists who regard it as the exclusive touchstone of rationality. But *no one* of these levels, I argue, is adequate by itself:

! Limitations of the formal level: Axiomatic derivations say that if you accept the axioms, you must accept the model (e.g., Bayesian inference, maximization of SEU); but if you don't accept the axioms, you need not accept the model. Yet every derivation in the literature requires one or more axioms that have no inherent plausibility. For example, Savage (1954) assumes that a decision maker always knows her preferences among gambles; she either prefers gamble A to gamble B, prefers gamble B to gamble A, or is indifferent - but she is never *ignorant* (Shafer, 1988). Lindley (1982) makes an equivalent assumption: that uncertainty is to be measured by a single number (although at least two numbers would be needed to measure the decision maker's *confidence* in her own probability assessments).

What do we gain by giving up the non-compelling axioms? Non-Bayesian systems may have attractive features, especially at the cognitive level, that Bayesian decision theory lacks: alternative, richer frameworks may more adequately represent types of uncertainty that characterize real-world problems, such as incompleteness of evidence, vagueness, and imprecision. Even Bayesians must implicitly give up formal justification (which requires the assumption that all probabilities and preferences are precisely known) in order to make sense of applied decision analysis as a tool for generating unknown probabilities and preferences from known ones.

How much do we lose by rejecting the non-compelling axioms? Systems that permit ignorance about beliefs and preferences (i.e., which measure uncertainty by more than one number), such as fuzzy logic or Shafer-Dempster belief functions, may still posses the attractive properties contained in the other axioms, e.g., independence of beliefs and preferences. Is a framework which is consistent with *all the other* axioms really *less justified* than decision theory? The only thing such frameworks lack, by comparison with Bayesian theory, is the demonstration of their uniqueness via axiomatic derivation. Being a doctrinaire Bayesian, from this point of view, is like preferring to live in a *uniquely tiny* house instead of in a comfortable house that is the same size as your neighbors'.

! Limitations of the cognitive level: Cognitive criteria emphasize the fit between the inputs required by a model and the decision maker's knowledge, and between the processing required by the model and the decision maker's judgments of plausibility: i.e., normative models are good "tools" for generating needed probabilities and preferences from known probabilities and preferences. But the tool metaphor, taken by itself, has trouble with a crucial feature of real decision aiding: the iterative process by means of which initial models are replaced by better ones until both analyst and decision maker are satisfied. What partially drives this process is the fact that decision makers often come to a problem with intuitions not only about the *inputs* to a model, but about the *answer* as well. When a model gives answers that are widely discrepant from intuitions, decision makers quite properly want to re-examine the model and improve it. A direct judgment or choice regarding the answer to the problem might sometimes capture the decision maker's knowledge more effectively than a more detailed analysis. It may sometimes make sense to resolve inconsistency by changing the model.

! Limitations of the behavioral level: According to the behavioral criterion, we should *always* abandon a model when we don't like the answers. But if this were true, decision making (under non-stressed conditions) could never be improved by analysis; the status quo would always be best. In fact, however, models can and do cause decision makers to change their minds. Models can be persuasive because they have certain very general attractive properties (e.g., independence of preferences and beliefs), or - perhaps more importantly - because they seem to organize all the relevant factors in a particular problem in a reasonable way. It is arbitrary to take intuitions about specific cases seriously, as L. J. Cohen (1981) urges, but to dismiss more general intuitions about the fit between the problem and a model (the cognitive level) and the more abstract intuitions about formally desirable properties. (L.J. Cohen (1983) himself seems to have come around to such a view.)

Formal, cognitive, and behavioral criteria are all, in the end, "intuitions" (Einhorn and Hogarth, 1981) - it is hard to justify any absolute epistemological priority to one set over the other. According to one view of justification (Goodman, 1965; Rawls, 1971; Daniels, 1979), normative judgments involve an equilibrium between general principles and performance: We amend the general principles if they imply a practice we are unwilling to accept, and we amend practice if it violates a principle we are unwilling to amend. "The process of justification is the delicate one of making mutual adjustments between (principles and practices); and in the agreement achieved lies the only justification needed for either" (Goodman, p. 64).

Some writers have been alarmed at the "relativistic" implications of this view; any practice is justified if it fits the principles a decision maker happens to accept (e.g., Stich and Nisbett, 1980). This criticism confuses the *process* of justification with *successful* justification. We need not regard a behavior as successfully justified simply because an equilibrium of this sort has been achieved. One equilibrium can be more convincing than another - if it draws on a wider range of more plausible theories (the formal level), if it organizes more of the decision maker's knowledge (the cognitive level), and if it fits a wider range of behavior. The success of the behavior in achieving real-world goals (cf., Thagard, 1988) is also a legitimate indicator of the validity of a normative framework; in fact, our intuitions may themselves be the product of an evolutionary past comprising a long series of actual successes and failures.

What are the implications of this view for decision biases? How people *actually* make decisions is a relevant consideration in evaluating a normative model of how they *ought* to decide. Is the normative evaluation of decision making, therefore, completely circular? Must systematic decision-making errors be dismissed as impossible? No. The reason is that other criteria besides the behavioral also contribute legitimately to the evaluation of normative models: i.e., at the formal level - desirable general properties; and at the cognitive level - a match of the model's required inputs and modes of processing to the knowledge representations and plausibility judgments of decision makers. Because of these other criteria, normative models may survive behavioral disagreement; if cognitive or formal criteria strongly support the model, behavioral disagreement may cause us to amend the behavior.

Nevertheless, L. J. Cohen (1981) was not all wrong. When behavior clashes with a normative model, there is always a force of at least some magnitude pulling away from the "error" verdict: *The more systematic the "errors," and the more prevalent they are among successful practitioners or experts, the greater the feeling that we should modify theory rather than condemn the practice.* Like other tradeoffs (see the discussion of challenge 3 above), precise and confident judgments of the relative importance of these criteria may simply not be realistic. In the real world (e.g., applications of decision aiding or training), we may have to decide on a case-by-case basis. The most constructive strategy is to use conflicts between behavior and model as a prompt for expanded understanding of both. For example, in our discussion of challenge 5, violations of behavioral consistency led to revised decision analytic models. Some of these new models proved to be unpersuasive on the cognitive level, however, because of their complexity and the difficulty of the inputs they required. The next step is to look at revisions at the formal level: i.e., non-Bayesian normative frameworks that may shed more light on decision making processes than complex Bayesian models.

(b) Alternative normative concepts. In the past two decades there has been a lively debate in the normative research community regarding alternative concepts and methods for reasoning with uncertainty (e.g., Smithson, 1989;

Cohen et al., 1985; Kanal and Lemmer, 1986). This on-going competition among normative concepts has been largely, though not entirely, disregarded by psychologists concerned with studying human decision biases (but see Shafer and Tversky, 1988). Yet one of the key issues of that debate has been decision making under varying conditions of knowledge and ignorance; some of the suggested solutions may, therefore, illuminate unaided decision making in the real world. We will focus here on one such approach: a framework for assumption-based reasoning that formalizes effective metacognitive control over the application of knowledge.

