The representation of the task: The case of the Lawyer-Engineer problem in probability judgment

Guy Politzer, Laura Macchi

In V. Girotto & P. N. Johnson-Laird (Eds.) ***The shape of reason. Essays in honour of Paolo Legrenzi*** **(pp. 119-135). Hove: Psychology Press.**

The representation of the task: The case of the Lawyer-Engineer problem in probability judgment

Guy Politzer (*)
CNRS, Saint-Denis, France
Laura Macchi (**)
Universitá Bicocca, Milan, Italy

(*) CNRS-Psychologie Cognitive,
Université de Paris VIII,
2 rue de la Liberté,
93526 Saint-Denis, France
e-mail: politzer@univ-paris8.fr

(**) Dipartimento di Psicologia,
Universitá degli Studi di Milano Bicocca

Piazza dell'Ateneo Nuovo, 1
20126 Milano
e-mail: laura.macchi@unimib.it

Numerous studies have investigated the use of base rates in probabilistic reasoning. Early results of these studies suggested that people generally ignore or neglect base rate probabilities. A famous problem initially proposed by Kahneman and Tversky (1973) was the following:

> 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. You will find on your forms a description, chosen at random from the 100 available descriptions. For each description, please indicate your probability that the person described is an engineer, on a scale from 0 to 100.

This information was followed by a personality description that presented the stereotype of an engineer:

> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.
>
> The probability that Jack is one of the 30 engineers in the sample of 100 is ____%.

And this description was followed in turn by four other descriptions that varied in their degree of representativeness (one of which, the *null* description, was completely nondiagnostic between the Engineer and the Lawyer stereotypes). These texts were submitted to one group of participants (the *low* base rate group) whereas a second group of subjects (the *high* base rate group) received identical texts except that the proportion of engineers and lawyers, which defines the base rate, differed (70 engineers vs. 30 lawyers). The aim of the study was to check the effect of the two different base rates on the probability judgment by comparing the evaluations given by the two experimental groups. The difference should indicate how wide the use of the base rate was. The prediction followed the representativeness hypothesis which posits that people "select or order outcomes by the degree to which the outcomes represent the essential features of the evidence" (Kahneman & Tversky, p. 237-238).

A consequence of this hypothesis for the description above is that the difference between the evaluations made by the two groups should be very small, because the participants should mainly consider the description that was identical for the two groups.

The answer considered normatively correct follows from Bayes's theorem. Its statement in terms of odds shows that the ratio of the posterior odds for the two groups depends on the base rate but is independent of the likelihood ratio, so that it is the same for all descriptions. Pooling across descriptions (except the null), the average of the mean probability estimate for each participant was only slightly higher for the *high* base rate group (55%) than for the *low* base rate group (50%). The authors concluded that, as the representativeness hypothesis predicted, the participants largely ignored base rates.

The concept of representativeness has been the target of much criticism (Olson, 1976; Evans & Pollard, 1982; Gigerenzer, 1991; Gigerenzer, Hell & Blank, 1988). It has been shown that the neglect of the base rate lacks robustness (for a review see Koehler, 1996). This question has been hotly debated (e.g. , Kahneman & Tversky, 1996; Gigerenzer, 1996) and the standard experimental paradigm might be clarified by a different approach. The present paper is devoted to such an approach.

The pragmatic approach to the psychology of thinking and reasoning (Politzer, 1986; Hilton, 1995; Politzer & Macchi, 2000; Politzer, in press a) has proved fruitful in various areas. It sheds new light on classic tasks (see Macchi, 1995, for Kahneman & Tversky's cab problem; Dulany & Hilton, 1991, Mosconi & Macchi, 2001, Politzer & Noveck, 1991, for the conjunction fallacy; Politzer, in press b, for conditional reasoning in context; Van der Henst, Rossi, & Schroyens, 2002, for Wason's 2-4-6 task); it raises and solves new questions about old paradigms (see Van der Henst, Sperber & Politzer, 2002, for relational reasoning; Mosconi, 1990, for problem solving at large); it even changes our understanding of some tasks altogether (see Politzer, 1993, for the class inclusion question in children; Sperber, Cara, & Girotto, 1995, and Girotto, Kemmelmeier, Sperber & Van der Henst, 2001, for Wason's Selection Task). By the "pragmatic approach", we do not refer to the mere acknowledgement of, or concern about, the effects of context or of world knowledge: It is now widely agreed

