Another Look at Reasoning Experiments: Rationality, Normative Models and Conversational Factors

ALEXANDER TODOROV

In the sixties the picture of human reasoning was clear and optimistic. People seemed to think according to statistical rules when solving inductive problems (Peterson & Beach, 1967)[1] and according to logical rules when solving deductive problems (Henle, 1962). The same positive image was accepted among the researchers in the field of causal attribution (Kelley, 1967). People detected covariations across pre-defined logical dimensions and made the right attribution.

In the seventies and eighties, things changed dramatically (Lopes, 1991). People no longer seemed to be so rational or logical (Evans, 1989; Kahneman, Slovic & Tversky, 1982; Nisbett & Ross, 1980). Since the Kahneman & Tversky seminal article (1973) *On the psychology of prediction*, the biases in human judgments have rapidly proliferated. Jussim (1991), criticizing this trend, has listed more than 15 phenomena of bias of social inference, only partially overlapping with the original 13 biases proposed by Tversky & Kahneman (1974). The idea underlying the biases approach was that they are systematic deviations from logical or statistical normative models. The research on biases portrayed a picture with "bleak implications for human rationality" (Nisbett & Borgida, 1975, p. 935). People seemed to think according to heuristics or rules of thumb (Nisbett & Ross, 1980).

However, the tension between these two views, the "statistical person" and the "heuristic person," still exists. In the recent years, many researchers, drawing upon Grice's conversational maxims (Grice, 1975), have addressed another possible determinant of human judgments (see for reviews Hilton, 1995; Schwarz, 1994, 1996). This recent research trend has put the problem in a new perspective. Many of the alleged biases inherent in human judgment turn out to be effects of the insensitivity of the experimenters to the conversational character of the experimental situation.

Other researchers (e.g. Cosmides & Tooby, 1996; Gigerenzer, 1991a; Gigerenzer & Hoffrage, 1995) have shown that when the information is presented

in frequency formats, lay judgments match the predictions of normatively correct probabilistic models. In a similar vein, if certain statistical assumptions, e.g. random sampling, are preserved in the experimental procedures, i.e. people are made to believe that the information was really randomly sampled, lay judgments are good approximations of the Bayesian probabilistic model (Gigerenzer *et al.*, 1988).

Still, the main questions about reasoning are with us. What is at the bottom of reasoning processes: an inferential system based on deductive rules, based on Bayesian type probabilistic rules, or rules of thumb? What is the contribution of the conversational factors to this inferential process? Do they act at the level of representations, i.e. the selection of information, or are they related to a different type of inference rules? How one will answer these questions depends on one's implicit assumptions.

The aim of this paper is to provide a general framework for evaluation of studies demonstrating biases in human reasoning and, hence, flaws in human rationality. All studies on human reasoning face two main problems: the selection of a normative model for assessment of subjects' performance and account of the subjects' interpretations of the reasoning task. In other words, the normative model should be appropriate for the reasoning task at hand and the subjects' interpretations of the task should match the experimenter's model of the task interpretation. If any of these problems is not acknowledged, the findings of the studies will suffer from indeterminacy.

This article is organized in three parts. In the first part, the assumptions underlying much of the research on human reasoning are considered. Two main criteria for evaluation of reasoning studies are developed and outlined. The first criterion is conceptual and concerns the selection of inferential normative models against which subjects' performance is assessed. The second criterion is empirical and concerns the experimental designs used in the reasoning studies. It includes the effects of conversational factors on subjects' performance in experimental situations. In the second part of the paper, these criteria are applied to studies on two particular tasks from the fields of deductive and inductive reasoning, correspondingly the Wason selection task and tasks related to the base-rate fallacy. In the final part of the article, the implications of the present account for research on reasoning are discussed.

CRITERIA FOR EVALUATION OF STUDIES ON REASONING: TWO WAYS TO SAVE RATIONALITY

Traditionally, reasoning research has been divided into two main areas, deductive and inductive reasoning (e.g. Evans, 1989; Evans *et al.*, 1993). In view of the different normative systems underlying both types of reasoning this division makes perfect sense. However, both fields of research face similar conceptual

and empirical problems. Besides, the normative "distinctiveness" of the two types of reasoning tasks, deductive and inductive, is not directly translatable in their psychological "distinctiveness." To put it in another way, the existence of different normative systems does not guarantee the plausibility of different psychological systems.

The Normative Models: The Conceptual Problem

Assessing subjects' performance in reasoning tasks requires the availability of a normative model. A description of an inference as correct or incorrect is always relative to a normative system. As Evans (1989) noticed the use of accuracy measures in memory and perception research is a common practice. However, the process of assessment of human performance in reasoning tasks is not the same as in the latter two fields. Take, for instance, the assessment of intuitive statistical judgments in the areas of descriptive and inferential statistics (e.g. Peterson & Beach, 1967). In the former case, the criterion is accuracy. The process consists of a simple comparison between the subject's judgment and a real number, say a sample size or mean. Correspondingly, the judgement may be described either as accurate or as inaccurate. In the case of inferential statistics, the subject's judgment may only be described as more or less optimal. This description is determined by the degree to which the judgment corresponds to the inference prescribed by a statistical model.[2]

   An error of inference is a deviation from a prediction derived from a certain normative system. Correspondingly, a bias is defined as a systematic deviation from a certain normative system. A bias is demonstrated when an error is consistently repeated in normatively similar tasks. In these definitions, a normative model is presupposed and the subject's performance is assessed against it. To draw an implication for human rationality on this basis, one first needs an epistemic justification of the normative model used.

   However, the implicit assumption in most of the reasoning studies (e.g. Nisbett & Ross, 1980) is that there is only one correct and uncontroversial normative model against which to assess subjects' judgments. But in the area of statistics we have, at least, three competing systems: Fisher's null hypothesis testing, Neyman & Pearson's testing of a priori specified alternative hypotheses, and the Bayesian statistics (Gigerenzer & Murray, 1987). The first two are related to a frequentist interpretation of probability whereas the third is related to a subjectivist interpretation. The former treats the probability of an event as its long run frequency or the proportion of the time it occurs in a long sequence of trials. The latter treats the probability of an event as the subjective degrees of belief that the event will occur or have occurred. The debate between these two interpretations of probability is the most heated, but they are not the only ones. There are at least three other interpretations (Cohen, 1982). In the system

of deductive logic, though much more developed than inductive logic, one can also find different logical subsystems, e.g. modal logic, or even many valued logics which accept more than two truth values (Rossere & Turquette, 1952). The upshot is that one performance interpreted in different normative frames of reference would bear different implications for human rationality.

Besides their multiplicity, the normative systems should be intelligible to their lay "users." L. Cohen (1981) has argued that all normative systems are ultimately based on lay intuitions. He has drawn an analogy with Chomsky's distinction between competence and performance. The reasoning intuitions are similar to the intuitions of grammatical well-formedness. Therefore, in the final account, the normative systems for evaluation of human reasoning are constrained by what is intuitively acceptable to lay adults. Formal deductive systems depend on lay interpretations of such logical particles as "if," "and," "or," "some," etc. The different interpretations of probability satisfy the formal axioms of classical calculus but have different semantics. To reveal which conceptions of probability are operative in a given case, one has to be sure which judgments of probability are intuitively acceptable. On this basis, Cohen has classified all reasoning studies demonstrating irrational judgments into four categories: studies of cognitive illusions, tests of intelligence or education, misapplications of appropriate normative theory, and applications of inappropriate normative theory. In the former two cases, though human judgments have been shown to be incorrect, the studies hardly reveal much about human rational competence. In the latter two cases, human judgments are misleadingly interpreted as incorrect or the fallacy is attributed to the experimenter.[3] Even if one is unwilling to accept Cohen's argument for the dependency of the normative systems on lay adult judgments, one must be concerned with the issue of their interpretation.

Other researchers in the area of personality judgments and social perception (Funder, 1987, 1993, 1995; Jussim, 1991; see also Fiske, 1992, 1993) have pointed out that human judgments are accurate or good-enough in the real world. Funder (1987) has compellingly argued that researchers in the "error-bias" tradition fundamentally confuse "error" and "mistake." An "error" is a deviation from a model, the experimenter's standard, whereas a "mistake" is an incorrect judgment in the real world. He makes an illuminating analogy with visual perception research. Perceptual illusions have been a topic of interest in this tradition for years, but they have not been viewed as mistakes of the perceptual system. They have illustrated only its adaptive workings[4]. In three-dimensional space, as opposed to the presentation of artificial stimuli in two dimensional space, the illusions are the right answers (for the role of wider social context in experimental situations see Orne, 1962). The upshot is that there are different criteria for "errors," produced in a laboratory situation, and "mistakes," produced in real life. Therefore, the generalization from the former to the latter is unjustified.

Although the normative models are attractive for modeling human performance (Gigerenzer, 1991b), it is questionable how realistic their computational

demands for the human cognitive system are (Cherniak, 1986; Oaksford & Chater, 1991; Simon, 1957). Anderson (1990) has recently developed the notion of rational analysis. He makes a distinction between normative and adaptive rationality. A behavior is rational if it is adaptive to its environment. Applied to reasoning studies, this means that the optimality of human judgments is relative to an adaptive criterion. The important thing to begin with is the structure of the environment. The mathematical model should match this structure (see also Einhorn & Hogarth, 1981; Gigerenzer, 199a; Gigerenzer *et al.*, 1991; Simon, 1957). A second important consideration is the constraints on the human cognitive system, e.g. the limited capacity of working memory. The emphasis is on the construction of models that fit the data of human performance. In other words, a priority is given to human performance.

To summarize, we have three arguments against the "error-bias" approach to human inference: epistemic, consisting of multiplicity of normative models and dependency of their interpretaion on lay adult judgments; ecological, emphasizing the different criteria for correctness of a certain judgment in the real world and the laboratory; and evolutionary or adaptive, assuming the adaptive rationality of people given the constraints on the environment and the cognitive system. The arguments are overlapping and based on the questioning of the normative models used in reasoning studies. The implication is that we need a rational justification of the normative models[5] against which human rationality is checked. This is our first criterion for evaluation of reasoning studies.