Classical Bayesian theory has no easy way to represent the amount of knowledge or ignorance underlying an uncertainty judgment. A .5 probability that Jones will beat Smith in a tennis match may represent thorough knowledge of the capabilities of the two players, leading to the conclusion that they are evenly matched, or it might reflect complete ignorance (Gardenfors and Sahlin, 1982). Similarly, a .9 probability might be based on a lot of evidence or very little. From a Bayesian point of view, choices should be unaffected by the decision maker's confidence or lack of confidence in his own beliefs: the decision maker whose .5 probability represents a large amount of knowledge should act in precisely the same way as the decision maker whose .5 probability represents virtually no knowledge at all. Nevertheless, real choices are affected not just by probability, but by degree of knowledge (Ellsberg, 1961). People may prefer to bet on *either* player in a tennis match in which the opponents are known rather than bet on *either* player in a match whose players are unknown, even though the "probabilities" should be 50-50 in each case. These choices violate the probability axioms: the less well-understood "probabilities" appear to sum to less than 1.0. Yet such choices can seem quite reasonable on reflection.

According to a number of normative theorists (e.g., Ellsberg, 1961; Gardenfors and Sahlin, 1982; Levi, 1986), under conditions of ignorance decision makers are entitled to fall back on non-decision-analytic criteria of choice. One such criterion, for example, involves comparison of options in terms of their worst-case outcomes. A worst-case strategy would reduce the desirability of betting on either of the unknown players in a tennis match: since both Smith and Jones are unknown, it is possible that Jones is much worse than Smith, *and* it is possible that Smith is much worse than Jones. An alternative, equally permissible, decision strategy is to evaluate unknown probabilities in terms of the best case. More generally, to the extent that a decision maker's knowledge does not specify exact probabilities, preferences, or actions, she is free to adopt *assumptions* within the range permitted by her knowledge.

Assumption-based reasoning has received considerable attention from the artificial intelligence community as a method for handling incomplete information (e.g., Doyle, 1979; deKleer, 1986). Unfortunately, in these assumption-based systems, the process of metacognitive control, e.g., revising assumptions when they lead to contradictory results, is largely arbitrary. Cohen (1986, 1989) has proposed a system that provides for higher-order reasoning (i.e., metacognition) about the assumptions in quantitative models; in turn, quantitative measures of the reliability of beliefs, the magnitude of the conflict, and the responsibility of particular assumptions for the conflict guide the processes by which assumptions are adopted, evaluated, and revised.

Techniques of this sort may capture aspects of naturalistic reasoning more successfully than the decision analytic paradigm. In particular, the notion of assumption-based reasoning fits several recently proposed models of naturalistic decision making quite well (see Lipshitz in this volume). According to Lipshitz (in his theory of decision making as Argument-Driven Action) and Klein (in his theory of Recognition-Primed Decision Making), decisions do not involve explicit comparison and choice among alternatives. Instead, a single action is generated by matching the current situation to known cases; this is followed, optionally, by a process of verifying or critiquing the generated action, rebutting the critique, and (if necessary) modifying or rejecting the option. Similarly, I have proposed (1989) that in inference problems, a "first-blush" or normal reaction to a piece of evidence is followed, optionally, by consideration of possible exception conditions or rebuttals, and possibly by a revised interpretation of the evidence. More generally, an action or inference, once it has been generated, is assumed appropriate, until reasons are found to believe otherwise.

In the last section, we observed patterns of behavior that violated simple decision analytic models, but which could be accommodated within far more complex (but less plausible) decision analytic models. For example, a first step toward handling "base rate neglect" was a Bayesian model that explicitly evaluates the reliability of the experimentally provided frequency data as evidence for the true base rate. This model fails, however, to accommodate the apparent inconsistency in evaluation from one condition to another: when individuating evidence (the witness) was available, frequency data were apparently regarded as unreliable; but when no individuating evidence was present, frequency data were apparently regarded as reliable. A far more complex model is required at this point, which anticipates every possible combination of evidence and frequency data. *An account in terms of ignorance and assumption-based reasoning, by contrast, stays simple: it divides the work of the complex model between simple first-level beliefs and simple meta-level rules.*

In the base-rate studies, it is reasonable to suppose that subjects are unsure about the reliability of the frequency data: they are not told the source of the information or how recent it is; they are told nothing about other relevant factors that could distinguish the two cab companies (e.g., competence of drivers, training not to leave the scene of an accident); finally, they have virtually no experience with information of this kind in comparable settings (either experimental or real-world). Nevertheless, in the absence of any other information, they may well be prepared to *assume* that the frequency data are reliable and relevant. Such an assumption, however, is not equivalent to a belief based on knowledge: it is subject to change. In particular, conflict between the frequency data (e.g., the proportion of Blue cabs in the city = .15) and individuating evidence (e.g., the witness says the guilty cab was Blue) triggers a process of problem solving in which the assumptions that contributed to the conflict are re-examined and possibly revised. In the presence of individuating data, therefore, subjects may retract their assumption that the frequency data are reliable. There is no inconsistency in beliefs, only a quite reasonable change in assumptions.

The so-called "belief bias," in which new evidence is reinterpreted to fit a prior hypothesis, is subject to a very similar analysis. This phenomenon can also be handled in a decision analytic model; within such a model, evidence items will not be independent of one another conditional on the hypothesis, as they are in standard Bayesian models; in fact, evidence and prior probabilities would not be independent either. Such a model, however, is often intractably complex and requires precise assessments of how the decision maker will respond to every combination of internal prior judgments and external events; the model must be complicated even further if the temporal order of the evidence is taken into account.

The most plausible way to handle "belief bias" behavior is not to impose complex Bayesian models of belief change, but to introduce a notion of assumption-based reasoning. Decision makers are not completely confident ahead of time about the meaning of every piece of evidence they are likely to encounter. In the normal course of events, assumptions are adopted in order to facilitate the smooth incorporation of new data into a pre-existing framework, or schema (as in Piaget's "assimilation" or Thomas Kuhn's "normal science"). Sometimes the requirement to assimilate data conflicts with the assumptions usually adopted (e.g., that a source of information is reliable until proven otherwise, or that the worse-case outcome will occur); the usual interpretation may then be overridden (either by an automatic, expectancy-driven process or by conscious reflection) and the data "explained away." The same piece of evidence may thus have a different interpretation as a function of the context of other evidence and beliefs in which it occurs. Assimilation, however, can be carried too far. In a dynamic environment, if a long series of apparently conflicting pieces of evidence has been explained away, the decision maker may grow uneasy. At some point, he may realize that a simpler overall set of beliefs and assumptions can be achieved by rejecting the favored hypothesis. Examples of this sort of "Gestalt shift" are the "scientific revolutions" referred to by Kuhn (1962).

According to Fischhoff and Beyth-Marom (1983), people choose to perform tests which, no matter what information they actually obtain, will be interpreted as confirming their favored hypothesis. Baron et al. (1988) found that subjects consistently overestimated the value of questions that were regarded as useless by a Bayesian value-of-information model. Another possibility, however, is that the Bayesian value-of-information model missed the point. In that model, the potential impact of an observation is based on the interpretation that the decision maker assigns to it *at the present time*. But the present interpretation of the evidence may depend on assumptions, which are subject to change. Suppose, for example, that subsequent evidence continues to be interpreted as confirming the hypothesis, but that more and more work to "explain away" is required to do this. At some point, the decision maker does change her mind: the cumulative effect of all this "confirming" evidence is a disconfirmation! The tests that produced that evidence clearly did have value, even if no single one of the tests could have caused any perceptible change in belief *at the time it was performed*.