that such factors affect performance. Our view is more radical: We believe that experimental tasks for which a normatively correct response can be defined should be submitted to a double examination. One, carried out at a micro-structure level, consists of a linguistic analysis of the premises or of the problem statement in order to make sure that they convey the meaning intended by the experimenter: A typical outcome of such an analysis is the identification of different possible interpretations due to the generation of conversational implicatures (Grice, 1989), either particularized (that is, implicatures generated in a specific context), or generalized (that is, implicatures that may accompany connectives or quantifiers). The other examination, at a macro-structure level, consists of identifying the representation of the task that participants are likely to build: A typical outcome of this examination is the identification of the kind of skill, knowledge, or ability that participants think they must exhibit in order to satisfy the experimenter's request. This latter analysis takes a serious view on the special relationship between experimenter and participant (akin to the one between teacher and student in testing situations); it is common knowledge that what is of interest to the experimenter is not the informational content of the answer, but its normative correctness. But what is of interest to the experimenter is not always clear to the participant; the participant may engage in a process of attribution whose outcome may be different from the experimenter's expectations. If this occurs without being detected by the experimenter, the participant's interpretation of the task will cause the experimenter to misinterpret the results. In this paper, we argue that such a misinterpretation is the case for the Lawyer-Engineer problem. In the 'fifties and 'sixties, some social psychologists were concerned about the demand characteristics of the task (Orne, 1969) but unfortunately, with few exceptions, investigators of thinking and reasoning have not paid enough attention to such worries. One exception is due to some educational psychologists who have introduced the notion of a *didactic contract,* that is, an implicit set of rules based on mutual knowledge between teacher and pupils that regulate the interpretation of the question and the type of responses that are appropriate (see Schubauer-Leoni & Perret-Clermont 1997).

Another notable exception is the work of Schwarz (1996) which highlights the importance of the conversational approach and its implications for the design of experiments. Norenzayam & Schwarz (1999) asked participants to attribute reasons for a crime. One group was offered cues to identify the experimenter as a social scientist, and the other group as a personality psychologist. The first group gave more situational attributions than dispositional attributions, whereas the second group showed the reverse pattern. This result demonstrates the role of the epistemic goals that the participants attribute to the experimenter.

Schwarz, Strack, Hilton & Naderer (1991) applied the same approach to the Lawyer-Engineer problem. They presented the psychological description either as a profile written by a psychologist or as a piece of information formulated by a computer. The probability estimate that the description was that of an engineer was higher in the former case than in the latter, indicating that participants relied more on the individuating information in the former case. According to the authors these results show that the participants' use of information depends on the communicative intention of the experimenter. While we agree with these authors' conversational approach, their dependent variable - as in most investigations of the problem - (the mean probability estimate) is not fully satisfactory. It yields only indirect conclusions such as, "the higher probability evaluation indicates greater use of individuating information" without allowing the measurement of the size of the effect. Hence, if one aims to show that pragmatic factors are the main determinants of performance in this task, it is necessary to use a different method based on a different dependent variable.

In this paper, we adopt the theoretical framework of Sperber & Wilson's (1995) relevance theory. The Lawyer-Engineer problem has an evident heterogeneous structure in that the base rate is in the form of a percentage while the individuating information is in the form of a description. Participants assume that the information provided to them is relevant. Indeed, the processing of the description has cognitive effects at a relatively low cost in terms of effort: It yields an estimate of typicality that is necessary for solving the problem. This estimate shows the ability of the participants to exploit a stereotype. Such an estimate, the participants will suppose, is of interest to the experimenter, and so, by

providing a numerical value that reflects their estimate of typicality, they will be satisfied that they have fulfilled their task. Our hypothesis, in brief, is that the participants take their task to be one in which they should infer a typicality estimate based solely on the description.

Following our hypothesis, it should be possible to alter performance by designing a situation in which participants would receive equivalent individuating information but without inferring it themselves. For example, if they were provided with just an evaluation of typicality without a psychological description, participants should correctly view the task as mathematical, which would lead them to use the numerical information available in combination with the indicated typicality. In the usual condition, however, many participants stop their reflection after they have worked out a typicality value of their own because, as claimed earlier, they think that, in exploiting the description, they have fulfilled their task.

Experiment

Method.

Materials and design. Each participant received a form that contained the problem statement followed by the usual request to indicate the probability that the file under consideration was that of one of the engineers.