> Criterion 1: Given that any bias is a deviation from a normative system, it is always possible either to find an alternative normative system relative to which the alleged bias is an optimal strategy of inference or to find violation in the application of the normative model used in the study. Therefore, any theory of reasoning should provide a justification of its choice and application of a normative theory. This justification process includes, at least, two components:
>
> 1.1 A priori conceptual justification of the appropriateness of the normative model for the specific task.
>
> 1.2 A demonstration that the bias is also a deviation from a reasonable range of alternative normative theories mapping the task domain.

This criterion should be considered as consisting of two interrelated parts: evaluation of reasoning studies and development of normative models for evaluation of human judgments. The evaluation part answers the questions: Is there a conceivable alternative system accounting for the results? and Are the assumptions underlying the normative model valid in the concrete situation? Whereas the former is rather a conceptual question the latter could be tested empirically (e.g. Gigerenzer *et al.*, 1988). This is the stringent part of the criterion. Its second part concerns the development and justification of normative models. The model should meet certain assumptions, e.g. its structure should correspond to the structure of the environment, more specifically the task environment including its social aspects. For instance, one of the alleged biases in human reasoning is that lay judgments are

insufficiently regressive (Kahneman & Tversky, 1973; Nisbett & Ross, 1980), i.e. people tend to give extreme predictions for future events. However, this would be a bias in a stable environment. If the underlying process is changing, the regressive predictions are suboptimal and the extreme predictions are closer to optimal (Einhorn & Hogarth, 1981). The model also should take into account the constraints on the human cognitive system. In the present paper, we are concerned only with the evaluation part of the criterion.

The criterion is stringent but not impossible to meet. First, the accepted normative systems, though multiple, are finite. Second, the fit of the data to an alternative normative model does not guarantee in itself that this is the correct model of human inference, but it generates new research questions. The criterion also fits what Cohen (1982) calls "the norm extraction method" as opposed to "the preconceived norm method." In the former, the investigator is trying to discover what conceptions of probability people endorse in solving a particular problem, whereas in the latter she imputes a predetermined theory to them. Similarly, in the rational analysis (Anderson, 1990) the priority is given to the observed human performance[6] (see also Gigerenzer, 1991a).

The Conversational Factors: The Empirical Problem

A normative model of human reasoning should match the rational competence of lay adults. Besides the distinction between competence and performance which is related to the epistemic problems of selection of normative models, there is another important distinction between representation and process. The latter is related to the use of information in the experimental situation. What may happen during a reasoning experiment is that the subjects may form representations based on the information provided in the experiment different from the experimenter's representations.[7] These subjects' representations are used in the subsequent judgment. In other words, the inferential process itself may be rational and with accordance to the assumed normative model. The trouble is not at the level of computation, but at the level of selecting of information for the computation. This is precisely the kind of logic M. Henle (1962) used in her explanation of the findings of fallacies in deductive reasoning.

The implication is that the research should focus on factors contributing to the differential selection of information. The latter process is related to the communication between the experimenter and the subject. Grice (1975) has convincingly argued that everyday conversation is guided by the cooperative principle. Conversation is viewed as a rational activity in which the speaker and the listener follow conversational maxims. These are the maxims of manner, of relation, of quality, and quantity. In short, both participants try to be clear, relevant, truthful, and informative. The violations of the maxims are interpreted as implicatures of the conversation. In this sense, the cooperative principle and

the maxims provide a frame of reference for interpretation of the ongoing utterances. The Gricean perspective provides us with an inferential theory of communication (Sperber & Wilson, 1995).

Though the experimental situation is much more constrained than everyday conversation, there is no reason to assume that the conversational maxims are not operative (Clark & Schober, 1991; Hilton & Slugoski, 1986; Schober & Conrad, under review; Schwarz, Strack & Mai, 1991; Strack & Schwarz, 1992; Strack, Martin & Schwarz, 1988; Strack, Schwarz & Wänke, 1991). Indeed, on closer inspection many of the biases in human reasoning turn out to reflect violations of the maxims that govern everyday conversation on the side of the experimenter (Adler, 1984; Hilton, 1990, 1995; Schwarz, 1994, 1996). However, the conversational character of the experimental situations has been largely ignored in the studies on human judgment.

One example will suffice. In a typical experiment demonstrating children's lack of mastery of the concept of number conservation (Piaget, 1952; Piaget & Inhelder, 1969), children are shown two rows with an equal number of objects. They are asked whether the rows have the same number of objects. They answer correctly "yes." Then the experimenter rearranges the objects in one of the rows, extending its length, and asks the children the same question. They answer wrongly "no." In an unusual replication (McGarrigle & Donaldson, 1974) instead of an experimenter rearranging the objects of one of the rows, a "naughty teddy bear" appears and tries to "spoil the game." The children answer correctly "yes" to the second question in the "bear" condition. This experiment shows that when the intentionality assumption which underlies the conversational maxims is precluded, the judgments are much closer to the presumed normative standard (see below Schwartz *et al.*, 1991 for the effect of a similar manipulation on the base rate fallacy). The act of rearranging the objects in the "bear" condition is rendered irrelevant to the normative question. However, in the normal condition, the "intentional" experimenter, the act is perceived as relevant. It should be, otherwise why would the same question be asked?

In sum, in any experimental situation the maxims of everyday conversation are operative. As a result, information, which is considered as irrelevant to the normative judgment at hand by the experimenter, is perceived as relevant by the subject. The effect is that the relevant "irrelevant" information changes the form of the premises and correspondingly the subsequent judgment. This gives our second criterion for evaluation of reasoning studies.

> Criterion 2: Given that any bias is produced in a certain experimental situation, it is possible to find an alternative explanation according to which the alleged bias is a result of the cooperative, in the sense of Grice (1975), behavior of the subjects. Therefore, any study on reasoning must acknowledge the role of the conversational factors by providing the following:
>   1.1 A model of the comprehension processes involved in the reasoning task or a set of explicit assumptions.
>   1.2 An experimental analysis of the subjects' interpretation of the task.

Notice that this criterion does not state that it is *always* possible to find an alternative explanation. The problem, however, is empirical and should not be ignored. Stated differently, the criterion postulates that there are two normative systems or two types of constraints operating in the experimental situation. These two models are related to what might be called conversational relevance and normative relevance of the information provided in the study. By analogy with the first criterion, the separation between normatively relevant, whether logically or statistically, and normatively irrelevant information is relative to a certain interpretation in the same way as a bias is relative to a certain normative system.

This logic may be illustrated with the so called dilution effect (Nisbett & Ross, 1980; Nisbett *et al.*, 1981; Zukier, 1982). The dilution effect, one of the many alleged biases, is the influence of nondiagnostic information on lay judgments. For instance, in Nisbett *et al.* (1981, Exp. 4 & 5) subjects rated the likelihood that a target person was a "known child abuser." They were presented with diagnostic information about the target, e.g. "He is aroused by sadomasochistic sexual fantasies," and nondiagnostic information, e.g. "He manages a hardware store" (the diagnocity of the information was determined in a pretest question-naire). The important finding was that the presence of nondiagnostic information influenced the subjects' judgments, i.e. "diluted" the initial extreme predictions based on diagnostic information. Moreover, increasing the pieces of nondiagnostic information increased the effect. These findings are interpreted as indicating a bias, specifically, as a result of the representativeness heuristic (see below). Compare this interpretation with the communicative principle of relevance of relevance theory. "Every act of ostensive communication communicates a presumption of its own optimal relevance" (Sperber & Wilson, 1995, p. 158). As Sperber, Cara & Girotto point out "whether or not the presumption of relevance is warranted, whether or not it is accepted, the very fact that it accompanies an utterance helps determine the utterance's intended interpreta-tion." (1995, p. 50). In other words, the mere inclusion of the normatively inappropriate information makes it conversationally relevant (see also Tetlock *et al.*, 1996). As Schwarz puts it "if the experimenter presents it, I should use it" (1994, p. 128).[8]

<center>APPLICATION OF THE CRITERIA: TWO CASE STUDIES</center>

Two particular phenomena were selected for an illustration of the present argument, the failure to use modus tollens in the Wason selection task and the underutilization of prior probabilities in inductive reasoning tasks. Both phenomena have a long history of investigation. Both have produced a large body of research and still are at the focus of research attention. Both reveal heavy fallacies of human judgments: in the case of the Wason selection task for

an inferential deductive system and in the case of prior probabilities for an inferential inductive system.

Deductive Reasoning: The Case of the Wason Selection Task

In a typical Wason selection task (Wason, 1968), subjects are presented with four cards. Two of the cards have letters on their front side and two have numbers. The subjects are told that the cards with letters have numbers on their back side and, correspondingly that the cards with numbers have letters. They are presented with a rule of the type "*if p, then q,*" say, if there is a vowel on one side (p), there is an even number on the other side (q). An example of the four cards is A (p), K (not-p), 4 (q), 7 (not-q). The task of the subjects is to select only those cards which need to be turned over in order to determine the truth or falsity of the rule. The normatively correct solution is to select the cards p and not-q. Both cards could falsify the rule. On the other hand, the cards not-p and q do not contribute anything to the solution. The not-p card does not provide any information relevant to the rule. The q-card could only confirm the rule but not falsify it. For example, if on its other side there is not-p, it will not show anything specific about the rule, "*if p, then q.*" In terms of deductive logic the solution requires the application of modus tollens: "*if p, then q,*" "*not-q, therefore not-p.*"

However the subjects' performance is far from this normative ideal (Johnson-Laird & Wason, 1970; Wason & Johnson-Laird, 1972). Subjects consistently select the p and q cards most often. The second preferred choice is only the p card, the third is the p, q, and not-q cards, and the least preferred choice is the normative correct one, the p and not-q cards. The finding is robust across a variety of tasks using abstract materials (Evans, 1989; Evans, Newstead & Byrne, 1993; Johnson-Laird & Byrne, 1991).