A similar approach might illuminate the dynamic aspects of choice behavior. According to multiattribute utility analysis, precise preferences among criteria exist in the decision maker's head waiting to be elicited; when the appropriate algorithms are applied to such preferences, one and only one option (barring exact ties) always turns out to be best. Choice in the real world, however, is often a dynamic process, in which changing assumptions about goals and options reflect increasing understanding of preferences. This process may involve rather sophisticated and active problem-solving steps. For example, when goals are in conflict (i.e., no single alternative satisfies all the criteria), the decision maker may re-examine assumptions: he may drop some goals or change their priority; he may relax or qualify the scoring of alternatives on various criteria; or he may try to discover alternative means to the same ends. By contrast, when evaluative criteria are incomplete (i.e., an insufficient number of options have been eliminated), the decision maker may look for additional plausible assumptions: he may explore reasons to strengthen one or more of the original goals; he may look for additional goals; he may scrutinize the performance of candidates on various criteria more closely; he may even find a way to make more precise tradeoff judgments. A multiattribute utility analysis, by giving a definitive answer prematurely, may forestall rather than facilitate improved understanding.

The same principles may apply to a variety of other biases in inference and choice. Assumptions regarding predictability and change in underlying processes may affect the so-called "regression fallacy," in which "overextreme" predictions of one quantity are made based on another quantity. Assumptions about various possible exceptions to a conclusion may influence "overconfidence" (such assumptions are necessary, under certain conditions, for the achievement of calibration). Assumptions about the types of non-randomness to be expected in particular domains may influence judgments of randomness. Assumptions about the representativeness of recalled instances may underlie availability effects. In many experiments, subjects may make assumptions about the relevance or credibility of experimentally provided information.

On what grounds do I claim that a system of assumption-based reasoning can be "normative"? Clearly, anyone for whom Bayesian decision theory and/or logic are normative *by definition* will be unmoved by such a claim. From a naturalistic point of view, however, we consider the cognitive and behavioral plausibility of a proposed normative framework in addition to its more formal virtues. From this point of view, assumption-based reasoning is entirely defensible:

- Formally, it retains the more compelling axioms underlying decision analysis, since many important inconsistencies are attributed to changes in assumptions rather than to firm beliefs and preferences (Levi, 1986; Kyburg, 1968/1988). Changes in preferences, for example, as the decision maker learns which options are feasible can be accommodated without sacrificing the independence of (firm) preferences and (firm) beliefs.

- Cognitively, assumption-based models require more natural inputs and provide more plausible processing of those inputs than decision analytic models. First, they do not demand assessments (e.g., of the reliability of information) that are more precise than the decision maker's knowledge or capacity permits. Assumption-based reasoning is tailored to limited-capacity processing, in which it is not possible to marshall all the information that may conceivably be relevant for every decision. Assumptions help a person make more effective use of the knowledge that is available, by permitting selective focus on hypotheses, outcomes, and dimensions about which she has the most information (by assuming, provisionally, that unexamined possibilities contain no surprises). Problems with the first-pass solution prompt activation of additional parts of the decision maker's knowledge, or additional external information collection.

  The assumption-based approach also seems inherently more plausible than the standard decision analytic model. Bayesian updating handles conflicting evidence, in effect, by taking an average; there is never a definitive picture of the situation, in which conflict is explained or resolved. Assumption-based reasoning, by contrast, takes conflict as a symptom that something might be wrong in the reasoning that led to the conflict, and uses it as a stimulus to find and correct mistakes (Cohen, 1986, 1989). It exploits the opportunity for self-correction in dynamic environments.

- Finally, the assumption-based model matches actual decision-making performance far more closely. If one doctor says a child has the mumps and the other says she has measles, parents might differ in their responses: some might assume one doctor was correct, and disregard the other; other parents would look for a third doctor, investigate the credentials of the two conflicting doctors, or explore the reasons for their diagnoses in more depth. But very few parents would take an average (i.e., settle for an inconclusive assignment of probabilities to the two possibilities). If both doctors happened to assign a very small probability to a third disease, a Bayesian model would assign full support to the "compromise" - even though neither doctor regarded it as an important possibility (Zadeh, 1984).

Assumption-based strategies also pervade highly successful "expert" reasoning. For example, scientists seek a coherent picture of nature rather than an assignment of probabilities to alternatives: they try to explain unexpected new observations by means of minimal adjustments in existing theory (Quine, 1960); and they adjust (or "calibrate") experimental procedures until the procedures produce theory-predicted results (Greenwald et al., 1986). The alternative to these practices is a "disbelief bias," i.e., abandoning a theory at the first sign of trouble, probably crippling science's ability to find (or impose) regularities in nature. The relatively sudden shifts that characterize "scientific revolutions" suggest a process of initially explaining away conflicting data, then reexamining and revising assumptions - as opposed to the continual changes in probability characteristic of Bayesian models.

By fitting a variety of actual behaviors into the framework of assumption-based reasoning, however, we by no means imply that all decision-making performance is normatively correct. The preponderance of evidence suggests that actual behavior will not perfectly fit *any* normative model that is also plausible on cognitive and/or formal grounds. *The goal is to devise normative models that are illuminating and insightful: i.e., they must provide a close enough fit to actual decisions and decision processes so that the discrepancies really do seem like errors that are worthy of attention.*

What kinds of decision errors, then, does an assumption-based theory of decision making identify? Biases, from this point of view, are defects in the metacognitive processes that control the verification and revision of conclusions. In the assumption-based model, as in Bayesian decision theory, errors involve formal inconsistency; but the system within which consistency is now defined has been tailored to fit the processes and constraints that affect real-world problem solving. For example,

! Belief bias: It is not automatically an error when a decision maker "explains away" an apparently conflicting piece of evidence; however, she may explain away "too much" conflicting evidence, and (especially in a time-stressed environment) fail to maintain a sense of "cumulative doubt" that triggers the reassessment of a favored hypothesis. Moreover, the opposite error is also possible: she may take evidence literally when she should have questioned it - i.e., explained it away.

! Overconfidence: In assessing uncertainty, the decision maker may overlook important exceptions to a conclusion ("overconfidence"); alternatively, however, she may take far-fetched possibilities too seriously, paralyzing effective inference and action.

! Satisficing, elimination-by-aspects: It is not automatically an error for a decision maker to adopt goals on evaluative dimensions and to neglect tradeoffs; however, it is an error if she fails to raise or lower her goals realistically in the light of their achievability. Once again, adjusting her goals too much or too soon may also be an error.

! Ellsberg's paradox: By focussing on the worst case, the decision maker may miss important opportunities; or by focusing on opportunities (the best case), she may overlook risks. However, she may also go wrong by trying to summarize multiple outcomes or dimensions by an abstract average (incorporating both worst and best cases) that is poorly correlated with successful action.

The assumption-based approach formalizes and refines our understanding of the kinds of errors we have already identified, in our discussion of the other challenges: failure to compensate for shortcomings of decision strategies, failure to make midcourse corrections in an inference or choice, failure to effectively exploit knowledge and to efficiently deploy capacity. Systematic errors of this type may well exist, but such "decision biases" are clearly not the ones to which we have grown accustomed. The "cure" may also be different from what is usually supposed: it need not require the imposition of decision analytic models. Decision making should be both understood and improved in its own terms.

Chapter ?: THE BOTTOM LINE: NATURALISTIC DECISION AIDING

Marvin S. Cohen
Cognitive Technologies, Inc.