Four kinds of forms were prepared, defining four conditions (three controls and one experimental condition). For all conditions, the base rate was defined by the same ratio: 30% of engineers and 70% of lawyers. The four conditions were the following (see the Appendix).

- the *Diagnostic description* condition (a) presented a translation and adaptation of Kahneman & Tversky's original problem.
- the *Nondiagnostic description* condition (b), is a control presenting a psychological description that could fit various professions and was nondiagnostic between the engineer and the lawyer stereotypes.

There were two novel conditions:

- the *No description + Statement of Typicality* condition (c) repeated the story about the panel of psychologists making evaluations and about choosing files randomly, but did not present any psychological description. In its

place, it contained this statement: "The description is typical of an engineer".
- the *Diagnostic description + Statement of Typicality* condition (d) was a combination of the standard condition (a) and of the crucial statement used in condition (c); in other words, it presented both the statement, "the description is typical of an engineer" and the psychological description. It was used as a control for the comparison between conditions (a) and (c). (The authors are indebted to Phil Johnson-Laird for suggesting this condition).

Predictions. The probability estimates should not only reflect the relative extent to which participants rely on the individuating information. As noted above, they should also indicate an absolute measure of this extent. An objective and accurate measure can be obtained in the case where the individuating information is totally disregarded in favor of the base rate, in which case the participants give the proportion of engineers (30%) as their answer. Conversely, this answer indicates that the participant is sensitive to the base rate exclusively. We chose a value commonly used in other studies. Likewise, the choice of 30% was motivated by the fact that an answer equal to .30 can unambiguously be attributed to the exclusive use of the base rate because a high probability estimate derived from the typical description is unlikely to yield the .30 value. This would not be the case with the 70% value (or any value higher than 50%).

Assuming that the only two sources of information for participants to work out their response are the base rate and the individuating information, when a response differs from .30 it can be inferred that the individuating information has been taken into account at least to some extent. In brief, the dependent variable, which will be called r, is the frequency of exact reproduction of the base rate. Readers might object that an estimate of 50% does not indicate a non-exclusive use of base rate but rather a "don't know" response. But the results of a pilot study in which participants were asked to justify their answers indicated that an estimate of 50% was a numerical way to express uncertainty based on a lack of cues.

The following predictions were made:

(i) The value of r should be lower with (a) *(Diagnostic description)* than with (b) *(Nondiagnostic description*), that is, there should be a greater use of the description in (a) than in (b). This prediction follows from the fact that the description is more relevant in (a) than in (b), i.e., the diagnostic description suggests the stereotype of one of the two professions. The use of this description yields an estimate of typicality and justifies the presumption of its relevance. In contrast, the nondiagnostic description is not the source of an estimate of typicality, which could be used as a basis for answering the question: It is less relevant in the technical sense, that is, few, if any, inferences can be made from it.

(ii) The value of r should not differ between (a) *(Diagnostic description)* and (d) *(Diagnostic description + Statement of Typicality.* In both conditions, the description is relevant because it enables one to make an inference: Producing an inference in condition (a) and verifying an inference in condition (d). The difference is that, in the latter condition, the description is information attesting to the reasonableness of the judgment of typicality. The inference is directed by given information. But, in the former condition, no such direction is provided: The description is information from which a judgment of typicality has to be inferred. Hence, in both cases, the description is roughly equally relevant because the participant makes use of it either to produce an inference or to make a verification. The values of r should accordingly be comparable in both conditions.

(iii) The main prediction is that the conditions that present a diagnostic description ((a) and (d)) will elicit a greater use of the individuating information (and therefore a less exclusive use of the base rate) than condition (c) *(No description + Statement of Typicality).* This is because in the conditions with a description, participants may suppose that the essence of their task is to infer a typicality value. If so, they will tend to focus on this value, which represents the end of the task (and consequently overlook the base rate). In contrast, the experimental condition contains no individuating information for participants to work out a typicality estimate. They cannot view the task as calling for such an estimate. They should represent the task as closer to a mathematical exercise than to a request for displaying their psychological skills, and consequently they will tend to look for numerical information. Also, this

condition is crucial to distinguish the relevance explanation from the typicality explanation. From the relevance point of view, a typicality value worked out from a description is not equivalent to one that is passively read. Only in the former case does the high degree of typicality render the description relevant for the task, which should be represented as an assessment of the typicality of the description. In the experimental condition, the task has a different representation, as the typicality value is given: The typicality description is relevant for estimating the probability that an engineer is in the sample. In contrast, for the typicality approach, the two conditions (with or without an explicit statement) are just two alternative ways of giving the same final information to the effect that the typicality value is high.