Initially, the Wason selection task (Wason & Johnson-Laird, 1972) was designed as a task tapping the process of hypotheses testing. From this point of view, the subjects' performance is interpreted as a confirmation bias. Subjects look for instances confirming the hypothesis, i.e. the rule, but avoid instances falsifying it (but see Fischhoff & Beyth-Marom, 1983 and Klayman & Ha, 1987 for an alternative interpretation of the confirmation bias). However, it turned out that in certain versions of the task the subjects' choices match the normative choices. The first factor eliciting correct performance was the use of negation rule of the type, "*if p, then not q*", e.g. "if there is an A on one side of the card, then there is not a 3 on the other side of the card" (Evans, 1972; Evans, 1989; Evans & Lynch, 1973). In this version of the task, the subjects predominantly select the logically correct p and q cards, i.e. the A and 3 cards in the example. Evans explained these findings as reflecting a matching bias. Subjects select only those cards which are explicitly mentioned in the rule. In this sense, their correct

performance is incidentally due to the form of the rule, not to a correct inferential use of modus tollens.

The second factor eliciting correct responses is related to the logical nature of the task. Johnson-Laird, Legrenzi & Legrenzi (1972) demonstrated this in a task with four envelopes. The rule was "If a letter is sealed, then it has a 5d stamp on it" and the subjects' task was to select those envelopes which need to be turned over to decide whether the rule is violated. The initial interpretation was that these results were due to the familiarity of the task content (Wason & Johnson-Laird, 1972) and after unsuccessful replications the familiarity of the experience depicted in the task (Griggs & Cox, 1982), but later studies (e.g. Manktelow & Over, 1991) demonstrated that the improvement in subjects' performance has to do with the logical nature of the task not with its content. What matters is not the familiarity of the material, the content of the rule, but its interpretation as a social contract rule. In other words, rules with unfamiliar content, e.g. "If a man eats cassava root, then he must have a tattoo on his face" (Cosmides, 1989), interpreted as social contract rules elicit "correct" responses. The tasks which elicit a correct performance, independent of the familiarity of their content, are deontic tasks. They state not what people are doing but what they *should* be doing. In such tasks the subjects look for evidence of violation of the rule (Cheng & Holyoak, 1985) or cheating (Cosmides, 1989; Gigerenzer & Hug, 1992), not of its truth.

Generally, the results of the Wason task studies throw serious doubts on the rationality of the human inferential system. In the two cases in which the performance has been optimal, their reasons have been incidental, a matching bias and a different logical task.

Criterion 1: The Normative Model. Imagine a more "real life situation" of card selection: subjects are presented with, say, 52 cards instead of four cards. They have to check the same type of rule "if p, then q." However, the cards with q on the front side are few, say 5, whereas the card with not-q are many, say 30. If we follow the normative logic how many cards do we have to turn over to check the rule? This logic is similar to Hempel's raven paradox (Hempel, 1965; see also Nickerson, 1996): if one tests the hypothesis "if it is a raven, then it is black," a much better strategy is to inspect ravens and check whether they are black rather than to inspect non-black things and check whether they are ravens or not. This line of reasoning may be called the rarity assumption, i.e. properties that are in causal relations in the environment are rare. Another way of framing the assumption is to say that the number of positively defined classes is smaller than their negative counterparts (e.g. Evans & Over, 1996; Nickerson, 1996; see also Kirby, 1994 for the effects of increasing the size of the p-set relative to the not-p set on the not-q card selection). This assumption is at the core of the rational analysis of the Wason task made by Oaksford & Chater (1984).

Remember that the selection task was introduced as a laboratory version of hypothesis testing. The normative logic of the selection/hypothesis testing task derives from Popper's falsificationist philosophy of science (Popper, 1959, 1979). According to Popper, the task of the scientists is to construct experiments which could falsify the theory. Hence, the correct choice in the selection task always contains the non-q card. This is the only card, except the p-card, which could falsify the rule, i.e. the hypothesis. However, besides the problems of the Popperian model to account for scientific practice (e.g. Kuhn, 1970; Lakatos, 1978), there is an alternative model of hypothesis testing based on a Bayesian probabilistic model of confirmation (Horwich, 1982). This approach provides a justification of the intuitions that the choice of a certain experiment is better than another if the experiment has a greater informational value for a decision between alternative hypotheses. In other words, the task is to construct such experiments that decide which hypothesis is more corroborated.

Oaksford & Chater (1994) have adopted this alternative approach of inductive hypothesis testing. They have reviewed and reassessed the studies on the selection task against a Bayesian normative model of optimal data selection. The key assumption of their model is the rarity assumption (see also Oaksford & Chater, 1996) which was mentioned above. On the basis of the latter and the Bayesian model of data selection, Oaksford & Chater have modeled the performance data from the Wason task. The alternative hypotheses in the Wason task are that the dependency "if p, then q" holds and that p and q are independent. In Oaksford & Chater's model, the subjects compute the expected gain information, which is the difference between the uncertainty before the card selection and the uncertainty after it about the validity of the dependency hypothesis, of the selections of the cards they made. In view of the rarity assumption, stated in terms of the model the subjects assume that the probabilities of p and q are low, the order of expected information gain (Ig) from the cards is as follows: Ig (p) > Ig (q) > Ig (not-q) > Ig (not-p). Indeed, this order reflects the order of the preferred choices of the subjects. Oaksford & Chater have reanalyzed 34 studies reporting an affirmative abstract version of the task. In all cases the model nicely fits the data. In cases where the rarity assumption is not preserved, the order of the cards in terms of expected information gain changes, e.g. when the probability of p is large the informativeness of the not-q card increases compared to the informativeness of the q-card. In such cases, the subjects' cards selections reflect this changed order.

With minor changes, the model also accounts for the performance in studies using a negation rule paradigm and deontic task studies. For instance, the performance in deontic task studies is related to rule use, not to rule testing. Correspondingly, the computations are in terms of expected utilities based on the different perspectives, e.g. enforcer of the rule or actor, in the deontic selection task (see Gigerenzer & Hug, 1992; Manktelow & Over, 1991). On the one hand, we have one parsimonious model accounting for all versions of the task. On the other hand, this model tells a different story about human rationality.

Criterion 2: The Conversational Factors. Consider in what way the Wason task is a deductive task. Strictly speaking, to be an unambiguous deductive task, subjects should be presented with the rule "*if p, then q*" as a major premise and each of the cards as minor premises, constituting four types of syllogisms. For example, they may be presented with a rule "if there is a vowel on one side (p), there is an even number on the other side (q)" and a card showing an odd number on its front side (not-q) and then asked "what should be on the other side of the card?" As Sperber, Cara & Girotto (1995) have pointed out the Wason task is deductive only by its optimal logical solution, but not by its instruction. If this is the case, then what is crucial is the subjects' interpretation of the task.

Sperber, Cara & Girotto (1995) have recently reanalyzed the Wason selection task in terms of relevance theory (Sperber & Wilson, 1995). At the core of the latter is the communicative principle of relevance: every utterance conveys a presumption of its own relevance. Sperber and Wilson (1995) define the relevance of a piece of information to an individual as guided by two factors, the cognitive contextual effects of the information and the processing effort required for it. Sperber, Cara & Girotto (1995) have argued that the selection task is simply a task of selection, not of hypothesis testing, guided by considerations for relevance.

In the Wason task, the subjects are presented with a general conditional rule ($\forall x$, $Px \rightarrow Qx$), i.e. for any item x, if x has the feature P, x has the feature Q. This form of the rule is not directly testable. In order to test the rule, the subjects have to derive testable consequences from it. Unlike Evans' heuristics approach (1984, 1989), Sperber *et al.* (1995) have argued that subjects' understanding of the task is driven by inferential comprehension processes. There are three interesting cases of interpretation of the rule which derive from ordinary communication. The first is a biconditional interpretation, in which the rule is interpreted as implying its converse ($\forall x$, $Qx \rightarrow Px$), e.g. a card with an even number (q) on one of its sides must have a vowel (p) on its other side. The second case is an interpretation of the rule as existential quantified conjunction ($\exists x$, $Px \& Qx$), i.e. there are items x that have the feature P and the feature Q, e.g. there are cards which have vowels on their one side and even numbers on their other side. These two interpretations are logically incorrect and lead to a predominant selection of the p and q cards, but they are admissable in ordinary discourse. For instance, a general conditional rule that does not imply specific instances will be irrelevant in everyday conversation. The third and logically correct case is an interpretation of the rule either as entailing (not $\exists x$, $Px \&$ not-Qx), there are no items of x that have the feature P and the feature not-Q, e.g. there are no cards which have vowels on their one side and odd numbers on their other side, or as contradicting ($\exists x$, $Px \&$ not-Qx), there are items of x that have the feature P and the feature not-Q, e.g. there are cards which have vowels on their one side and odd numbers on their other side. As Sperber, Cara & Girotto (1995) have pointed out these two interpretations are not representationally identical though they are logically equivalent. In the third case, the

interpretation of the rule will lead to the "correct" choices of p and not-q cards. Hence, to elicit correct responses of the Wason task one should make the subjects interpret the rule as a denial of occurrence of p-and-not-q cases. In normal circumstances, this interpretation would be the least accessible, because it contains two negations. However, making it readily accessible is not impossible if one varies the expected cognitive effects and the processing effort of the task information.

In abstract versions of the Wason task, the expected cognitive effects are almost completely lacking. In such cases, the subjects' comprehension is driven by considerations for least processing effort. In thematic versions of the task, however, the expectations for relevance may be increased either by introducing a specific perspective which the subjects are instructed to adopt, e.g. in deontic versions of the task, or by making the rule uttered by some of the characters in the story preceding the rule.