It is appealing to suppose that technology has the means for improving decisions. Computer-based systems to advise decision makers have incorporated decision analysis, expert knowledge, and/or mathematical optimization. Success, however, has been limited; the very features of real-world environments that are stressed in this volume, e.g., their ill-structuredness, uncertainty, shifting goals, dynamic evolution, time stress, multiple players, and so on, typically defeat the kinds of static, bounded models provided by all three technologies. Each decision involves a unique and complex combination of factors, which seldom fits easily into a standard decision analytic template, a previously collected body of expert knowledge, or a predefined set of linear constraints. Users are sometimes ahead of the aids in their ability to recognize and adapt to such complex patterns.

The literature on decision biases (as described in chapters 3 and 4) has reinforced the tendency to regard users as passive recipients of assistance: unaided decision making is presumed to be subject to fundamental flaws, which can be corrected only by adoption of "normative" methods such as Bayesian decision analysis. The rationalist tradition has encouraged a sort of arrogance toward actual decision makers that can only make their acceptance of decision aids (and the aids' success if accepted) less likely (Berkeley and Humphreys, 1982; Lopes, 1988). Such an approach may force decision makers to adopt highly unfamiliar modes of reasoning; as a result, aids may not be used, or if used, may be poorly understood; worse yet, they may fail to exploit user knowledge or expertise that might facilitate adaptation to complex, novel situations. Although there is lip service to "supporting the user rather than replacing him," in technology-driven approaches (whether based on decision analysis, optimization, or expert systems) *the user's approach to the problem*, if not the user himself, *is* replaced: at best, the user may provide probability and utility inputs for a standard decision analytic model.

Should we give up hope of advising decision makers? At the other extreme are less ambitious (and more organizationally acceptable) "status-quo-driven" approaches, which merely automate the more tedious aspects of a task without modifying it in any essential respect. Such aids do not correct any flaws in traditional procedures, fail to exploit potential synergies between humans and computers, and - ironically - may be just as unacceptable to users as technology-driven aids.

An alternative approach is to start with the user's preferred way of solving the problem and to examine its strengths and weaknesses carefully. Attention is paid to how decision makers actually solve problems (including consideration of individual differences and changes over time) and the cognitive strategies and knowledge representations underlying performance, as well as to normative models as sources of potential insight for improvements. Aids are then designed which support more optimal *variants* of the user-preferred strategy. We have called this methodology Personalized and Prescriptive Aiding (Cohen, Bromage, Chinnis, Payne, and Ulvila, 1982; Cohen, Laskey, and Tolcott, 1987). The naturalistic framework encourages aiding that is *user-driven* (or personalized) - i.e., tailored to user knowledge representations and processing strategies, but not necessarily to the status-quo procedure - and simultaneously *problem-driven* (or prescriptive) - i.e., able to safeguard against errors and pitfalls to which the user-preferred approach is susceptible, but not necessarily wedded to traditional normative models.

The reader need not subscribe to non-Bayesian normative models (chapter 4, challenge 6), or to alternative Bayesian models (challenge 5), to be persuaded about the value of a more adaptive approach to decision aiding. The reader need not even accept the claim that knowledge (challenge 3) and limited capacity (challenge 4) sometimes justify non-Bayesian decision processes. The case for adapting aids to users can be made purely in terms of outcomes (challenges 1 and 2): by arguing, for example, that deviations from optimality often don't matter much (von Winterfeldt and Edwards, 1973), and that where they do matter, specific safeguards can be provided. The argument, however, gets stronger, and the associated concept of aiding gets richer, as one's acceptance of the naturalistic point of view moves from outcomes to processes to decisions. Let us look briefly at the implications for aiding, together with some examples, from these different perspectives:

An aid developed for the command staff of an attack submarine illustrates the role of decision analytic models as advisors who step in only when the user's approach is likely to produce an unfavorable outcome (Cohen et al., 1982). As a "hunter-killer" submarine stalks an enemy submarine, the commander must balance competing goals of improving the accuracy of localization and probability of kill (thus taking time and getting closer), while minimizing the chances of being detected (thus shooting early and far way). A decision analytic model can integrate all the factors involved in the time-of-fire decision into a single aggregated figure of merit (subjectively expected utility) for each tactical option. Such a measure, however, requires highly ambiguous uncertainty estimates (e.g., the detection capabilities of the opposing submarine), difficult preference tradeoffs (e.g., the value of own ship versus the target), and numerous assumptions (e.g., about what each commander will do if the first shot misses). By contrast, the personalized and prescriptive aid allows the commander to evaluate options in terms of goals and constraints at whatever level of concreteness he chooses (e.g., achieve a particular probability of kill on the first shot; get within x yards of the target but no closer than y yards). At the same time, in the background, the aid creates a decision analytic model of the problem; it also creates a model of the user: inferring his decision strategy by monitoring his information requests and goal specifications. Finally, the aid compares the recommendations of its model with the implications of the user's strategy. The user is prompted if and only if they are significantly different (at a threshold set by the user): for example, the aid might point out that an option, which has been rejected by the user because it just misses the desired 90% chance of kill, achieves a far better chance of avoiding counterdetection. Prompts are framed in terms of the specific factors that cause the discrepancy, described at the level of concreteness the user prefers. As a result, the user can benefit from the decision analytic model without being forced to abandon his own way of thinking.

A dyed-in-the-wool Bayesian might accept such an aid (grudgingly), because the *outcomes* it leads to should not be significantly worse than those expected from a more orthodox aid. A naturalist is more likely to focus on the advantages of supporting the user's familiar decision *processes*: those processes may produce outcomes that are not merely just as good, but *better* than Bayesian procedures - because they more effectively exploit the decision maker's knowledge and capacity. Traditional decision analysis takes knowledge for granted, assuming, in the extreme, that the required judgments of probability and preference pre-exist somehow in the decision maker's head and are accessible on demand; traditional decision analysis thus neglects the metacognitive processes of inquiry and reflection by means of which a problem-specific model is constructed (Levi, 1986). In response, new *formal* methods have been developed, including the assumption-based reasoning framework that I discussed in chapter 4, that explicitly consider the amount of knowledge or ignorance underlying a decision, and formally incorporate the dynamic processes by which beliefs and preferences are tentatively adopted and subsequently revised.

An aid has recently been developed that focuses directly on the dynamic aspects of crystallizing one's own knowledge. D-CIDER (Cohen, Laskey, and Tolcott, 1987) is based on the premise that users differ in how much they know about their preferences and in the way they know it, and that users' understanding may evolve as they work the problem. D-CIDER enables users to express preferences in a variety of qualitatively different formats, including direct judgments of a sample of options, setting goals on different dimensions, rank ordering dimensions, and assessments of exact or inexact importance weights. Implications of inputs in any one format are displayed in all the other formats, to prompt and constrain further judgments. D-CIDER also provides a choice of decision strategies: e.g., elimination-by-aspects (which screens options by user-set goals in order of importance), dominance-structuring (in which users work backwards from a tentative choice, determining whether the choice can be justified by comparison to other options), and maximization of utility (which takes whatever tradeoff information the user has provided, however partial and incomplete, and calculates which options could be best). Prompts help users shift strategies, and add, drop, strengthen, or weaken their goals, when the information provided is inadequate to make a choice from the available options.

Whether we focus on decision outcomes, decision processes, or the decisions themselves, then, the conclusion is that aiding should start with the user's preferred approach. Normative models are appropriate only when they fit the basic contours of the decision maker's knowledge and preferred method of problem solving. From the naturalistic point of view, the goal of aiding is not to radically alter an experienced decision maker's decision-making style, but to mitigate specific potential weaknesses, and to amplify her decision-making strengths.

