Memory & Cognition
2000, 28 (1), 79-91

Not only base rates are neglected in the Engineer–Lawyer problem: An investigation of reasoners' underutilization of complementarity

JEAN BARATGIN
Centre de Recherche en Epistémologie Appliquée, Paris, France

and

IRA A. NOVECK
Centre de Recherche en Epistémologie Appliquée, Paris, France
and Université de Grenoble II, Grenoble, France

The standard Engineer–Lawyer problem (Kahneman & Tversky, 1973) points to the failure of reasoners to integrate mentioned base-rate information in arriving at likelihood estimates. Research in this area nevertheless has presupposed that participants respect complementarity (i.e., participants ensure that competing estimates add up to 100%). A survey of the literature leads us to doubt this presupposition. We propose that the participants' non-normative performance on the standard Engineer–Lawyer problem reflects a reluctance to view the task probabilistically and that normative responses become more prominent as probabilistic aspects of the task do. In the present experiments, we manipulated two kinds of probabilistic cues and determined the extent to which (1) base rates were integrated and (2) the complementarity constraint was respected. In Experiment 1, six versions of an Engineer–Lawyer-type problem (that varied three levels of cue to complementarity and two base rates) were presented. The results showed that base-rate integration increased as cues to complementarity did. Experiment 2 confirmed that Gigerenzer, Hell, and Blank's (1988) random-draw paradigm facilitates base-rate integration; a second measure revealed that it also prompts respect for complementarity. In Experiment 3, we replicated two of our main findings in one procedure while controlling for the potential influence of extraneous task features. We discuss approaches that describe how probabilistic cues might prompt normative responding.

When one needs to consider the likelihood of an outcome, base rates (prior probabilities) become critical pieces of information. For instance, if a survey indicates that a candidate for public office is preferred by 80% of a particular voting bloc, it would then be important to know what the base rate of this bloc is with respect to the general voting population. Obviously, the candidate's chances for election would be considerably higher if the bloc's base rate were 70% of the voting population rather than 30%.

Much research in the psychological literature points out that participants often neglect base-rate information. In this paper, we focus on one well-known task that highlights this neglect—the Engineer–Lawyer problem (Kahneman & Tversky, 1973). In the original version of Kahneman and Tversky's problem, participants are presented with five written portraits of people drawn randomly from a population of 70 lawyers and 30 engineers. In one condition, the participants are asked to estimate the probability that each person described is one of the 70 lawyers. In another condition, the task is the same but the base rates are reversed (i.e., there are 30 lawyers and 70 engineers) and the participants are asked to estimate the probability that the person described is one of the 30 lawyers.

Kahneman and Tversky drew two claims from their findings. The first, which is well known in the probabilistic reasoning literature, is that participants generally neglect base-rate information and establish their estimates by using the *representativeness* heuristic (i.e., their estimates are based on how well the portraits match the stereotype for an engineer or a lawyer; Kahneman & Tversky, 1972, 1973; Tversky & Kahneman, 1982). Supported by findings such as these, heuristic accounts of reasoning have been very influential in the psychological literature (see, e.g., Piatelli-Palmarini, 1993; Slovic, Lichtenstein, & Fischhoff, 1988; von Winterfeldt & Edwards, 1986).

This first claim has led to some controversy. Many argue that other factors, having little to do with a heuris-

This work was partially funded by a grant from DRET (La Direction de la Recherche et de la Technologie, 94/074) in France. The authors thank the members of le GRICE (Groupe de Recherche sur l'Inférence et la Compréhension Elémentaire) for their generous comments: Daniel Andler, Francesco Cara, Guy Politzer, Dan Sperber, and Jean-Baptiste van der Henst. We also thank Gerd Gigerenzer, John Wixted, and an anonymous reviewer for their helpful critiques of an earlier version of this manuscript. Special thanks to Dominique Boutet and the students at IUT of Bretigny sur Orge for their participation in the experiments J. Baratgin is now at l'Université de la Méditerranée (Marseilles). I. Noveck is now at l'Institut des Sciences Cognitives (Lyon). Correspondence concerning this article may be directed to J. Baratgin, CREA-École Polytechnique, 1 rue Descartes, 75005 Paris, France (e-mail: baratgin@ poly.polytechnique.fr).