This is the Sperber, Cara & Girotto (1995) recipe for constructing an easy selection task: on the effort side of the task, select a pair of simple features p and q (e.g. p is "being of working age," q is "has a job," Exp. 3, relevance condition) such that the complex feature p-and-not-q (e.g. unemployed) is easier to represent, or could be made, than the complex feature p-and-q; on the effect side, create a context where knowing whether there are p-and-(not-q) cases would have greater cognitive effects than knowing whether there are p-and-q cases (e.g. the presence of unemployed young people in the context of a claim that there is no unemployment in a particular country); present the rule "if p, then q" in a pragmatically felicitous manner (e.g. the rule "if a person is of working age then this person has a job" is uttered by a reigning Prince who claimed that there is no unemployment and social problems in his country, notice that the rule is uttered as a factual statement). Varying the cognitive effects and the processing effort of the content of the tasks they used, Sperber, Cara & Girotto (1995) have demonstrated in four experiments that the subjects' performance is guided by considerations for relevance.

In the light of the "easy" descriptive selection task, the distinction between abstract descriptive versions and easy deontic versions becomes problematic (see also Johnson-Laird & Byrne, 1992; Kirby, 1994). The relevance account also explains the performance in deontic versions of the task and in tasks using a negative consequent rule, "if p, then not-q." Unlike in the standard version of the task with a rule "if p, then q," the most easily inferable interpretation of "if p, then not-q" is that it contradicts the occurrence of p-and-q cases. More formally, the general conditional statement $(\forall x, Px \rightarrow \text{not-}Qx)$ contradicts the assumption $(\exists x, Px \& Qx)$. Correspondingly, in that case, people predominantly choose the "correct" p and q cards. In deontic versions of the task, the most accessible interpretation is that the rule forbids the occurrence of p-and-not-q cases, i.e. its violations. In both cases, the use of a negative consequent rule and deontic tasks, the crucial condition for logically correct performance, that is

interpretation of the rule as a denial of the occurrence of p-and-not-q cases, is readily available.

In a similar vein, Liberman & Klar (1996) have demonstrated that a cheating version and perspective change in deontic versions of the task are neither sufficient nor necessary to elicit "correct" performance. What is important is the subjects' understanding of the task. They have showed that when the deontic tasks in the experiments of Gigerenzer & Hug (1992) and Cosmides (1989) were not confounded with a deterministic interpretation of the rule and a perspective suggesting a falsification strategy as opposed to a confirmation strategy, the subjects' performance was similar to the performances in descriptive versions of the task, i.e. people predominantly chose the p and q cards. Liberman & Klar (1996) have argued that the crucial factors eliciting logically correct performance are a deterministic interpretation of the rule, a context implying that p-and-not-q cases are relevant violating instances, and adoption of looking for violation strategy. The content of the task critically affects reasoning performance, but this is not domain specific content. As in the case of Oaksford and Chater's (1994) analysis, the relevance account, and more generally the interpretation of the task account, provides a general explanation of all versions of the task.

To summarize, the Wason selection task has been intensively studied over the last thirty years. The reason is that it throws a serious doubt on human rationality. However, these "negative" implications for human inferential capabilities have been based on a particular normative model and ignorance of the conversational factors operating in the experimental situation. Changing the normative model against which the subjects' performance is assessed and taking into account the relevance considerations put the question of human rationality in a more favorable light.

Inductive Reasoning: The Case of the Base-rate Fallacy

The phenomenon of underutilization of prior probabilities, or base-rate fallacy,[9] was introduced by Kahneman & Tversky (1973). Consider one of their original problems: subjects are presented with information about a sample of engineers and lawyers. Specifically, they have information about prior probabilities, seventy persons are lawyers and thirty are engineers (0.7 Vs 0.3 respectively), and individual descriptions of five randomly chosen stimulus persons. For each description, they have to indicate the probability of the described person being an engineer. The subjects in this task consistently underutilize the prior probabilities and base their judgments on the individual information. A normatively correct judgment in a Bayesian probabilistic framework should be influenced by the prior probabilities of the event.[10]

Kahneman & Tversky (1972, 1973; Kahneman, Slovic & Tversky, 1982) have interpreted these results as reflecting the use of a representativeness heuristic.[11]

This is a judgment of the similarity of a target event to its parent population. In the above example, the subjects' judgments are guided by their respective stereotypes of engineers and lawyers and not by normative considerations. The more similar the description to the stereotype held, the higher the probability of the target being a member of the respective stereotyped group. Nisbett & Ross (1980) have taken this lead and have argued that lay people's judgments and explanations are guided by a priori theories, often having nothing to do either with reality or with the cognitive processes which produce these judgments (e.g. Nisbett & Wilson, 1977).

Peterson & Beach (1967) claimed that the Bayesian model provides a good first approximation of human judgment. However, the findings of Kahneman & Tversky have suggested that "man is apparently not a conservative Bayesian: he is not Bayesian at all" to use their phrase (Kahneman & Tversky, 1972, p. 450). However, later studies have demonstrated that the subjects take into account the base rates when they are causally relevant as opposed to incidentally relevant (Ajzen, 1977; Tversky & Kahneman, 1982). For instance, when the base rates are presented in the following way "a final exam was given in a course at Yale University. About 75% of the students failed the exam" and the task is to judge the probability of success of a target person, the prior probabilities affect the subject's judgment in the normatively appropriate way. Still, these findings do not do full justice to the rationality of lay inductive judgments (see for the "covered" effects of sex base rates on gender related judgments Biernat & Manis, 1994; Biernat *et al.*, 1991).

Criterion 1: The Normative Model. L. Cohen (1979) has argued that the Kahneman & Tversky claims are based on the assumption that there is only one legitimate normative model of reasoning about uncertain events. He has opposed the Pascalian notion of probability, which is at the core of the Bayesian model, to a Baconian type of probability. The main difference is that the former grades the probabilities "on the assumption that all relevant facts are specified in the evidence," whereas the latter grade them "by the extent to which all relevant facts are specified in the evidence" (Cohen, 1979, p.389). In view of this Baconian perspective, Cohen has reinterpreted some of the Kahneman & Tversky results. For instance, in the case of the engineer/lawyer problem the idea is that the subjects assume some causal connection between the target person's profession and his interests, abilities and whatever happens to be in his personal description. Given this hypothesis, the individual information is inductively more relevant to the judgment than the base rate information. This interpretation is also consistent with the findings for the effects of causal base rates. In both cases, the judgment is guided by causal hypotheses, but the Baconian concept of probability provides a broader framework. Cohen's conclusion is that people are Baconian in their reasoning and therefore their inductive judgments are rational. Indeed, one does not have to refer to Baconian probability but simply to specification of the

normatively relevant base rate.[12] If the latter is not specified, subjects may assume the principle of indifference and ignore the base rates (see Gigerenzer & Murray, 1987).

Another exploited problem in the studies demonstrating the neglect of base rates is the so called cab problem (Barr-Hillel, 1980; Tversky & Kahneman, 1982). Subjects were presented with the following information: a cab was involved in a hit-and-run accident at night, 85% of the cabs in the city are green and 15% are blue, a witness identified the cab as blue and he makes correct identification 80% of the time. The question is: what is the probability that the cab involved in the accident was blue rather than green? The "correct" Bayesian solution is .43. However, the subjects' modal answer is .80, ignoring the base rates according to the original investigators (Barr-Hillel, 1980; Tversky & Kahneman, 1982). Laying aside the normative relevance of the base rate provided in the study (see below), this problem may have more than one normative solution. Birnbaum (1983), using signal detection theory, which derives from the Neyman & Pearson model of hypothesis testing, demonstrated three such solutions. The main difference is that in the original Bayesian solution it is assumed that the ratio of hit rate, the witness reports blue given the cab is blue, to false alarm, the witness reports blue given the cab is green, is independent of the proportion of each cab color, i.e. the base rate, whereas in signal detection theory this assumption is not valid. As a result, the solution is dependent on an accepted decision criterion. In other words, the final solution depends on a theory of judgments. Interestingly for one of the three theories Birnbaum used, the optimal observer theory where the goal is maximization of the probability of correct identifications, the correct answer is .82.

The use of Bayes rule is only legitimate in conditions of random sampling, i.e. to perform a Bayesian calculation one must be sure that the sample was randomly drawn from the respective population. It is true that in Kahneman & Tversky's engineer/lawyer study, the subjects were told that the sample of the descriptions was randomly drawn, but did they believe that it was really random, given that the descriptions were preselected for the experiment? Gigerenzer, Hell & Blank (1988, Exp. 1) had their subjects observe and participate in a random sampling procedure. The subjects were shown 10 sheets of paper corresponding to the descriptions of either seven engineers and three lawyers or of three engineers and seven lawyers, depending on the base rate condition. The sheets were put in an empty urn and then the subjects drew one of the descriptions. In these conditions, the use of base rates was significantly enhanced. Indeed, the performance was more similar to the optimal Bayesian solution than to the base rate neglect hypothesis.

Base rate information is always relative to a specific population. However, as Einhorn & Hogarth (1981) have pointed out there is no general agreed upon procedure for specifying the normatively appropriate population (see also note 12). In the cab problem, for instance, the subjects were told that 85% of the

cabs in the city are green and 15% are blue. In a causal base rate condition, in which the subjects used the base rate information, the subjects were told that 85% of the incidents in the city involve green cabs and 15% involve blue cabs (Tversky & Kahneman, 1982). The latter base rate information seems more appropriate, but the "correct" answer to the two problems is the same! Even the causal base rate information requires additional assumptions such as that it approximates the base rate specifying the percent of reckless drivers from the green and blue companies leaving the place of an incident (see Gigerenzer & Murray, 1987, pp. 157–162).