REFERENCES

Allais, M. The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American School. In M. Allais and O. Hagen (Eds.), *Expected utility hypothesis and the Allais paradox*. Dordrecht: Reidel, 1979 (French original: 1953).

Anderson, J.R. Acquisition of cognitive skill. *Psychological Review*, 1982, *89*(4), 369-406.

Anderson, N.H. A cognitive theory of judgment and decision. In B. Brehmer, H. Jungermann, P. Lourens, and G. Sevón (Eds.), *New directions in research on decision making*. North Holland: Elsevier Science Publishers B.V., 1986.

Atkinson, R.C. and Shiffrin, R.M. Human memory: A proposed system and its control processes. In W. K. Spence and J.T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2). New York: Academic Press, 1968.

Bar-Hillel, M. The base rate fallacy in probability judgments. *Acta Psychologica*, 1980, *44*, 211-233.

Barclay, S., Beach, L.R., and Brathwaite, W.P. Normative models in the study of cognition. *Organizational Behavior and Human Performance.* 1971, *6*, 389-413.

Baron, J., Beattie, J., and Hershey J.C. Heuristics and biases in diagnostic reasoning II. *Congruence, Information, and Certainty, Organizational Behavior and Human Decision Processes*, 1988, *42*, 88-110.

Beach, L.R. and Mitchell, T.R. A contingency model for the selection of decision strategies. *Academy of Management Review*, 1978, *3*, 439-449.

Beach, L.R., Christensen-Szalanski, J., and Barnes, V. Assessing human judgment: Has it been done, can it be done, should it be done? (Chapter 3). In G. Wright and P. Ayton (Eds.), *Judgmental forecasting*, John Wiley & Sons Ltd., 1987, 49-62.

Bell, D.E. Components of risk aversion. In J.P. Brans (Ed.) *Proceedings of the Ninth IFORS Conference*, Amsterdam: North Holland, 235-242, 1981.

Bell, D.E. Regret in decision making under uncertainty. *Operations Research*, September-October 1982, *30*(5).

Bell, D.E. Disappointment in decision making under uncertainty. In D.E. Bell, H. Raiffa, and A. Tversky (Eds.), *Decision Making: Descriptive, Normative, and Prescriptive Interactions*, Cambridge University Press, 1988.

Bell, D.E., Raiffa, H., and Tversky, A. Descriptive, normative, and prescriptive interactions in decision making. In Bell, D.E., Raiffa, H., and Tversky, A. (Eds.) *Decision making: Descriptive, normative, and prescriptive interactions*, New York: Cambridge University Press, 1988.

Bellman, R.E. and Giertz, M. On the analytic formalism of the theory of fuzzy sets. *Information Sciences*, 1973, *5*, 149-156.

Ben Zur, H. and Breznitz, S.J. The effect of time pressure on risky choice behavior. *Acta Psychologica*, 1981, *47*, 89-104.

Berkeley, D. and Humphreys, P.C. Structuring decision problems and the `bias heuristic'. *Acta Psychologica 50*, 201-252. North Holland Publishing Co., 1982.

Beyth-Marom, R. and Fischhoff, B. *Direct measures of availability and judgments of category frequency* (Technical Report PTR-1042-77-3). Eugene, OR: Decision Research, 1977.

Birnbaum, M.H. Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology*, Spring 1983, *96*(1), 85-94.

Birnbaum, M.H. and Mellers, B.A. Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, 1983, *45*, 792-804.

Bransford, J.D., and Franks, J.J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, *2*, 331-350.

Brown, A.L. and Deloache, J.S. Skills, plans, and self-regulation. In R.S. Siegler (Ed.), *Children's thinking: What develops?* Hillsdale, NJ: Erlbaum, 1978.

Brown, R.V. Commentary, on "Implementing Decision Analysis," by H. Thomas. In I. Horowitz, (Ed.), *Organization and Decision Theory*. Norwell, MS: Kluwer Academic Publishers, 1989a.

Brown, R.V. Toward a prescriptive science and technology of decision aiding. *Annals of Operations Research*, 1989b, *19*, 467-483.

Brown, R.V., Kahr, A.S., and Peterson, C.R. *Decision analysis for the manager.* New York: Holt, Rinehart, and Winston, 1974.

Bruner, J.S. On perceptual readiness. *Psychological Review*, 1957, *64*, 123-152.

Centor, R.M., Dalton, H.P., and Yates, J.F. *Are physicians' probability estimates better or worse than regression model estimates?* Paper presented at the Sixth Annual Meeting of the Society for Medical Decision Making. Bethesda, Maryland, 1984.

Chase, W.G., and Simon, H.A. The mind's eye in chess. In W.G. Chase (Ed.), *Visual Information Processing*. NY: Academic Press, l973.

Cheng, P.W. and Holyoak, K.J. Pragmatic reasoning schemas. *Cognitive Psychology*, 1985, *17*, 391-416.

Cherniak, C. *Minimal rationality*. Cambridge, MS: MIT Press, 1986.

Chi, M., Feltovich, P., and Glaser, R. Categorization and representation of physics problems by experts and novices. *Cognitive Science*, l98l, *5*, l2l-l52.

Chi, M.T.H., Glaser, R. and Rees, E. Expertise in problem solving. In R. Steinberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, 1982, 7-75.

Christensen-Szalanski, J.J.J. Improving the practical utility of judgment research. In B. Brehmer, H. Jungermann, P. Lourens, and G. Sevón, (Eds.), *New Directions in Research on Decision Making*, North Holland: Elsevier Publishers, 1986.

Christensen-Szalanski, J.J.J. and Beach, L.R. The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologists*, 1984, *39*, 75-78.

Cohen, L.J. *The probable and the provable*. Oxford: Oxford University Press (Clarendon Press), 1977.

Cohen, L.J. Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences*, September 1981, *4*(3).

Cohen, L.J. The controversy about irrationality. *The Behavioral and Brain Sciences*, September 1983, *6*(3), 510-517.

Cohen, M.S. An expert system framework for non-monotonic reasoning about probabilistic assumptions. In L.N. Kanal and J.F. Lemmer, *Uncertainty in artificial intelligence: Machine intelligence and pattern recognition* (Vol. 4), Elsevier Science Publishers B.V., 1986, 279-293.

Cohen, M.S. Decision making "biases" and support for assumption-based higher-order reasoning. *Proceedings of the Fifth Workshop on Uncertainty in Artificial Intelligence*. Windsor, Ontario: August 18-20, 1989, 71-80.

Cohen, M.S. *Conflict resolution as a knowledge elicitation technique* (Working Paper). Reston, VA: Decision Science Consortium, Inc., 1990.

Cohen, M.S., Bromage, R.C., Chinnis, J.O., Jr., Payne, J.W., and Ulvila, J.W. *A personalized and prescriptive attack planning decision aid* (Technical Report 82-4). Falls Church, VA: Decision Science Consortium, Inc., July 1982.

Cohen, M.S., Laskey, K.B., and Tolcott, M.A. *A personalized and prescriptive decision aid for choice from a database of options (Revised)* (Technical Report 87- 18). Reston, VA: Decision Science Consortium, Inc., December 1987.

Cohen, M.S., Leddo, J.M., and Tolcott, M.A. *Cognitive strategies and adaptive aiding principles in submarine command decision making* (Working Paper). Reston, VA: Decision Science Consortium, Inc., October 1988.