 Copyright 2000 Psychonomic Society, Inc.

tic explanation, may be responsible for base-rate neglect (for a survey, see Girotto, 1994; Koehler, 1996). These factors focus mostly on the presentation of the task. Base-rate information is heeded more reliably, for example, when portraits are presented before the base rates (Krosnick, Li, & Lehman, 1990) or when the experimental context is more conversationally cooperative (Schwarz, Strack, Hilton, and Naderer, 1991; Zukier & Pepitone, 1984; for a review, see Hilton, 1995). Thus, many researchers are dubious about the implications of Kahneman and Tversky's studies, most notably Gigerenzer and his colleagues (Gigerenzer, 1991, 1993, 1994, 1996; Gigerenzer et al., 1989). In one study exemplary of Gigerenzer's objections, Gigerenzer, Hell, and Blank (1988) showed how the participants' estimates revealed consideration of base rates significantly more often when the participants actually drew the portraits from urns whose contents were known to have specific a priori base rates. Our aim in the present work is not to take sides in this debate but to better understand what prevents participants from using base-rate information on problems of this sort. This leads us to Kahneman and Tversky's second, lesser-known, claim—that participants nevertheless comply with the task's complementarity constraint.

When two groups of participants were presented identical base rates (e.g., 30 lawyers and 70 engineers) and one group was asked to estimate the probability that a given portrait was one of the 30 lawyers and the other group was asked to estimate the probability that the portrait was one of the 70 engineers, Kahneman and Tversky found that the participants' estimates of the two groups to be complementary (i.e., estimates from these two groups added up to 100%). This makes sense because laws of probability indicate that $P(\text{Lawyers}) = 1 - P(\text{Engineers})$. That is, if participants were to be asked to estimate the likelihood that an engineer-like portrait was one of the 30 engineers, one would find median estimates to be, say, 90%, and when asked whether the portrait was one of the 70 lawyers, median estimates would be 10%. This indicated that the participants did respect the complementarity constraint, even as they ignored base rates. From that moment forward, Kahneman and Tversky requested participants to supply estimates for only one of the occupations (say, only for engineers) because it was assumed that estimate requests in the complementary condition would have produced complementary results. Most researchers in the field have followed this practice (e.g., Carroll & Siegler, 1977; Gigerenzer et al., 1988; Schwarz et al., 1991; Wells & Harvey, 1978), which we call an *asymmetric* approach. As we will show, this practice may well be based on a false positive result, and it prevents one from fully understanding the reasoning processes involved in the Engineer–Lawyer task.

Kahneman and Tversky's complementarity claim is questionable for the following two reasons. First, evidence for complementarity—when reported—has not been verified. Swieringa, Gibbins, Larsson, and Sweeney (1976) did not replicate the complementarity finding with calculated medians, and Krosnick et al. (1990) failed to replicate the finding with means. Second, there is direct evidence revealing the tendency of participants not to comply with the complementarity constraint on the Engineer–Lawyer task when an estimate for each of the two hypotheses is requested. In a *symmetric* design, Davidson and Hirtle (1990) required participants to give two likelihood judgments for each portrait: one that determined the likelihood that the person in question was an engineer, and another that the person was a lawyer. Before normalizing their data, they found that of 120 pairs of estimates, 24 pairs (20%) added up to amounts other than 100% and that the sums ranged from 90% to 120%. Despite such findings, authors have not drawn any link between their results and the reasoning processes involved; research done with this task typically ignores or normalizes this sort of violation. We assume that the failure to respect the complementarity constraint—what we will call the *neglect to respect complementarity*—occurs frequently when participants make estimates in the standard Engineer–Lawyer problem and that this indicates that there is more going on with the task than simply a failure to consider base rates.

The tendency of participants to neglect considerations of complementarity has been investigated independently (Alberoni, 1962; Asare & Wright, 1995; Cohen, Dearnaley, & Hansel, 1956; Einhorn & Hogarth, 1985; Marks & Clarkson, 1972; Teigen, 1974, 1983; Van Wallendael & Hastie, 1990). When there are two alternatives, one finds both supra-additivity, where the sum of participants' estimates is over 100%, and subadditivity, where the sum of their estimates is below 100%. The literature reveals that neglect grows with increased task complexity and that neglect can be eliminated. Although neglect to respect complementarity does not appear to be as severe as base-rate neglect, it is important to bear in mind that once an experiment tests for complementarity it implicitly cues it. Thus, a tendency to neglect complementarity on standard versions of the Engineer–Lawyer task may well be more common than one is led to believe by findings such as Davidson and Hirtle's (1990).