Similarly, in medical problems (e.g. Casscells *et al.*, 1978) people are presented with information about the base rate of a disease in a population and the percentage of times a test for the disease is positive and valid. In these problems people hardly use the base rate information. But should the subject use the base rate provided for the general population? The main conceptual problem is how to relate a single probability for a particular person to a frequentist probability. In this case, one must be sure that the appropriate base rate is specified. Is it the base rate of the whole population, of the same age group of the person, or of the same blood group? Indeed, when the information is presented in frequency formats (Gigerenzer & Hoffrage, 1995), i.e. people decide not for a single person but for, say, how many persons from 100 will be affected by the disease when positively tested, the observed performance may be enhanced from 76% to 92% correct Bayesian answers (e.g. Cosmides & Tooby, 1996). Compare this result with 2% "correct" answers in the original problems (Casscells *et al.*, 1978).

Besides, Christensen-Szalanski & Beach (1982) have demonstrated that when people experience the relation between the base rate and the diagnostic information, they use the base rate information in an appropriate fashion (see also Manis *et al.*, 1980). Their subjects observed 100 slides indicating whether a person had a certain disease and his positive or negative results from a test for the disease. Indeed, Gigerenzer *et al.* (1988, Exp. 2) demonstrated that in highly familiar task environments the subjects' performance is almost indistinguishable from a Bayesian performance. They used soccer problems (the experiment was done in Germany) in which the base rate information was the previous proportion of wins and the new information was the half-time result of a match.

Criterion 2: The Conversational Factors. Consider again the original engineer/lawyer task, its instruction begins as follows: "A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written . . ." In a replication of the original study, Schwarz, Strack, Hilton & Naderer (1991, Exp. 1) replaced "on the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written" with information that "a computer drew a random sample of descriptive sentences bearing on the target person." The

probability estimate of the subjects that the target is an engineer dropped significantly from .76 to .40 (prior probability .30). As we noted above, the assumption of intentionality underlies the maxims of conversations. When this assumption is undermined the maxims are no longer operative. In the original study, subjects were led to believe that the individual information was provided by experts. Correspondingly, its relevance was perceived as higher than the relevance of the base rate information. Ginossar & Trope (1987, Exp.6) have demonstrated similar effects when the information was presented to the subjects as generated by a card game (see also Nisbett *et al.*, 1983 for chance framing of the information).

Other changes in the instruction also affect the use of base rate information. Ginossar & Trope (1987, Exp. 5) replaced "a panel of psychologists" who "have interviewed and administered personality tests to . . ." with either "a student who interviewed the target person in one of the first meetings of a course on interviewing" or "a palm reader who examined the target person's palm." In these conditions the subjects used the base rate information in a Bayesian fashion.

Schwarz *et al.* (1991) also varied the presentation format of the task. They manipulated the presentation of the problem in a procedure similar to the one used by Zukier & Pepitone (1984). In the latter study (Exp. 1), half of the subjects were instructed to take a clinical stance and half scientific. These different orientations affected the use of base rates. The "clinical psychologists" relied more on individual information than on the base rates whereas the "scientists" used the base rate information. Schwarz *et al.* (1991) presented the problem as solved either by psychologists or by statisticians, changing the reference to experts in the last part of the instruction:"The same task has been performed by a panel of *experts* who were highly accurate in assigning probabilities to the various descriptions." As predicted, the "statistics" problem enhanced the use of the base rate information.

Another drawback of the original study (Kahneman & Tversky, 1973) is related to the way of presentation of the information. Base rate information was presented as a between-subjects factor, whereas the individual information as a within-subject factor (but see Kahneman & Tversky, 1996 for the use of these experimental designs), i.e. each subject was presented with five different personal descriptions and the same base rate information. This procedure implies greater relevance of the individual information as a judgment cue. Schwarz, Strack, Hilton & Naderer (1991, Exp.2) used the same procedure but varied the base rates and held constant the individual information. Not surprisingly in those conditions, the subjects' judgments were affected by the base rate information. Fischhoff, Slovic & Lichtenstein (1979, Exp. 1) demonstrated the same effect for the cab problem and one additional "factory quality control" problem with the same formal structure (see also Birnbaum & Mellers, 1983).

Ginossar & Trope (1980) demonstrated use of base rate information even in a between-subjects design. Their subjects rated probabilities for only one

personality description. The use of base rate increased as the diagnocity of individuating information decreased (see also Beyth-Marom & Fischhoff, 1983, Exp. 5). Gigerenzer (1991a), reanalyzing the studies on base rate use in the engineer/lawyer problem, has shown that the base rate use depends on the number and order of informative and uninformative descriptions presented to each subject. If the uninformative description is only one or comes first, people use the base rates. If it comes later, especially among several informative descriptions, people neglect the base rates.

Krosnick, Li & Lehman (1990) varied the order of presentation of the base rate and the individual information. Across several experiments, they have demonstrated that when the base rate is presented after the individual information it affects the subjects' judgments. These authors have interpreted their findings in terms of the perceived relevance of the information. The second piece of information is perceived as more relevant and correspondingly is assigned greater weight in the subsequent judgment (cf. the given-new contract, Clark & Haviland, 1977). This effect disappears when subjects are told that the order of information is arbitrary (Krosnick *et al.*, 1990, Exp. 4). Using a somewhat different procedure, Ginossar & Trope (1987, Exp. 2) increased the salience of the base rates and correspondingly its use. They presented the task information in a list format, i.e. every sentence was presented on a separate line. As predicted this presentation enhanced the use of base rates.

All these studies demonstrate that people are quite perceptive to minor changes in the presentation of the information. It seems that the written information is assessed quite precisely for its intended relevance. As Birnbaum points out "it may be that interest in mathematical puzzles is not highly diagnostic of being an engineer or lawyer, but the *report* of interest in mathematical puzzles may be highly diagnostic of a rare engineer in a population consisting mostly of lawyers." (1983, p. 92, italics in the original).

Bar-Hillel (1980) has suggested that the base rate fallacy should be entirely explained in terms of the perceived relevance of the information to the judgment.[13] This is her recipe for constructing a task in which the base rate fallacy is not observed: present the base rate information as referring to a set smaller than the overall population to which the individual item refers (e.g. married couples as opposed to the whole population), but of which the judged case is a member (e.g. husband), i.e. make the information more specific, or make the base rate information causally linked to the judged outcome in the absence of such a link on behalf of the individual information (e.g. a facility is operated by two identical motors but with a different history of breakdown – 85% Vs 15%, there is a breakdown and both motors are tested by a mechanical device which identifies the faulty motor four times out of five, problem 8).

As in the case of the Wason selection task, the base rate fallacy turns out not to be a "real" fallacy on closer inspection. Assessed against alternative probabilistic models, the subjects' performance was perfectly rational. Even in a Bayesian

framework, when the necessary assumptions were preserved, the performance was optimal. Moreover, in tasks related to familiar environments, people's performance was almost identical with the Bayesian predictions. When the conversational factors operating in the experimental situation were controlled, the base rate fallacy largely disappeared.

IMPLICATIONS FOR HUMAN REASONING STUDIES: RELATING THE CRITERIA

Two criteria for evaluation of studies on reasoning have been outlined. First, human performance is always assessed against a certain normative model and therefore any claim about fallacies or biases in human judgment rests on the presumptions of the only validity of the particular model used and its correct application. Hence, any "error-bias" approach has to justify its choice and use of a normative model. Second, in any experimental situation there are tacit assumptions which govern everyday conversation. As a result, many of the alleged biases reflect lack of control of the conversational factors. These criteria were applied to two tasks from the fields of deductive and inductive reasoning. In both cases, the original claims for biases have been demonstrated to be largely unsupported when the original studies were reanalyzed in view of alternative normative models, as well as the conditions of application of the original model of assessment, and relevance considerations. In a weaker sense, the two criteria of evaluation may be considered as constraints on interpretations of experiments allegedly demonstrating biases in human reasoning.

The relevance and normative analyses were converging in the case of the base rate fallacy. When the normatively appropriate base rates are specified, human performance is a good approximation of the Bayesian model. In the case of the Wason selection task, Oaksford & Chater (1995a) have argued that their information gain theory provides a quantitative measure of relevance in the task. They have demonstrated that the Sperber, Cara & Girotto (1995) data fit their model. In both cases, the normative models are compatible with the relevance explanation.

At the same time, the two criteria proposed in the present paper imply somewhat different types of explanations. In its empirical part, the first criterion is related to the application of normative models. These types of "statistical" explanations tend to be grounded in adaptive or evolutionary arguments (e.g. Anderson, 1990; Cosmides, 1989; Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). On the other hand, the second criterion is related to the implicit conversational rules in the experimental situation. How do the relevance and the evolutionary explanations fit together?

Indeed, at some general level it would be difficult to distinguish the two explanations, at least for some representatives of the two trends. Compare the cognitive principle of relevance of relevance theory, also called the first principle

of relevance, and the general principle of rationality of the rational analysis perspective. The former is "Human cognition tends to be geared to the maximization of relevance" (Sperber & Wilson, 1995, p. 260), whereas the latter is "The cognitive system operates at all times to optimize the adaptation of the behaviour of the organism" (Anderson, 1990, p. 28). At first sight, the principles may appear different. However, as was noted above the relevance in relevance theory is defined as a trade-off of cognitive effects and processing efforts. Furthermore, the cognitive principle of relevance, as the general principle of rationality, is grounded in evolutionary considerations (Sperber & Wilson, 1995, pp. 260–263). At this general level, the guiding principles are almost identical. Does it follow that the implied explanations function at different levels of analysis or they are simply different "notational" variants of one underlying principle?