Cohen, M.S., Schum, D.A., Freeling, A.N.S., and Chinnis, J.O., Jr. *On the art and science of hedging a conclusion: Alternative theories of uncertainty in intelligence analysis* (Technical Report 84-6). Falls Church, VA: Decision Science Consortium, Inc., December 1985.

Connolly, T. and Wagner, W.G. Decision Cycles. *Advances in information processing in organizations*, *3*, 183-205, JAI Press, Inc., 1988.

Daniels, N. Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy*, 1979, *76*, 256-282.

Davidson, D., Suppes, P., and Siegel, S. *Decision making: An experimental approach*. Stanford, CA: Stanford University Press, 1957.

Dawes, R.M. The robust beauty of improper linear models in decision making. *American Psychologist*, 1979, *34*, 571-582.

Dawes, R.M. and Corrigan, B.C. Linear models in decision making. *Psychological Bulletin*, 1974, *81*, 95-106.

Dawkins, R. Adaptationism was always predictive and needed no defense. *The Behavioral and Brain Sciences*, September 1983, *6*(3), 360-361.

de Finetti, B. Foresight: Its logical laws, its subjective sources. English translation in H.E. Kybert, Jr., and H.E. Smokler (Eds.), *Studies in Subjective Probability.*. NY: Wiley, 1964. (Original: 1937).

De Groot, A.D. *Thought and choice in chess*, 2nd ed.). New York: Mouton.

1965/1978.

deKleer, J. An assumption-based truth maintenance system. *Artificial Intelligence*, 1986, *28*, 127-162.

Dennett, D.C. Intentional systems in cognitive ethology: The "Panglossian paradigm" defended. *The Behavioral and Brain Sciences*, September 1983, *6*(3).

Doyle, J. A truth maintenance system. *Artificial Intelligence*, *12*(3). North Holland Publishing Co., 1979, *12*, 231-272.

Duhem, P. *The aim and structure of physical theory*. NY: Atheneum, 1962. (Originally published 1914.)

Ebbesen, E.G. and Konecni, V.J. On the external validity of decision-making research: What do we know about decisions in the real world? In T. Wallsten (Ed.), *Cognitive Processes in Choice and Decision Behavior.* Hillsdale, NJ: Erlbaum, 1980.

Edwards, W. Conservatism in human information processing. In B. Kleinmentz (Ed.) *Formal Representation of Human Judgment*. New York: Wiley, 1968.

Einhorn, H.J. Learning from experience and suboptimal rules in decision making. In T.S. Wallsten (Ed.), *Cognitive processes in choice and decision behavior*. Hillsdale, NJ: Lawrence Erlbaum and Associates, Inc., 1980.

Einhorn,, H.J., and Hogarth, R.M. Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, 1981, *32*, 53-88.

Einhorn, H.J., Kleinmuntz, D.N., and Kleinmuntz, B. Linear regression and process-tracing models of judgment. *Psychological Review*, 1979, *86*, 465-485.

Ellsberg, D. Risk, ambiguity, and the Savage axioms. *Quarterly Journal of Economics*, 1961, *75*, 643-649.

Fischhoff, B. Predicting frames. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1981.

Fischhoff, B. Debiasing. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* (pp. 422-444). New York: Cambridge University Press, 1982.

Fischhoff, B. and Beyth-Marom, R. Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 1983, *90*, 239-260.

Fischhoff, B., Slovic, P., and Lichtenstein, S. Subjective sensitivity analysis. *Organizational Behavior and Human Performance*, 1979, *23*, 339-359.

French, S. *Updating of belief in the light of someone else's opinion*. University of Manchester, April 1978.

Gardenfors, P. and Sahlin, N. Unreliable probabilities, risk taking and decision making. *Synthese*, 1982, *53*.

Gavelek, J. and Raphael, T.E. Metacognition, instruction, and the role of questioning activities (Chapter 3). In D.L. Forrest-Pressley, G.E. MacKinnon, and T.G. Waller, *Metacognition, cognition, and human performance: Volume 2 - Instructional practices*. Academic Press, Inc., 1985, 103-136.

Gigerenzer, G. and Murray, D.J. *Cognition as intuitive statistics*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1987.

Glaser, R. Expertise and learning: How do we think about instructional processes now that we have discovered knowledge structures? (Chapter 10). In D. Klahr and K. Kotovsky (Eds.), *Complex information processing: The Impact of Herbert A. Simon*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1989, 269-282.

Glymour, C. *Theory and evidence*. Princeton, NJ: Princeton University Press, 1980.

Glymour, C., Scheines, R., Spirtes, P., and Kelly, K. *Discovering causal structure: Artificial intelligence, philosophy of science, and statistical modeling*. New York: Academic Press, Inc., 1987.

Goldberg, L.R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 1968, *23*, 483-496.

Goodman, N. *Fact, fiction and forecast*. Indianapolis: Bobbs-Merrill (2nd edition), 1965.

Gould, S.J. and Lewontin, R. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society* (London), 1979, B205.581-98.

Greenwald, A. Pratkanis, A., Lieppe, M., and Baumgardner, M. Under what conditions does theory obstruct research progress? *Psychological Review*, 1986, *93*, 216-229.

Grice, H.P. Logic and conversation. In D. Davidson and G. Harman (Eds.) *The logic of rramman*. Encino: Dickenson, 1975.

Hammond, K.R. Judgment and decision making in dynamic tasks. *Information and Decision Technologies* 1988, *14*, 3-14.

Hempel, C.G. (1945) Studies in the logic of confirmation. Reprinted in *Aspects of scientific explanation*. Free Press, 1965.

Hoch, S.J. Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1985, *11*(4), 719-731.

Hoch, S.J., and Tschirgi, J.E. Cue redundancy and extra logical inferences in a deductive reasoning task. *Memory and Cognition*, 1983, *11*, 200-209.

Hogarth, R.M. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. *Psychological Bulletin*, 1981, *90*(2), 197-217.

Hogarth, R.M. and Makridakis, S. Forecasting and planning: An evaluation. *Management Science*, February 1981, *27*(2), 115-138.

Horwich, P. *Probability and evidence*. Cambridge University Press, 1982.

Huber, G.P. *Managerial decision making*. Glenview, IL: Scott Foresman, 1980.

Johnson-Laird, P. N., Legrenzi, P., and Legrenzi, M.S. Reasoning and a sense of reality. *British Journal of Psychology*, 1972, *63*, 305-400.

Jungermann, H. The two camps on rationality. *Decision Making Under Uncertainty*, R.W. Scholz (Ed.), North Holland: Elsevier Publishers, 1983.

Kadane, J.B., and Lichtenstein, S. *A subjectivist view of calibration* (Report 82-6). Eugene, OR: Decision Research, l982.

Kahneman, D. and Tversky, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 1972, *3*, 430-454.

Kahneman, D. and Tversky, A. On the psychology of prediction. *Psychological Review*, 1973, *80*, 237-251.

Kahneman, D. and Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica*, March 1979, *47*(2), 263-291.

Kahneman, D. and Tversky, A. On the study of statistical intuitions. *Cognition*, 1982, *11*, 123-141.

Kahneman, D., Slovic, P., and Tversky, A. (Eds.) *Judgment under uncertainty: Heuristics and biases.* NY: Cambridge University Press, 1982.

Kanal, L.N. and Lemmer, J.F. (Eds.). *Uncertainty in artificial intelligence: Machine intelligence and pattern recognition* (Vol. 4). Elsevier Science Publishers B.V. (North-Holland), 1986.