A normative analysis of the problem shows how an underappreciation of alternativeness can affect participants' estimates on tasks such as the Engineer–Lawyer problem.[1] Bayes's Theorem in Equation 1, which is the standard reference for such problems, indicates that there are two places that require normative reasoners to consider competing hypotheses in approaching the Engineer–Lawyer problem. In the event that participants do not consider the alternative hypothesis, the denominator is incomplete, causing their estimates to depart from Bayesian norms. Similarly, in the event that participants do not consider complementarity, $P(\text{H}\,|\,\text{D}) + P(\text{not H}\,|\,\text{D}) = 1$, their estimates will also vary from those predicted by Bayesian norms.

$$P(\mathrm{H}\,|\,\mathrm{D}) = \frac{P(\mathrm{D}\,|\,\mathrm{H})P(\mathrm{H})}{P(\mathrm{D}\,|\,\mathrm{H})P(\mathrm{H})+P(\mathrm{D}\,|\,\text{not-H})P(\text{not-H})} \quad (1)$$

As stated earlier, we find it untenable to suppose, as many in the literature do, that participants do not appreciate the alternativeness implicit in base rates but that they do appreciate the alternativeness implicit in the complementarity constraint. This is why the goal of the present study is first to verify that complementarity to 100% tends to be neglected and then to investigate how efforts to counter this neglect can facilitate normative responses on the Engineer–Lawyer problem.

We propose that the consideration of complementarity on the one hand and base rates on the other are part and parcel of a unitary cognitive ability to think probabilistically. Furthermore, we propose that effort directed toward the portraits has the potential to draw away the resources required to reason probabilistically (Baratgin, 1999). The standard Engineer–Lawyer problem, in our view, leads to a low occurrence of base-rate integration because participants focus their resources on the portraits and not on the problem's probabilistic aspects. Therefore, we predict that attention paid to relevant probabilistic notions (whether based on complementarity or on base rates) should prompt higher rates of normative responding on such tasks. This interpretation motivates our experiments.

In Experiment 1, we exploited the participants' accessibility to probabilistic reasoning by providing two levels of a cue to complementarity on the Engineer–Lawyer-type problems. We proposed that such cues would induce normative responses because they would compel the participants to take into account probabilistic aspects of the task in general. These two cues differed in degree. Therefore, we expected a strong cue to complementarity to prompt higher rates of normative responding than a less compelling one. In Experiments 2 and 3, we took the converse approach: We investigated a version of the problem that had been previously shown to readily prompt base-rate integration—the Urn problem (Gigerenzer et al., 1988)—so that we could determine the extent to which it would prompt the participants to respect complementarity. We anticipated that the Urn problem, which highlights probabilistic features, would prompt relatively few complementarity violations.

EXPERIMENT 1

We adapted the Engineer–Lawyer problem to French participants in order to preserve its high degree of diagnosticity (Fischhoff & Bar-Hillel, 1984). We chose two occupations whose stereotypes are clearer to French sensibilities, that of a mathematics teacher and that of a French literature teacher. This variation of Kahneman and Tversky's Engineer–Lawyer problem was presented to participants under one of three conditions. In the *standard* condition, the participants were required to estimate the likelihood that a description reflected one hypothesis (i.e., math teacher). In the *suggested-complementarity* condition, the participants were required to estimate the likelihood that the given description reflected one of the hypotheses (math teacher) and then the other (French literature teacher). The *induced-complementarity* condition was identical to the suggested-complementarity condition except that the participants were further reminded that their estimates had to add up to 100%. The induced-complementarity condition was expected to yield estimates closest to those anticipated by Bayesian reasoning.

Method

Participants. One hundred twenty 1st-year students enrolled in a general communication course at the Institut Universitaire Technique (IUT) de Bretigny sur Orge, France, took part in the experiment. They had no knowledge about studies on base-rate neglect or Bayesian theory. They were tested in small groups.