Participants and procedure. Participants were 113 undergraduate students of Psychology at the University Bicocca in Milan. Each of them was randomly allocated to one of the four conditions.

Results

Table 1 indicates the frequency of responses equal to the base rate for each condition. The *No description + Statement* condition (c) yielded more productions of the base rate, that is, less use of the individuating information, than the *Diagnostic description* condition (a) (chi-square = 29.9, $p < .001$), and also than the *Diagnostic description + Statement* condition (d), (chi-square = 16.2 , $p < .001$). In addition, the results corroborated the prediction that the *Diagnostic description* condition (a) would give rise to more use of individuating information than the *Nondiagnostic description* condition (b) (chi-square = 8.58, $p < .01$). Likewise, as we predicted, performance with the *Diagnostic description* (a) and with the *Diagnostic description + Statement* (d) did not differ significantly.

INSERT TABLE 1 ABOUT HERE

In brief, the main prediction was confirmed: Participants who received the individuating information indirectly in the form of a psychological description (the standard *Diagnostic description* condition (a)) from which they had to infer a judgment of typicality used this information more often than those who received it in the form of an explicit statement of typicality (in the *No description + Statement* condition (c)). This result is paradoxical: One would expect the information that is less explicit and harder to exploit to be used less frequently, so that after

the description, participants would use it less often (because the typicality must be inferred) than those to whom this information is explicitly offered in a direct statement. The contrary occurred, and our theoretical approach explains why: When they are presented with a psychological description, participants' representation of the task is that of a test of their own ability to identify the psychological features of a description that can distinguish an engineer from a lawyer. They are consequently diverted from making use of the other sources of information such as the base rate which belongs to a task which is a mathematical exercise: The representation of the two kinds of task are alien to each other.

Readers might object that the less frequent use of individuating information in the *No description + Statement* condition as compared to the *Diagnostic description* condition could be attributed to the presence of the typicality statement rather than to the absence of a description. But, on the one hand, it is hard to see how a statement of typicality could decrease the use of individuating information: If it had any effect in isolation, it could only increase it. And, on the other hand, the absence of a significant difference between the *Diagnostic description* and the *Diagnostic description + Statement* conditions indicates that the presence of the typicality statement had no effect of its own.

One result is at variance with the results originally reported by Kahneman &Tversky (1973) and replicated in other studies (e. g. , Gigerenzer, & al., 1988) concerning the nondiagnostic descriptions. These studies found that exactly the same median estimate was given by the groups that had different base rates (30-70 and 70-30) and they took this observation as evidence that the base rates are totally disregarded. But this is a very coarse indicator that conceals the distribution of the participants' answers. The value of the median is compatible with the earlier interpretation, but there is another interpretation that is at least as likely: The same median value was observed because participants have used the base rate in the same manner (*not* because they did not use it). A possible source for the discrepancy with our results (in which the exclusive use of the base rate for the nondiagnostic condition was 54%) could be the difference in the procedure: Our participants received only one description whereas in the standard procedure participants received the nondiagnostic description mixed with diagnostic ones. The results of a number of studies

(in Ginossar & Trope, 1987, 1987; Gigerenzer & Murray, 1987) point to the existence of order effects. They support the view (equally endorsed by Bar-Hillel, 1983) that Kahneman & Tversky's (1973) result for the nondiagnostic description was due to the repeated measures design that they used, so that participants focused their attention on the descriptions that varied across problems. Likewise, Fischhoff, Slovic & Lichtenstein (1979) showed that when the base rate is varied within-subjects instead of the description performance reflected the base rates.

General Discussion

Three main conclusions can be drawn from the experiment. They concern the relevance-based explanation of performance, the integration of information, and the representativeness explanation. We take up these issues in turn.

Relevance and alternative explanations

Our first conclusion is that the use of the base rate depends on how the task is mentally represented, which in turn depends on considerations of relevance. Knowledge of the typicality of the psychological description is exploited only to the extent that participants consider it to be relevant. About three quarters of the participants (in the *Diagnostic description* and *Diagnostic description + Statement* conditions) took the diagnostic information into account when they had to extract it from a typical description. However, when the diagnostic information was condensed in the form of a verbal statement *(No description + Statement* condition), only 15% of the participants took it into account. The percentage was intermediate (about 50%) for those who were presented with the nondiagnostic description (suggesting that up to one half of this latter group assumed it was diagnostic in order to render it relevant). These results extend the observations made by Schwarz & al. (1991). While the size of the effect of the manipulation reported by these authors cannot be assessed, the present experiment reveals a shift of about two thirds of the population in the use of individuating information between the two critical conditions, namely the *Diagnostic description* and the *No description + Statement* conditions.