Keeney, R.L., and Raiffa, H. *Decisions with multiple objectives: Preferences and value tradeoffs.* NY: Wiley and Sons, l976.

Klayman, J. and Ha, Y-W. Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, 1987, *94*, 211-228.

Klein, G.A. Automated aids for the proficient decision maker. *IEEE Transactions on Systems, Man, and Cybernetics*, l980, 30l-304.

Klein, G.A. Do decision biases explain too much? *Human Factors Society Bulletin*, May 1989, *32*(5).

Koriat, A., Lichtenstein, S., and Fischhoff, B. Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, March 1980, *6*(2), 107-118.

Kuhn, T. *The structure of scientific revolutions*. Chicago, IL: University Press of Chicago, 1962.

Kuhn, D., Amsel, E., and O'Loughlin, M. *The development of scientific thinking skills*. Academic Press, Inc., 1988.

Kyburg, H.E. Bets and beliefs (Chapter 6). In P. Gardenfors and N.-E. Sahlin (Eds.), *Decision, probability, and utility: Selected readings*. Cambridge University Press, 1988. (Reprinted from *American Philosophical Quarterly* 1968, (5), 63-78.)

Larkin, J.H. *Problem solving in physics* (Technical Report). Berkeley: University of California, Group in Science and Mathematics Education, 1977.

Larkin, J.L. Enriching formal knowledge: A model for learning to solve textbook physics problems. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition*. Hillsdale, NJ: Erlbaum, 1981.

Larkin, J., McDermott, J., Simon, D.P., and Simon, H.A. Expert and novice performance in solving physics problems. *Science*, 20 June 1980, *208*, 1335-1342.

Leddo, J. Abelson, R.P., and Gross, P.H. Conjunctive explanations: When two reasons are better than one. *Journal of Personality and Social Psychology*, 1984, *47*(5), 933-943.

Leddo, J., Chinnis, J.O., Jr., Cohen, M.S., and Marvin, F.F. *The influence of uncertainty and time stress on decision making* (Technical Report 87-1). Falls Church, VA: Decision Science Consortium, Inc., January, 1987.

Lee, W. *Decision theory and human behavior*. New York: John Wiley & Sons, Inc., 1971.

Leon, M. and Anderson, N.H. A ratio-rule from integration theory applied to inference judgments. *Journal of Experimental Psychology*, 1974, *102*, 27-36.

Levi, I. *Hard choices: Decision making under unresolved conflict*. New York: Cambridge University Press, 1986.

Lichtenstein, S., Fischhoff, B., and Phillips, L.D. Calibration of probabilities: The state of the art to l980. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* NY: Cambridge University Press, l982.

Lindblom, C.E. The science of "muddling through." *Public Administration Review*, Spring 1959, 19 *2*, 79-88.

Lindley, D.V. Scoring rules and the inevitability of probability. *International Statistical Review*, 1982, *50*, 1-26.

Lopes, L.L. Doing the impossible: A note on induction and the experience of randomness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1982, *8*(6), 626-636.

Lopes, L.L. Observations: Some thoughts on the psychological concept of risk. *Journal of Experimental Psychology: Human Perception and Performance*, 1983, *9*(1), 137-144.

Lopes, L.L. *The rhetoric of irrationality*. Madison: University of Wisconsin. In revision, April 1988.

Lopes, L.L. and Oden, G.C. The rationality of intelligence. In E. Eells and T. Maruszewski (Eds.), *Rationality and Reasoning*. Amsterdam, Rodopi, in press.

Luce, R.D. *Individual choice behavior. A theoretical analysis*. New York: Wiley, 1959.

Luce, R.D. The choice axiom after twenty years. *Journal of Mathematical Psychology*, 1977, *15*, 215-233.

Macdonald, R.R. Credible conceptions and implausible probabilities. *British Journal of Mathematical and Statistical Psychology*, 1986, *39*, 15-27.

Makridakis, S. and Hibon, M. Accuracy of forecasting: An empirical investigation. *J. Royal Statistical Society, A*, 1979, *142*(2), 97-125.

March, J.G. Bounded rationality, ambiguity, and the engineering of choice. In D.E. Bell, H. Raiffa, and A. Tversky (Eds.), *Decision making: descriptive, normative, and prescriptive interactions*. Cambridge University Press, 1988.

Marks, R.W. The effect of probability, desirability, and `privilege' on the stated expectations of children. *Journal of Personality*, 1951, *19*, 332-351.

May, R.S. Inferences, subjective probability and frequency of correct answers: A cognitive approach to the overconfidence phenomenon. In B. Brehmar, H. Jungermann, P. Lourens, and G. Sevón (Eds.), *New directions in research on decision making*. North Holland: Elsevier Science Publishers, 1986.

Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision*. NY: McGraw-Hill, l975.

Montgomery, H. Decision rules and the search for a dominance structure: Towards a process model of decision making. P. Humphreys, O. Svenson and A. Vari (Eds.), *Advances in Psychology*. Amsterdam: North-Holland, l983.

Mosteller, F. and Nogee, P. An experimental measurement of utility. *Journal of Political Economics*, 1951, *59*, 371-404.

Nau, R.F. *A new theory of indeterminate probabilities and utilities*. Fuqua School of Business working paper #8609, 1986.

Newell, A. *The knowledge level* (Report No. CMU-CS-81-131). Pittsburgh, PA: Carnegie-Mellon University, 1981.

Niiniluoto, I. L.J. Cohen versus Bayesianism. *The Behavioral and Brain Sciences*, 1981. *4*(3), 349.

Nisbett, R. and Ross, L. *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1980.

Nisbett, R.E., Borgida, E., Crandall, R., and Reed, H. Popular induction: Information is not necessarily informative. In J.S. Carroll and J.W. Payne (Eds.), *Cognition and Social Behavior*. Hillsdale, NJ: Lawrence Erlbaum and Associates, 1976.

Noble, D., Truelove, J., Grosz, C., and Boehm-Davis, D. *A theory of information presentation for distributed decision making* (Final Report). Vienna, VA: Engineering Research Associates, October 15, 1989.

Payne, J.W. Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior and Human Performance*, 1976, *16*, 366-387.

Payne, J.W., Bettman, J.R., and Johnson, E.J. *The Adaptive Decision-Maker: Effort and Accuracy in Choice*. ONR Technical Report 89-1. Durham, NC: Duke University, The Fuqua School of Business, January 1989.

Pearl, J. *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan Kaufmann Publishers, Inc, 1988.

Pennington, N. and Hastie, R. Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1988, *14*(3), 521-533.

Peterson, C.R. and Beach, L.R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, *68*, 29-46.

Phillips, L., IPL/AMRD. Applications of inference theory and analysis of artificial intelligence to problems in intelligence analysis. *Intelligence Production Laboratory Conference Proceedings*. Rosslyn, VA: Analytic Methodology Research Division, Office of Research and Development, Central Intelligence Agency, November, 1982.

Pitz, G.F. Subjective probability distributions for imperfectly known quantities. In L.W. Gregg (Ed.), *Knowledge and cognition*. New York: Wiley, 1974.

Polya, G. *How to solve it*. Princeton, NJ: Princeton University Press, 1945.

Popper, K. *The logic of scientific discovery*. NY: Basic Books, 1959.

Quine, W.V.O. Two dogmas of empiricism. *From a logical point of view* (Second Edition, revised). NY: Harper and Row, 1961, 20-46.

Quine, W. *Word and object*. Cambridge, MA: MIT Press, 1960.