Procedure and Materials. As in Kahneman and Tversky's (1973) study, the participants in the low-base-rate condition received the following instructions (translated from French):

> The ministry of education conducted a research project concerning the psychological profile of school teachers. A panel of psychologists interviewed and administered personality tests to 100 teachers. There were 30 teachers of mathematics and 70 teachers of French literature. For each teacher, a portrait was written by a psychologist. You will find on your forms five portraits, chosen at random from the 100 available portraits. For each portrait, please indicate the chance (in percent) that the person described is one of the 30 math teachers. You are asked to give your answer on a scale from 0% to 100%.

The high-base-rate condition differed only in that the participants were told there were 70 math teachers and 30 French literature teachers and that they had to indicate the chances that the person described was one of the 70 math teachers. In the the remainder of this section, the low-base-rate condition will be used as an example.

The introductory instruction was followed by the presentation of five psychological portraits tested previously in a pilot study. Two portraits were designed to have a strong informative message and thus to be highly diagnosing. One portrait implied that the person described (Jacques) was a math teacher; a second (Anne) indicated a French literature teacher. A third portrait (Françoise) was designed to be less diagnostic yet suggestive of a French literature teacher. Her portrait was drawn from a real portrait (in Gigerenzer et al., 1988). A fourth (Paul) was uninformative. The fifth portrait (Raphaël) was informative but inconsistent (as in Ginosar & Trope, 1980). Questionnaires concerning the five portraits were prepared in one of five different random orders. Below are the five portraits that were presented:

> Jacques is 45 years old. He is married and has four children. He has a moderate interest in politics and current affairs. He spends most of his spare time doing carpentry, crosswords, and yachting.

> Anne is 28 years old. She is single, cheerful, and enthusiastic about her job. She has a special interest in fine arts. She runs a small gallery in her spare time where she exhibits the work of young artists. She also likes the seaside and to take walks.

> Françoise is 48 years old and has remained single. She is down to earth, earnest, and seems to be appreciated by her pupils. She is very politically engaged and is active in a trade union. Her main hobbies are mountain hiking and traveling to distant countries.

> Paul is 50 years old. He is married and has two children. He is a man of a great intellectual capacity and is very motivated. He is very successful at his job. He is appreciated by his colleagues.

> Raphaël is 35 years old. He is separated from his wife and is raising his daughter by himself. He is rather outgoing and likes jokes and plays on words. He enjoys going to the theater and visiting museums. He is a good chess player, as well. He has started a chess club in his school.

The participants were required to estimate for each portrait the chance (in percent) that the person described was a mathematics teacher. For example, with respect to the portrait of Anne in the low-base-rate condition, the participants' estimates were requested in the following manner:

> The chance (in percent) that Anne is one of the 30 teachers of mathematics is __%.

In the suggested-complementarity/low-base-rate condition, the procedure was identical except that the second to last sentence of the introductory paragraph ended with "and the chance (in percent) that the person described is one of the 70 teachers of French literature." The task instruction read as follows (e.g., for Anne):

> The chance (in percent) that Anne is one of the 30 teachers of mathematics is ___%.
>
> The chance (in percent) that Anne is one of the 70 teachers of French literature is ____%.

The induced-complementarity condition was identical to the suggested-complementarity condition except that the participants were given the following instructions that required them to provide probabilities that added up to 100%.

> Please note that the sum of the two percentages must be 100%. For example, Anne is either a mathematics teacher or a French literature teacher.

Finally, on a separate page, all the participants were asked to assess their confidence in their answers. They were told that some experts (not the psychologists who had written the portraits) had been given the same exercise with the same five portraits. The participants had to rate on a scale, ranging from 0% to 100%, the chance that they had given the same answers as the experts for each portrait and more generally over the entire task. They were entitled to review their original estimates but were not allowed to change them. Most of the participants, however, provided their confidence ratings without a review.

Design. We used a 2 (high [70%] vs. low [30%] base rate of mathematics teachers) × 3 (task instructions: standard, suggested complementarity, and induced complementarity) between-subjects design. Each participant was presented the set of portraits and was assigned randomly to one of the six conditions. There were 20 participants in each condition.