Is there an alternative explanation of the performance in the experimental group? It may be tempting to attribute the results to the difference in salience of the individuating information between the

conditions with a description and the *No description + Statement* condition. There are two reasons to reject this hypothesis. First, in order to accommodate all the results, the salience hypothesis should also explain why most participants apparently disregarded the statement in the *No description + Statement* condition. An explanation in terms of a lack of salience seems highly implausible because it implies that nearly all participants presented with the two sentences, "A description has been chosen at random from the 100 available descriptions. The psychological description is typical of an engineer", systematically overlooked the second sentence. Rather, we believe that participants did process the second sentence, but in the end discarded it because its content makes the problem question irrelevant. In effect, participants are facing a request to express numerically the probability that a typical description of an engineer is that of an engineer. For those who understand the difference between likelihood and posterior probability, the question is trivial (although awkward to answer numerically). For those who mix up these concepts, the question contains the answer: In both cases, the problem question lacks relevance. In such a situation, participants will try to reinterpret the question. The most relevant way (which follows a least effort path to producing an answer) is to exploit any other cue provided by the experimenter: The base rate is the most conspicuous piece of information to use. The second and more important reason why the salience hypothesis fails is that it does not make a distinction between the diagnostic (salient) information and the typicality estimate. The diagnostic information, whether it is salient or not, is not the basis for the final probability estimate. The probability estimate is inferred from the typicality estimate, which acts as a mediator between the diagnostic information (whatever its degree of salience may be) and the probability estimate.

There is an alternative version of the salience hypothesis that *does* take into account the distinction between diagnostic information and typicality estimate. It claims that participants who have paid attention to salient information will attach greater importance to the subsequent typicality estimate. But this claim is just an imprecise and pre-theoretical version of the relevance explanation which says that the description carries with it a presumption of relevance, that is, a guarantee for the

reader that it is worth processing the description in order to infer new information (or in the present case, in order to arrive at the solution). So, not only does the relevance approach have much to say about the task, it also explains the effect of information salience.

Another possible explanation of the results is based on the notion of vividness: The *Diagnostic description* presented information that is vivid, concrete and specific, whereas the *No description* condition presented pallid, abstract and general information. The salience hypothesis concerns the focusing of attention. But the difference in vividness concerns the subjective weight of the information, which should be greater when there is a description than when there is no description. There are two consequences. First, up to here the analysis of the evidence could result in a degree of belief in the typicality of the description that is greater in the *description* conditions than in the *No description* condition. But it is not sure at all that a belief inferred by the participant's own consideration of the information should be greater than the belief in a statement provided by the experimenter: one could claim quite the contrary. Second, the typicality estimate itself would be higher; for instance, participants in the *description* conditions would infer a belief that the description is *highly* typical, while those in the *No description* conditions were just told that "the description is typical". But it is very doubtful that a necessarily limited difference of this sort (given that "typical" is linguistically marked) could account for the sharp difference in performance that was found. Finally, if the vividness hypothesis has something to say about the belief in the typicality of the description, it cannot explain, contrary to the relevance explanation, performance in the *No description* condition for the same reasons that the salience hypothesis fails: The statement must be processed, unless it is passed unnoticed, which is very implausible.

Integration of information

The second conclusion has two facets, one of which sheds new light on an old paradigm. On the one hand, in the pilot study referred to earlier, in which justifications were asked (and that was otherwise nearly identical to the present experiment), the analysis indicates that nearly all the responses that differ from the base rate (that is, responses that take the individuating information into account) were justified without any reference to the base rate: This confirms the base rate neglect but our