Even if the two principles are identical, the implied explanations at the algorithmic level of human understanding and performance in reasoning tasks are different. For instance, the normative account could fare well without the help of the conversational model and vice versa. In the former, the key notions are that the applicability conditions of the normative model proper should be preserved (e.g. Gigerenzer *et al.*, 1988; Ginossar & Trope, 1987; Kruglanski & Ajzen, 1983; Kruglanski *et al.*, 1984) and that the task information should be presented in frequency formats (e.g. Cosmides & Tooby, 1996; Gigerenzer, 1991a; Gigerenzer & Hoffrage, 1995). In this case, people activate and apply statistics-like rules. In the conversational model, people "assess" the relevance of the information presented in the study and use it according to its degree of relevance. The latter may be influenced by simple wording of the task or of the instruction (e.g. Schwarz *et al.*, 1991), the number of pieces of information and tasks, and many other "extra" logical socially contextual factors (e.g. Hilton, 1995; Schwarz, 1994, 1996).

There are at least two general possibilities how these types of explanations could be related. The first and theoretically uninteresting one is that they are reducible to one underlying explanation whether statistical or pragmatic. The second possibility is that people face two different types of constraints in the experimental situation related to the notions of normative and pragmatic relevance. In that case when the constraints are converging, we should observe good performance in terms of the normative model proper. When they are diverging, we should expect poor performance. This implies that people have intuitive notions of normative and conversational relevance.

One line of reasoning is to propose a two stage model in terms of pragmatic and normative demands of the task. For instance, Evans (1984, 1989) has proposed a two stage model of human reasoning. The first stage consists of heuristics processes which select information for the second stage. The latter consists of content specific analytic processes which generate inferences. The model is plausible, but it still considers the first stage, selection of items of information as a process largely guided by biases. Hilton (1995) has also proposed

a two stage model of judgments in experimental tasks. In the first stage of his model, the subjects interpret the information in a rational way using a criterion of consistency with higher order assumptions about conversation and attributions about the speaker. In the second stage, the subjects apply normative inferential rules to the representations formed. Sperber, Cara & Girotto (1995) also implicitly endorse a two stage model: "Relevance Theory focuses on the psychological processes that guide the selection of information relevant to inferential processes" (p.39). In short, the selection of the information in the first stage is guided by relevance considerations, whereas the computation stage, i.e. the judgment proper, is guided by probability considerations.[14] This is also consistent with a suggestion made by Oaksford & Chater that "pragmatics affects reasoning via its impact on people's subjective probabilities" (1995, p. 102, note 1).

The two stage model notion has a certain heuristic value, but it still requires validation of these two types of constraints. From the present analysis, we can speculate for a Bayesian inductive model of human reasoning. However, the rational analysis does not commit one to a definite view of what is going on in the head (Anderson, 1990). In an earlier article, Oaksford & Chater (1992) even argued that the notion of Bayesian analysis in the head is not plausible because the Bayesian computation would be intractable. The way out of this impasse is to look at the Bayesian analysis as a computational-level theory (cf. Marr, 1982; see also Oaksford & Chater, 1995b), and to try to locate the specific algorithms achieving the same computational goal (e.g. Gigerenzer & Hoffrage, 1995). Indeed, Chater (personal communication, 12/96) points out that a more plausible version of inference processes would describe heuristics approximating Bayesian standards.

However, the implementing algorithms may be described in two different ways related to the statistical and relevance explanations. For instance, in the experiment of Gigerenzer *et al.* (1988, see above), subjects participated in a random sampling procedure and as a result their use of base rate information was enhanced. This experimental manipulation may be interpreted either as a condition which increases the relevance of the base rate or as a condition which activates the use of Bayes like rules. The same is true for the experiment of Schwarz *et al.* (1991, see above), the replacement of experts descriptions with computer random sampling of descriptions may be interpreted either as undermining the intentionality of the information or as activating statistical like rules. In other words, the pragmatic explanation states that what happens is that different experimental manipulations increase the relevance of a specific piece of information and people use it. On the other hand, the normative explanation states that the manipulations activate statistical procedures. Which account is more descriptively correct is an empirical question.

In a similar vein, in deontic versions of the Wason task, the effect of the adopted perspective may either increase the relevance of certain cards (e.g. Liberman & Klar, 1996; Sperber *et al.*, 1995) or activate domain specific knowledge (e.g. Cosmides, 1989; Gigerenzer & Hug, 1992). This consideration

is also valid for base rate studies which manipulate the perspective of the subjects (e.g. Ginossar & Trope, 1987; Zukier & Pepitone, 1984).

If we assume two types of normative and pragmatic constraints present in the experimental situation, their degree of influence on the subjects' judgments would be contingent upon the nature of the problem. Consider again the tasks used for assessment of the use of base rate information. Generally, there are three types of widely used problems: the engineer/lawyer problem, the cab problem, and the medical diagnosis problem. In the former the likelihood ratio is not specified, i.e. the diagnocity of the individuating information (that is why need for two groups with base rates as a between-subjects factor). In such problems it is plausible to assume that subjects will heavily rely on pragmatic cues, such as "a panel of psychologists," "expert evaluation," etc. In the other situations, the mathematical information is provided and the task is to find some solution combining the presented numbers.

This type of task, of course, does not exclude effects of the perceived relevance of the mathematical information on the performance. In one of the studies of Bar-Hillel (1979, Exp. 3), subjects were presented with information about the populations of two cities, i.e. 1,000,000 and 50,000, and the respective sample size of two polls in these two cities, 1000. They were asked which results would be trusted more. The subjects apparently based their answers on the ratio of the sample size to the population, but in situations where the population size is either infinite or much larger than the sample size (sampling with replacement) the normative appropriate solution is to use only the absolute sample size (but see Cohen, 1982). However, the mere inclusion of the population size information makes it relevant unless subjects are explicitly told that not all information is normatively appropriate. In these "relevance" circumstances, it seems a reasonable thing to compute the ratio of the two pieces of information. Other "biases", such as the conjunction fallacy, could be almost entirely explained in terms of pragmatic effects of the framing of the task (e.g. Fiedler, 1988; Dulany & Hilton, 1991).

It seems that a fruitful perspective for future research would be to explore the relations between the statistical type and the relevance type explanations. The error-biases tradition has produced a large body of research and has inspired a lot of new and exciting ideas. As a result, we are at the level of more elaborated, and ironically more rational, processing models (Gigerenzer, 1996). Still, much remains to be done.

Alexander Todorov
Department of Psychology
AL-340 Cognitive Science Lab.
New School for Social Research
65 Fifth Avenue
New York
NY 10003
USA

NOTES

[1] Certain deviations of the judgments from the inferential statistical model were observed, i.e. intuitive inferences were more conservative than normative, but they were not interpreted as biases. Moreover, the results demonstrated that subjects' performance was more optimal in conditions of greater task complexity.

[2] One may argue that even in memory and perception research, as well as in the case of descriptive statistics, the accuracy notion implies a normative model. In memory, for instance, before the invention of writing and literacy (Ong, 1967) the notion of accuracy would have been quite different from the contemporary notion. However, the use of accuracy measures in the mentioned fields is not controversial, i.e. there is a general consensus and no need for supporting assumptions about the normative models used.

[3] Cohen's argument has been harshly criticized (see the accompanying comments of his paper, 1981, and for a recent critical review Stein, 1996, Chapter 5). For instance, Kahneman (1981) has objected that the concern with normative models confuses what is with what should be, or the descriptive and the normative sides of the judgments. However, his heuristics are explicable entirely in terms of Bayesian statistics. Without a normative probabilistic model the heuristics are not deviations, i.e. error prone strategies. On the other hand even the heuristics could be described in terms of the Bayesian model. For instance, the representativeness heuristics is the exclusive reliance on the likelihood ratio or re-description of it (see Gigerenzer & Murray, 1987, pp. 152–156). Further, the use only of the likelihood ratio could be described as a kind of Fisherian analysis (see Gigerenzer & Hoffrage, 1995).

[4] It is ironic that Tversky & Kahneman (1974) used a similar visual analogy and were guided by reasoning about the adaptive value of the heuristics. However, the main emphasis in their and related research in social cognition has been on the systematic errors people make in their judgments implying the irrationality of the latter. There is nothing wrong with research focused on errors in the sense of Funder (1987) as far as they are viewed in the context of descriptive accuracy of the experimenter's model. A deviation from a normative theory may imply not that human reasoning is flawed but that the normative theory is an inappropriate computational-level theory (Oaksford & Chater, 1995b; see also Kyburg, 1983). Besides, a substantial part of the errors could be described as rational (e.g. see Ben-Zeev, 1995, 1996 for mathematical reasoning).

[5] The term 'normative model' refers to models derived from statistical theory in the case of inductive reasoning and from logical theory in the case of deductive reasoning, models which are used as a standard against which lay reasoning is assessed, and not to physical models of reality. For an account of scientific reasoning in terms of use of iconic analogical models based on familiar phenomena and classification of these different types of models see Harré (1970).

[6] There is an apparent contradiction between the adaptive rationality notions and the insistence on the use of normative models in the text. Cherniak (1986), for instance, cites the literature on heuristics as supporting his conception of minimal rational agent which is based on the critique of the plausibility and even applicability of the traditional doctrine of ideal rational agent in philosophy. However, the normative models are

indispensable tools in research on reasoning. The availability of constraints on human reasoning does not mean that the normative models are inappropriate for assessment of human performance, but it means that the constraints should be added as parameters in the models. On the other hand, at the level of the question what counts as rationality, the present argument is consistent with Cherniak's ideas.

[7] The use of the term 'representation' is not intended to imply an affiliation to the traditional AI symbolic approach as opposed to connectionist networks which do not use representations (cf. Rumelhart & McClelland, 1986). The idea simply stated is that different kinds of information may enter into the computation depending on the interpretation of the task. To put it another way, the form of the premises may be changed. This does not mean that people necessarily form an explicit representation of the premises.

[8] It should be noted that Kahneman & Tversky (1982) have explicitly recognized the possible role of conversational factors in the empirical studies on judgment: "We conclude that the conversational aspect of judgment studies deserves more careful consideration than it has received in past research, our own included" (p. 135, reprinted also in Kahnerman, Slovic & Tversky, 1982, p. 504).