Raiffa, H. *Decision analysis: Introductory lectures on choices under uncertainty.* Reading, MA: Addison-Wesley, l968.

Rawls, J. *A theory of justice*. Cambridge, MA: Bellknap, 1971.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., and Boyes-Braem, P. Basic objects in natural categories. *Cognitive Psychology*, 1976, *8*, 382-439.

Savage, L.J. *The foundations of statistics*. New York: Dover Publications, Inc., 1972. (Original publication: 1954).

Schank, R.C., and Abelson, R.P. *Scripts, plans, goals and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum, 1977.

Schneider, S.L. and Lopes, L.L. *Reflection in preferences for multioutcome lotteries* (WHIPP 22). Madison, WI: University of Wisconsin, Department of Psychology, June 1985.

Scholz, R.W. Biases, fallacies, and the development of decision making. In R.W. Scholtz (Ed.), *Decision making under uncertainty*. Elsevier Science Publishers B.V. (North-Holland), 1983, 3-18.

Shafer, G. *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press, 1976.

Shafer, G. Jeffrey's rule of conditioning. *Philosophy of Science,* 1981, *48*, 337-362.

Shafer, G. Savage revisited. In D.E. Bell, H. Raiffa, and A. Tversky (Eds.), *Decision Making: Descriptive, normative, and prescriptive interactions*. Cambridge University Press, 1988.

Shafer, G. and Tversky, A. Languages and designs for probability judgment. In D.E. Bell, H. Raiffa, and A. Tversky (Eds.), *Decision making: Descriptive, normative, and prescriptive interactions*. Cambridge University Press, 1988, 237-265.

Shanteau, J. Cognitive heuristics and biases in behavioral auditing: review, comments and observations. *Accounting Organizations and Society,* 1989, *14*(1/2), 165-177.

Simon, H.A. A behavioral model of rational choice. *Quarterly Journal of Economics*, 1955, *69*, 99-118.

Simon, H.A. Theories of bounded rationality. In C.B. Radner and R. Radner (Eds.), *Decision and organization*, pp. 161-176. Amsterdam: North-Holland Publishing Company, 1972.

Simon, H.A. and Hayes, J.R. The understanding process: Problem isomorphs. *Cognitive Psychology*, 1976, *8*, 165-190.

Slovic, P. and Tversky, A. Who accepts Savage's axiom? *Behavioral Science*, 1974, *19*, 368-373.

Slovic, P., Fischhoff, B. and Lichtenstein, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, *28*, 1-39.

Smithson, M. *Ignorance and uncertainty: Emerging paradigms*. NY: Springer-Verlag, Inc., 1989.

Stich, S.P. and Nisbett, R.E. Justification and the psychology of human reasoning. *Philosophy of Science*, 1980, *47*, 188-202.

Sternberg, R.J. A componential approach to intellectual development (Chapter 9). In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Volume 1), Hillsdale, NJ: Lawrence Erlbaum Associates, 1982, 413-463.

Svenson, O. Process descriptions of decision making. *Organizational Behavior and Human Performance,* l979, *23,* 86-ll2.

Swets, J.A. Is there a sensory threshold? *Science*, 1961, *134*, 168-177.

Thagard, P. *Computational philosophy of science*. Cambridge, MS: MIT Press, 1988.

Thorngate, W. Efficient decision heuristics. *Behavioral Science,* 1980, *25*, 218-225.

Tolcott, M.A., and Marvin, F.F. *Reducing the confirmation bias in an evolving situation* (Technical Report 88-11). Reston, VA: Decision Science Consortium, Inc., August 1988.

Tolcott, M.A., Marvin, F.F., and Lehner, P.E. *Effects of early decisions on later judgments in an evolving situation* (Technical Report 87-10). Falls Church, VA: Decision Science Consortium, Inc., July 1987.

Tribe, L.H. Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, April 1971, *84*(6), 1329-1393.

Tversky, A. Additivity, utility, and subjective probability. *Journal of Mathematical Psychology*, 1967, *4*, 175-201.

Tversky, A. Elimination by aspects: A theory of choice. *Psychological Review,* l972, 79*(4),* 28l-299.

Tversky, A. and Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, *4*, 207-232.

Tversky, A. and Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, *185*, 1124-1131.

Tversky, A. and Kahneman, D. Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in Social Psychology*, Volume I. Hillsdale, NJ: Lawrence Erlbaum Assoc., Inc., 1980.

Tversky, A. and Kahneman, D. The framing of decisions and the psychology of choice. *Science*, January 20, 1981, Vol. 211, 453-458.

Tversky, A. and Kahneman, D. Evidential impact of base rates. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases.* NY: Cambridge University Press, l982.

Tversky, A. and Kahneman, D. *Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment*. Technical Report. Stanford University Department of Psychology, June 1983.

Tyszka, T. Contextual multiattribute decision rules. In L. Sjöberg, T. Tyszka, and J. Wise, (Eds.), *Decision Processes and Decision Analysis*. Doxa, Lund, 1980.

Tyszka, T. Simple decision strategies vs. multi-attribute utility theory approach to complex decision problems. *Praxiology Yearbook*, 1981, *2*, 159-172.

Ulvila, J.W. and Brown, R.V. Decision analysis comes of age. *Harvard*

Business Review, September-October, l982, l30-l40; also reprinted in *A sampling of quantitative methods for managers*, Harvard Business Review Book No. 13016, 3-14.

Wagenaar, W.A. Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin*, 1972, *77*, 65-72.

Wallsten, T.X. The theoretical status of judgmental heuristics. In R.W. Scholz (Ed.), *Decision making under uncertainty*. New York: North-Holland, 1983.

Wason, P.C. On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, 1960, *12*, 129-140.

Wason, P.C. Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 1968, *20*, 273-281.

Winterfeldt, D. v., and Edwards, W. *Flat maxima in linear optimization models*. (Technical Report 011313-4-T). Ann Arbor, MI: University of Michigan Engineering Psychology Laboratory, Institute of Science and Technology, November 2, 1973.

Winterfeldt, D. von, and Edwards, W. *Error in decision analysis: How to create the possibility of large losses by using dominated strategies*. University of Southern California Social Science Research Institute (Research Report 75-4), April 1975.

Winterfeldt, D. von, and Edwards, W. *Decision analysis and behavioral research*. Cambridge University Press, 1986.

Woodcock, A.E.R., Cobb, L., Familant, M.E., and Markey, J. *I&W Applications of Catastrophe Theory*. Technical Report RADC-TR-88-212. Rome, NY: Synectics Corporation, October 1988.

Wright, P. The harassed decision maker: Time pressures, distractions and the use of evidence. *Journal of Applied Psychology*, 1974, *59*(5), 555-561.

Wright, J.C., and Murphy, G.L. The utility of theories in intuitive statistics: The robustness of theory-based judgments. *Journal of Experimental Psychology: General*, 1984, *113*, 301-322.

Yates, J.F. External correspondence. Decomposition of the mean probability score. *Organizational Behavior and Human Performance*, 1982, *30*, 132-156.

Zadeh, L.A. Fuzzy sets. *Inf. and Contr.,* 1965, *8*, 338-353.

Zadeh, L.A. Review of Shafer's, A mathematical theory of evidence. *AI Magazine*, l984, *5*(3), 8l-83.

Zimmer, A. Verbal versus numerical processing of subjective probabilities. In R.W. Scholz (Ed.), *Decision making under uncertainty*. North Holland: Elsevier Science Publishers, 1983.