explanation of it differs sharply from the heuristic explanation. In our view, the neglect does not imply an intrinsic bias, but an effect of the interpretation of what the task is about; actually, the base rate is widely considered when the participants are not diverted from the objective of the task as is the case in the standard problem. On the other hand, the experiment has revealed a symmetrical effect that seems to have passed unnoticed so far. In effect, the proportion of responses equal to the base rate, that is, the proportion of responses showing disregard of the individuating information, is surprisingly high. In the two conditions that had a diagnostic description (with or without the statement of typicality) the proportion of responses equal to the base rate reached about one quarter of the responses, even though the description instantiated the social stereotype. In the pilot study in which a justification was asked, about three quarters of the estimates equal to the base rate were justified by a mention of the proportion of engineers (or by an appeal to a procedure of computation), that is, without any reference to the individuating information which, it is reminded, was always present either in the form of a description, or in the form of a statement of typicality, or both. In other words, the present manipulation has revealed a mirror effect to the neglect of the base rate, namely the neglect of the individuating information. We claim that this occurs because the individuating information becomes irrelevant when it is realized that it looks like a direct answer to the question. There is a slight asymmetry between the two cases: In one case, the base rate is considered irrelevant because participants think that they have fulfilled the task; in the other case, the individuating information is considered irrelevant because if it were taken into consideration, there would be no task to fulfill. Upon being told that the description is typical, participants infer a probability estimate commensurate with the value verbally provided to them. But then, this too easy inference looks very much like an answer to the question, which makes it lack relevance. Except for those who were sophisticated enough to try to combine this qualitative information with the quantitative one, they reinterpreted the question as a request for an unconditioned probability, which enabled them to render the base rate information relevant and to fulfill the task, so that they gave the base rate as their response.

The notion that information should be perceived as relevant in order to be used has already been considered to explain base rate neglect (Bar-Hillel, 1980; Krosnick, Li, & Lehman, 1990). In our opinion, this notion applies to the individuating information as well. The present investigation contributes to redress the balance in a paradigm traditionally considered from a single point of view, demonstrating both the neglect of the base rate and the neglect of the individuating information.

We believe that these neglects are the two sides of the same coin: They reflect the lack of integration of two disconnected sources of information. The individuating information is qualitative and actively extracted because covertly provided. The base rate information is quantitative and passively exploited because overtly provided. The one or the other neglect occurs depending on the task representation, that is, presumed relevance. In the original problem, the emphasis is put on the description to exploit, so that participants interpret the task as a request to act as a psychologist and once they have extracted the individuating information and inferred a typicality estimate, the implicit contract by which they are linked with the experimenter has been fulfilled. Further information, especially mathematical, is irrelevant, both in the intuitive and theoretical senses. Similarly, in the new version of the problem introduced here, which resembles a mathematical problem, the statement of typicality provided to participants looks like an answer to the question, which would break the same contract if it were taken into consideration; they consider it as irrelevant, and fall back to the numerical information (the base rate) which, in their turn, they fail to combine with the other source of information.

The notion that there is a lack of integration of different sources of information is not new. Ginossar & Trope (1987) obtained increased use of base rate when the base rate value and the description were integrated and presented in the form of a list. But their data, based on the traditional dependent variable (the mean probability estimate) do not indicate which proportion of participants was affected by the manipulation. Kahneman and Tversky (1973, p. 243) themselves commented on the lack of integration:

> Our subjects, however, failed to integrate prior probability with specific evidence. [. . .] The failure to appreciate the

> relevance of prior probability in the presence of specific evidence is perhaps one of the most significant departures of intuition from the normative theory of prediction.

We concur with this quote; but whereas its authors just made a correct *description* using the pre-theoretical term of *relevance,* we have offered an *explanation* for this observation using the same term in its technical, theoretical sense.

Representativeness

Could representativeness account for the difference in performance between participants in the *No description + Statement* condition and those in the *Diagnostic description* condition? If the operation of the representativeness heuristic consists in the "assessment of the degree of correspondence between [. . .] an outcome and a model" (Tversky and Kahneman, 1983, p. 295), performance in the *No description + Statement* condition clearly contradicts this view. All the participants in this condition were provided with an explicit statement of strong correspondence between the outcome and the model. Therefore all of them should have opted for a high probability and in particular none of them should have given an answer equal to .30. Of course, one might object that there is a difference between (i) basing the judgment of correspondence with the model on a synthetic statement that provides the assessment ready made (as was necessarily the case in the *No description + Statement* condition if representativeness is to be used) and, (ii) working out the degree of correspondence by considering the various diagnostic features. In this view, representativeness would require to actively work out the final judgment of similarity, and would be blocked otherwise. But it is not quite clear why. Furthermore, this is close to our main claim that need no representativeness hypothesis: People who have been active in exploiting the description do not feel the need to use extra information.