[9] In a recent article, which I read after having completed the present paper, Koehler (1996) makes an excellent survey of the research on the so called base rate fallacy and draws conclusions similar to the conclusions drawn in the present paper. His analysis is entirely consistent with the analysis under the heading of criterion 1 in the present article.

[10] Basically, in studies on base rate use the subjects' answers are compared with the solution derived from Bayes' theorem. The latter is a rule for combining old and new information or for revision of one's beliefs in view of new evidence. Stated in terms of hypothesis testing, the theorem specifies how a hypothesis should be revised in the light of new data. The formula is the following: $P(H/D) = P(H) \times P(D/H)/P(D)$, where $P(H/D)$ is the probability of the hypothesis given the data, $P(H)$ is the prior probability of the hypothesis, $P(D/H)$ is the probability of the data given the hypothesis, and $P(D)$ is the probability of the data.

[11] Wells & Harvey (1978) have pointed out that Kahneman & Tversky (1973) found significant effects for base rate information and that the question should be to what extent people use base rates. Besides, they have revealed certain inadequacies in the data analysis of Kahneman & Tversky. For instance, Kahneman & Tversky used median and means comparison between the two base rates conditions over all 5 target cases. Given the individual differences between the subjects and the differences between the target cases, this analysis is not completely representative for the base rate use. Furthermore, averaging is a linear mathematical transformation whereas the Bayesian model predicts a curvilinear relation between the two base rate conditions. Wells & Harvey did a replication of the original study for three of the target cases and used a more fine-grained method of data analysis. Bayes' rule allows for a prediction what a person who estimated a target case should predict in the other base rate condition for the same target. In their study, they computed these scores and compared their means for each target case with the observed means. Wells & Harvey demonstrated significant effects for base rate information.

[12] Recent critics of Cohen (see Koehler, 1996 and the accompanying comments by Kyburg; Levi; McKenzie & Soll) have pointed out that the problem of specification of the normatively relevant base rate is far from resolved. Indeed, the use of more and more specific base rates may lead to less statistical reliability because of the small sample size. Meehl's (1954) proposal of dealing with this problem is to use the smallest reference class which can generate stable frequencies. Koehler (1996) suggests that the use of different base rates is an empirical matter.

[13] Despite this explanation, Bar-Hillel still treats the base rate effect as fallacy. Barr-Hillel, like Evans (1984), views the perceived relevance of information as a kind of heuristic, and hence, as a hindrance for appropriate use of normatively pertinent information.

[14] It should be noted that according to Sperber & Wilson (1995) inferences are performed by a deductive device, but see Oaksford & Chater (1991) for the computational intractability of this version.

REFERENCES

ADLER, J.E. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behaviour*, **14**, 165–181.

AJZEN, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology*, **35**, 303–314.

ANDERSON, J.R. (1990). *The Adaptive Character of Thought*. Hillsdale, NJ: Erlbaum.

BAR-HILLEL, M. (1979). The role of sample size in sample evaluation. *Organizational Behavior and Human Performance*, **24**, 245–257.

BAR-HILLEL, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, **44**, 211–233.

BEN-ZEEV, T. (1995). The nature and origin of *rational errors* in arithmetic thinking: Induction from examples and prior knowledge. *Cognitive Science*, **19**, 341–376.

BEN-ZEEV, T. (1996). When erroneous mathematical thinking is just as "correct". The oxymoron of rational errors. In R.J. Sternberg & T. Ben-Zeev (Eds.), *The Nature of Mathematical Thinking*. Mahwah, NJ: Lawrence Erlbaum Associates.

BEYTH-MAROM, R. & FISCHHOFF, B. (1983). Diagnocity and Pseudodiagnocity. *Journal of Personality and Social Psychology*, **45**, 1185–1195.

BIERNAT, M., MANIS, M. & NELSON, T. (1991). Stereotypes and standards of judgment. *Journal of Personality and Social Psychology*, **60**, 485–499.

BIERNAT, M. & MANIS, M. (1994) Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, **66**, 5–20.

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

BIRNBAUM, M.H. & MELLERS, B.A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, **45**, 792–804.

CASSCELLS, W., SCHOENBERGER, A. & GRABOYS, T.B. (1978). Interpretations by physicians of clinical laboratory results. *New England Journal of Medicine*, **299**, 999–1001.

CHENG, P.W. & HOLYOAK, K.J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, **17**, 391–416.

CHERNIAK, C. (1986). *Minimal Rationality*. Cambridge, MA: MIT Press.

CHRISTENSEN-SZALANSKI, J.L. & BEACH, L.R. (1982). Experience and the base rate fallacy. *Organizational Behavior and Human Performance*, **29**, 270–278.

CLARK, H.H. & HAVILAND, S.E. (1977). Comprehension and the given-new contract. In R.O. Freedle(Ed.), *Discourse Production and Comprehension*. Norwood, NJ: Ablex.

CLARK, H.H. & SCHOBER, M.F. (1991). Asking questions and influencing answers. In J.M. Tanur (Ed.), *Questions about Questions: Inquiries into the Cognitive Bases of Surveys*. NY: Russel Sage Foundation.

COHEN, L.J. (1979). On the psychology of prediction: Whose is the fallacy? *Cognition*, **7**, 385–407.

COHEN, L.J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, **4**, 317–370.

COHEN, L.J. (1982). Are people programmed to commit fallacies? Further thought about the interpretation of experimental data on probability judgment. *Journal for the Theory of Social Behaviour*, **12**, 251–274.

COSMIDES, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, **31**, 187–276.

COSMIDES, L. & TOOBY, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, **58**, 1–73.

DULANY, D.E. & HILTON, D.J. (1991). Conversational implicatures, conscious representation, and the conjunction fallacy. *Social Cognition*, **9**, 85–110.

EINHORN, H.J. & HOGARTH, R.M. (1981). Behavioral decision theory: Processes of judgment and choice. *Annual Review of Psychology*, **32**, pp. 53–88.

EVANS, J. St. B.T. (1972). Interpretation and "matching bias" in a reasoning task. *Quarterly Journal of Experimental Psychology*, **24**, 193–199.

EVANS, J. St. B.T. (1984). Heuristic and analytical process in reasoning. *British Journal of Psychology*, **75**, 451–468.

EVANS, J. St. B.T. (1989). *Bias in human reasoning: Causes and consequences*. Hillsdale, NJ. Erlbaum.

EVANS, J. St. B.T. & LYNCH, J.S. (1973). Matching bias in the selection task. *British Journal of Psychology*, **64**, 391–397.

EVANS, J. St. B.T. NEWSTEAD, S.E. & BYRNE, R.M.J. (1993). *Human Reasoning: The psychology of deduction*. Hillsdale, NJ: Erlbaum.

EVANS, J. St. B.T. & OVER, D.E. (1996). Rationality in the selection task: Epistemic utility versus uncertainty reduction. *Psychological Review*, **103**, 356–363.

FIEDLER, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, **50**, 123–129.

FISCHHOFF, B. & BEYTH-MAROM, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, **90**, 239–260.

FISCHHOFF, B., SLOVIC, P. & LICHTENSTEIN, S. (1979) Subjective sensitivity analysis. *Organizational Behavior and Human Performance*, **23**, 339–359.

FISKE, S. (1992). Thinking is for doing: Portraits of social cognition from daguerreotype to laserphoto. *Journal of Personality and Social Psychology*, **63**, 877–899.

FISKE, S. (1993). Social cognition and social perception. *Annual Review of Psychology*, **44**, pp. 155–194.

FUNDER, D.C. (1987). Errors and mistakes: Evaluating the accuracy of social judgment. *Psychological Bulletin*, **101**, 75–90.

FUNDER, D.C. (1993). Judgments as data for personality and developmental psychology: Error versus accuracy. In D. Funder, R. Parke, C. Tomlinson-Keasay & R. Widaman (Eds.), *Studying lives through time: Personality and development*. Washington, DC: American Psychological Association.

FUNDER, D.C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, **102**, 652–670.

GIGERENZER, G. (1991a). How to make cognitive illusions disappear. Beyond "heuristics and "biases." In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology*, Vol. 2. Chichester, England: Wiley.

GIGERENZER, G. (1991b). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, **98**, 254–267.

GIGERENZER, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, **103**, 592–596.

GIGERENZER, G., HELL, W. & BLAND, H. (1988). Presentation and content: The use of base rates as a continuous variable. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 513–525.

GIGERENZER, G. & HOFFRAGE, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, **102**, 684–704.

GIGERENZER, G. & HOFFRAGE, U. & KLEINBOLTING, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, **98**, 506–528.

GIGERENZER, G. & HUGG, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, **43**, 127–171.

GIGERENZER, G. & MURRAY, D.J. (1987). *Cognition as Intuitive Statistics*. Hillsdale, NJ: Erlbaum.

GINOSSAR, Z. & TROPE, Y. (1980) The effects of base rates and individuating information on judgments about another person. *Journal of Experimental Social Psychology*, **16**, 228, 242.

GINOSSAR, Z. & TROPE, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, **52**, 464–474.

GRICE, H.P. (1975). Logic and conversation. The William James lectures. Harvard University, 1967–68. In: P. Cole & J. Morgan (Eds.), *Syntax and Semantics* (Vol. 3, *Speech acts*). New York: Academic Press.

GRIGGS, R.A. & COX, R.J. (1982). The elusive thematic material effect in Wason's selection task. *British Journal of Psychology*, **73**, 407–420.

HARRÉ, R. (1970). *The principles of scientific thinking*. Chicago: The University of Chicago Press.

HEMPEL, C.G. (1965). *Aspects of scientific explanation*. New York: Free Press.

HENLE, M. (1962). On the relation between logic and thinking. *Psychological Review*, **69**, 366–378.

HILTON, D.J. (1990). Conversational processes and causal explanation. *Psychological Bulletin*, **107**, 65–81.