There are further reasons which render the representativeness explanation inadequate. As we have shown, it cannot explain the neglect of the individuating information: The representativeness explanation is incomplete and not parsimonious. More seriously it begs the question because in order to solve the Lawyer-Engineer problem, it is necessary to make a comparison between the description and the stereotype: The most

serious conceptual flaw in this approach is the use of the description of a necessary step in the resolution of the problem as an explanatory concept.

A tentative outlook of how people solve the problem.

We will propose an overview of how participants process the problem; although it has not been directly tested, the results are compatible with the idea that solving the standard problem requires a two-step process.

Step one requires to decide to which extent the set of features is characteristic of an engineer. It results in a degree of typicality, which all participants can arrive at. Step two consists in answering the question proper, that is, given the degree of fit to the stereotype worked out at step one, what is the probability that the description is that of an engineer? Once participants have arrived at a high estimate of typicality (step one), the question can receive two different interpretations.

(i) The high typicality value suggests that the description is that of an engineer, in other words the question seems to have been answered; participants stop on the way and the typicality value is taken to be the final response. This is possible because participants feel that, in exploiting the description, they have fulfilled the task requirement and satisfied the presumption of relevance of the description. The base rate neglect reflects a lack of interest for further information that is viewed as irrelevant. Notice that once the response is given, step one can be described in terms of the representativeness heuristic since the response is based on the working out of a typicality estimate. But we believe that this step is a necessary (and insufficient) step on the way to the solution, the bias which consists in, and is explained by, stopping at this step being artificially induced by the task representation.

(ii) The question amounts to asking what is the probability that a description typical of an engineer is that of an engineer: It is viewed as pointless or tautological. But the question can be normalized into "What is the probability to draw the description of an engineer". Participants readily find an answer in the base rate which has not been used, and this satisfies the presumption of its relevance. They make an exclusive use of the base rate, failing to integrate the two sources of information. Again their lack of interest for one of the sources of information (the individuating information) stems from its apparent irrelevance. The classic

control problem (the non-diagnostic description) leads, after step one, to the absence of a typicality value. This cannot provide the basis for an answer and participants interpret the question as "What is the probability that the description is that of an engineer" based on no other information than the base rate. The problem in the present *No description* condition leads participants to step two directly and to the second interpretation of the question.

Of course, the lack of integration of the two sources of information that results from task representation depends on an inherent difficulty in integrating them. Participants untutored in probability theory have difficulty to cope with problems that require probabilistic revision: Performance on less deceptive tasks is not perfect, even after the defects in their formulation have been corrected (Macchi, 1995, 2000). One can only regret that so much research has been devoted to a task that artificially enhances a difficulty, the origin of which could be studied better by trying to control it, for instance by drawing, to various degrees, participants' attention to the necessity of combining the information.

Our theoretical approach is based on the notion of task representation. This notion hinges upon processes of attribution made by the participant to the experimenter, which is but a specific example of the effects of social factors on thinking and reasoning. In view of Paolo Legrenzi's long standing interest and contribution to that stream of research, we are happy to dedicate the present work to him.

References

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

Bar-Hillel, M. (1983). The base rate fallacy controversy. In R. W. Scholz (Ed.), *Decision making under uncertainty* (pp. 39-61). Amsterdam: North-Holland.

Dulany, D. E. , & Hilton, D. J. (1991). Conversational implicature, conscious representation, and the conjunction fallacy. *Social Cognition, 9,* 85-110.

Evans, J. St. B. T. , & Pollard, P. (1982). Statistical judgment: A further test of the representativeness construct. *Acta Psychologica, 51,* 91-103.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: beyond "heuristics and biases". *European Review of Social Psychology, 2,* 83-115.

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. , & Blank, 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.

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,* 474-484.

Girotto, V. , Kemmelmeier, M. , Sperber, D. , & Van der Henst, J.-B. (2001). Inept reasoners or pragmatic virtuosos? Relevance and the deontic selection task. *Cognition, 81,* B69-B76.

Grice, P. (1989). *Studies in the way of words.* Cambridge: Harvard University Press.

Hilton, D. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118,* 248-271.

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

Kahneman, D. , & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review, 103,* 582-591.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral 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.

Macchi, L. (1995). Pragmatic aspects of the base-rate fallacy. *Quarterly Journal of Experimental Psychology, 48A,* 188-207.

Macchi, L. (2000). Partitive formulation of information in probabilistic problems: beyond heuristics and frequency format explanations. *Organizational Behavior and Human Decision Processes*, *82,* 217-236.