HILTON, D.J. (1995). The social context of reasoning. Conversational inference and rational judgment. *Psychological Bulletin*, **118**, 248–271.

HILTON, D.J. & SLUGOSKI, B.R. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, **93**, 75–88.

HORWICH, P. (1982). *Probability and Evidence*. Cambridge, England: Cambridge University Press.

JOHNSON-LAIRD, P.N. & BYRNE, R.M.J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

JOHNSON-LAIRD, P.N. & BYRNE, R.M.J. (1992). Modal reasoning, models, and Manktelow and Over. *Cognition*, **43**, 173–182.

JOHNSON-LAIRD, P.N., LEGRENZI, P. & LEGRENZI, M.S. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, **63**, 395–400.

JOHNSON-LAIRD, P.N. & WASON, P.C. (1970). A theoretical analysis of insight into a reasoning task. *Cognitive Psychology*, **1**, 134–148.

JUSSIM, L. (1991). Social perception and social reality: A reflection-construction model. *Psychological Review*, **98**, 54–73.

KAHNEMAN, D. (1981). Who shall be arbiter of our intuitions? *Behavioural and Brain Sciences*, **4**, 339–340.

KAHNEMAN, D., SLOVIK, P. & TVERSKY, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

KAHNEMAN, D. & TVERSKY, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, **3**, 430–454.

KAHNEMAN, D. & TVERSKY, A. (1973). On the psychology of prediction. *Psychological Review*, **80**, 237–251.

KAHNEMAN, D. & TVERSKY, A. (1982). On the study of statistical intuitions. *Cognition*, **11**, 123–141.

KAHNEMAN, D. & TVERSKY, A. (1996). On the reality of cognitive illusions. *Psychological Review*, **103**, 582–591.

KELLEY, H.H. (1967). Attribution theory in social psychology. In: D. Levine (Ed.), *Nebraska Symposium on Motivation*, Vol. 15, pp. 192–230.

KIRBY, K.N. (1994). Probabilities and utilities of fictional outcomes in Wason's four card selection task. *Cognition*, **51**, 1–28.

KLAYMAN, J. & HA, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, **94**, 211–228.

KOEHLER, J.J. (1996). The base rate fallacy reconsidered. Descriptive, normative, and methodological challenges. *Behavioural and Brain Sciences*, **19**, 1–53.

KROSNICK, J.A., LI, F. & LEHMAN, D.R. (1990). Conversational conventions, order of information acquisition, and the effect of base rates and individuating information on social judgments. *Journal of Personality and Social Psychology*, **59**, 1140–1152.

KRUGLANSKI, A. & AJZEN, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology*, **13**, 1–44.

KRUGLANSKI, A., FRIEDLAND, N. & FARKASH, E. (1984). Lay persons' sensitivity to statistical information: The case of high perceived applicability. *Journal of Personality and Social Psychology*, **46**, 503–518.

KUHN, T.S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

KYBURG, H.E. (1983). Rational belief. *Behavioral and Brain Sciences*, **6**, 231–273.

KYBURG, H.E. (1996). Probabilistic fallacies. *Behavioral and Brain Sciences*, **19**, 31.

LACATOS, I. (1978). *Methodology of Scientific Research Programmes*. Cambridge: Cambridge University Press.

LEVI, I. (1996). Fallacy and controversy about base rates. *Behavioural and Brain Sciences*, **19**, 31–32.

LIBERMAN, N. & KLAR, Y. (1996). Hypothesis testing in Wason's selection task: Social exchange cheating detection or task understanding. *Cognition*, **58**, 127–156.

LOPES, L.L. (1991). The rhetoric of irrationality. *Theory and Psychology*, **1**, 65–82.

MANIS, M., DOVALINA, I., AVIS, N.E. & CARDOZE, S. (1980). Base rates can affect individual predictions. *Journal of Personality and Social Psychology*, **38**, 231–248.

MANKTELOW, K.I. & OVER, D.E. (1991). Social roles and utilities in reasoning with deontic conditionals. *Cognition*, **39**, 85–105.

MARR, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: Freeman.

MCGARRIGLE, J. & DONALDSON, M. (1974). Conversation accidents. *Cognition*, **3**, 341–350.

MCKENZIE, C.R.M. & SOLL, J.B. (1996). Which referene class is evoked? *Behavioural and Brain Sciences*, **19**, 34–35.

MEEHL, P.E. (1954). *Clinical versus Statistical Prediction: A Theoretical Analysis and a Review of the Literature*. Minneapolis: University of Minnesota Press.

NICKERSON, R.S. (1996). Hempel's paradox and Wason's selection task: Logical and psychological puzzles of confirmation. *Thinking and Reasoning*, **2**, 1–31.

NISBETT, R. & BORGIDA, E. (1975). Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, **32**, 932–943.

NISBETT, R., KRANTZ, D.H., JEPSON, C. & KUNDA, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, **90**, 339–363.

NISBETT, R. & ROSS, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

NISBETT, R. & WILSON, T. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, **84**, 231–259.

NISBETT, R., ZUKIER, H. & LEMLEY, R.E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, **13**, 248–277.

OAKSFORD, M. & CHATER, N. (1991). Against logicist cognitive science. *Mind and Language*, **6**, 1–38.

OAKSFORD, M. & CHATER, N. (1992). Bounded rationality in taking risks and drawing inferences. *Theory and Psychology*, **2**, 225–230.

OAKSFORD, M. & CHATER, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, **101**, 608–631.

OAKSFORD, M. & CHATER, N. (1995a). Information gains explains relevance which explains selection task. *Cognition*, **57**, 97–108.

OAKSFORD, M. & CHATER, N. (1995b). Theories of reasoning and the computational explanation of everyday inference. *Thinking and Reasoning*, **1**, 121–152.

OAKSFORD, M. & CHATER, N. (1996). Rational explanation of the selection task. *Psychological Review*, **103**, 381–391.

ONG, W.J. (1967). *The Presence of the Word.* NY: Simon and Schuster.

ORNE, M. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, **17**, 776–783.

PETERSON, C.R. & BEACH, L.R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, **68**, 29–46.

PIAGET, J. (1952). *The child's conception of number.* London, England: Routledge & Kegan Paul.

PIAGET, J. & INHELDER, B. (1969). *The psychology of the child.* New York: Basic Books.

POPPER, K.R. (1959). *The logic of scientific discovery.* London: Hutchinson. (Originally published 1935.).

POPPER, K. (1979). *Objective Knowledge. An Evolutionary Approach.* Oxford Clarendon Press.

ROSSER, J.B. & TURQUETTE, A.R. (1974). *Many Valued Logics.* Amsterdam: North-Holland Publishing Co.

RUMELHART, D.E. & MCCLELLAND, J.L. (Eds.). (1986). *Parallel Distributed Processing* (2 vols.). Cambridge, MA: MIT Press.

SCHOBER, M.F. & CONRAD, F.G. (under review). Reducing survey measurement error through conversational interaction.

SCHWARZ, N. (1994). Judgment in a social context: Biases, shortcomings, and the logic of conversation. In: M.P. Zanna (Ed.), *Advances in Experimental Social Psychology.* New York: Academic Press.

SCHWARZ, N. (1996). *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation.* Mahwah, NJ: Lawrence Erlbaum Associates.

SCHWARZ, N., STRACK, F., HILTON, D.J. & NADERER, G. (1991). Judgmental biases and the logic of conversation: the contextual relevance of irrelevant information. *Social Cognition*, **9**, 67–84.

SCHWARZ, N., STRACK, F. & MAI, H.P. (1991). Assimilation and contrast effects in part-whole question sequences: A conversational logic analysis. *Public Opinion Quarterly*, **53**, 3–23.

SIMON, H.A. (1957). *Models of Man.* NY: John Wiley and Sons.

SPERBER, D., CARA, F. & GIROTTO, V. (1995). Relevance theory explains the selection task. *Cognition*, **57**, 31–95.

SPERBER, D. & WILSON, D. (1995). *Relevance: Communication and Cognition.* (2nd Ed.) Oxford: Blackwell.

STEIN, E. (1996). *Without Good Reason: The Rationality Debate in Philosophy and Cognitive Science.* Oxford Clarendon Press.

STRACK, F., MARTIN, L.L. & SCHWARZ, N. (1988). Priming and communication: Social determinants of information use in judgments of life satisfaction. *European Journal of Social Psychology*, **18**, 429–442.

STRACK, F. & SCHWARZ, N. (1992). Communicative influences in standardized question situations: The case of implicit collaboration. In G.R. Semin & K. Fiedler (Eds.), *Language, Interaction and Social Cognition*. London: SAGE Publications.

STRACK, F., SCHWARZ, N. & WÄNKE, M. (1991). Semantic and pragmatic aspects of context effects in social and psychological research. *Social Cognition*, **9**, 111–125.

TETLOCK, P.E., LERNER, J.S. & BOETTGER, R. (1996). The dilution effect: judgmental bias, conversational convention, or a bit of both? *European Journal of Social Psychology*, **26**, 915–934.

TVERSKY, A. & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, **185**, 1124–1131.

TVERSKY, A. & KAHNEMAN, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

WASON, P.C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, **20**, 273–281.

WASON, P.C. & JOHNSON-LAIRD, P.N. (1972). *The psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

WELLS, G.L. & HARVEY, J.H. (1978). Naive attributers' attributions and predictions: What is informative and when is an effect an effect? *Journal of Personality and Social Psychology*, **36**, 483–490.

ZUKIER, H. (1982). The dilution effect: The role of the correlation and the dispersion of predictor variables in the use of nondiagnostic information. *Journal of Personality and Social Psychology*, **43**, 1163–1174.

ZUKIER, H. & PEPITONE, A. (1984). Social roles and strategies in prediction: Some determinants of the use of base-rate information. *Journal of Personality and Social Psychology*, **47**, 349–360.