Mosconi, G. (1990). *Discorso e Pensiero*. [Discourse and thought]. Bologna: Il Mulino.

Mosconi, G. & Macchi, L. (in press). The role of pragmatic rules in the conjunction fallacy. *Mind and Society, 3.*

Norenzayan, A. , & Schwarz, N. (1999). Telling what they want to know: participants tailor causal attributions to researchers' interests. *European Journal of Social Psychology, 29,* 1911-1020.

Olson, C. L. (1976). Some apparent violations of the representativeness heuristic in human judgment. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 599-608.

Orne, M. T. (1969). Demand characteristics and the concept of quasi-controls. In R. Rosenthal & R. I. Rosnow (Eds.), *Artifact in behavioral research* (pp. 143-179). New York: Academic Press.

Politzer, G. (1986). Laws of language use and formal logic. *Journal of Psycholinguistic Research, 15,* 47-92.

Politzer, G. (1993). *La psychologie du raisonnement: lois de la pragmatique et logique formelle.* [The psychology of reasoning: Laws of pragmatics and formal logic]. Doctorate Dissertation. University of Paris VIII.

Politzer, G. (in press a). Reasoning, judgment, and pragmatics. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics.* London: Palgrave.

Politzer, G. (in press b). Premise interpretation in conditional reasoning. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment, and decision making.* London: Wiley.

Politzer, G. , & Macchi, L. (2000). Reasoning and pragmatics. *Mind and Society, 1,* 73-93.

Politzer, G. , & Noveck, I. A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research, 20,* 83-103.

Schubauer-Leoni, M. L. , & Perret-Clermont, A.-N. (1997). Social interactions and mathematics learning. In T. Nunes & P. Bryant (Eds.), *Learning and teaching mathematics. An international perspective* (pp. 265-283). Hove: Psychology Press.

Schwarz, N. (1996). *Cognition and communication: Judgmental biases, research methods, and the logic of conversation.* Mahwah, N. J. : Lawrence Erlbaum.

Schwarz, N. , Strack, F. , Hilton, D. , & Naderer, G. (1991). Base rates, representativeness, and the logic of conversation: The contextual relevance of "irrelevant" information. *Social Cognition, 9,* 67-84.

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 Edition. London: Blackwell.

Tversky, A. & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90,* 293-315.

Van der Henst, J.-B. , Sperber, D. , & Politzer, G. (2002). When is a conclusion worth deriving ? A relevance-based analysis of indeterminate relational problems.*Thinking and Reasoning, 8,* 1-20.

Van der Henst, J.-B. , Rossi, S. , & Schroyens, W. (2002). When participants are not misled, they are not so bad after all: A pragmatic analysis of a rule discovery task. *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. Mahwah, N. J. : Lawrence Erlbaum.

Table 1. Percentage of responses equal to the base rate for each condition and level of significance of the differences in frequency.

statement of typicality			
present (novel condition) description		absent description	
none (c)	diagnostic (d)	nondiagnostic (b)	diagnostic (a)
(N = 26)	(N = 29)	(N = 28)	(N = 30)
.85	.31	.54	.17

<. . . . . p <. 001 . . . . . >

<. . . . p <. 001 . . . .>

<. . . . . p <. 01 . . . . . >

<. . . . . N.S. . . . . .>

Appendix

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.

Condition (a): *Diagnostic description*

> You will find below a description, chosen at random from the 100 available descriptions. Read this description and indicate your probability that the person described is an engineer.
>
> > Paolo is a 45 years old man. He is married and has two children. He is generally conservative, careful, and ambitious. He does not care very much about his look. He shows no interest in political and social issues and he is addicted to his computer. He designed his Hi-Fi set himself. He spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.
>
> The probability that the description is that of an engineer is ____ (indicate a numerical value).

Condition (b): *Nondiagnostic description* (control).

> Andrea is a 30 years old man. He is married with no children. A man of high ability and high motivation, he promises to be quite successful in his field. He is well liked by his colleagues.

Condition (c): *No description + Statement of Typicality*.

> A description has been chosen at random from the 100 available descriptions. The psychological description is typical of an engineer. Indicate your probability that the person described is an engineer.

Condition (d): *Diagnostic description + Statement of Typicality* (control).

> You will find below a u, chosen at random from the 100 available descriptions. The psychological description is typical of an engineer, as you can see. Indicate your probability that the person described is an engineer. *[description follows like in condition (a)].*