Results

We followed Wells and Harvey (1978) and Gigerenzer et al. (1988) by applying a set of fine-grained measures based on the participants' actual performance and by determining how the same participants would have performed under idealized Bayesian conditions. Although descriptive statistics often prove to be sufficiently powerful, we provided a few critical statistical tests to back up the claims and to provide appropriate landmarks. We then analyzed the participants' performance with respect to complementarity violations (where applicable) and with respect to provided confidence judgments. A summary of the results is presented in Table 1.

In order to determine the extent to which the participants used base-rate information, we assessed the degree to which the two associated groups (high/low base rates) in a given task-instruction condition yielded different

Table 1
Means of Participants' Estimates (Chance in Percent That the Portrait Is That of a Math Teacher) and Scores Based on Bayesian-Derived Estimates (Last Six Columns) (N = 120)

Portraits	P_{Low}	P_{High}	DMHL	BE−AE$_{Low}$	BE−AE$_{High}$	BE−AE$_{Mean}$	b_{Low}	b_{High}	b_{Mean}
Standard Condition									
Anne	20.4	29.9	9.5	10.2	21.7	16.0	48	30	39
Françoise	38.2	44.2	6.0	23.0	25.1	24.1	21	19	20
Jacques	64.9	73.2	8.3	23.7	15.0	19.4	26	35	31
Paul	52.9	59.1	6.3	25.2	25.1	25.1	20	20	20
Raphaël	40.6	44.5	4.0	22.9	29.0	26.0	15	12	13
Mean	43.4	50.2	6.8	21.0	23.2	22.1	24	23	24
Suggested-Complementarity Condition									
Anne	19.1	32.1	13.0	8.6	17.9	13.3	60	42	51
Françoise	40.0	49.6	9.6	22.1	20.4	21.3	30	32	31
Jacques	60.3	74.3	14.1	18.5	8.3	13.4	43	63	53
Paul	46.0	57.0	11.0	17.0	21.1	19.1	39	34	37
Raphaël	37.5	45.9	8.4	16.3	25.9	21.1	34	24	29
Mean	40.6	51.8	11.2	16.5	18.7	17.6	40	37	39
Induced-Complementarity Condition									
Anne	17.8	34.5	16.8	5.7	13.4	9.5	75	56	65
Françoise	37.5	53.3	15.8	16.7	17.1	16.9	49	48	48
Jacques	56.8	76.3	19.5	14.2	7.3	10.7	58	73	65
Paul	41.0	60.0	19.0	11.8	9.4	10.6	62	67	64
Raphaël	36.3	52.8	16.5	12.1	16.8	14.4	58	50	54
Mean	37.9	55.4	17.5	12.1	12.8	12.4	59	58	58

Note—The *Low* and *High* subscripts refer to an experimental condition's base rate of mathematics teachers. Thus, the P_{Low} column contains the participants' (mean) estimate that the row's portrait is that of a mathematics teacher in the low-base-rate condition. P_{High} presents such mean estimates in the high-base-rate condition. The DMHL is the difference between the means of the high and low base rates. BE−AE is the difference between the Bayesian estimate and the actual estimate. The index b varies from 0 (no indication of base-rate utilization) to 100 (perfect Bayesian treatment of base-rate information).

scores for each portrait. In the event that the participants did not use base rates, the two groups' estimates should have been similar. The first column of Table 1 provides the mean of the participants' estimates that the person described was a mathematics teacher, given the low-base-rate condition. The second column presents the mean of participants' estimates in the high-base-rate condition. For each portrait, we compared the difference between the mean based on the percentages given by the participants in the high-base-rate group and the one worked out for the low-base-rate group. We call this the *difference between the means of the high- and low-base-rate groups*, or *DMHL*. The DMHLs appear in the third column of Table 1.

In the standard condition, the overall DMHL (across the five portraits) was 6.8%. This difference indicates that our findings are in line with results from previous studies; Kahneman and Tversky (1973) found a DMHL of 5%, and Gigerenzer et al. (1988) found a DMHL of 9.8%. In the suggested-complementarity condition, the difference averaged across the five portraits was substantially higher (DMHL = 11.2%) than in the standard condition. This indicated that the moderate cue to consider both hypotheses led to a greater percentage of participants to include base-rate information. We analyzed separately the results of the participants who had complied with the complementarity constraint in the suggested-complementarity condition. The results are presented in Table 2. This subset of participants ($n = 29$) revealed a DMHL of 16.4%, and the difference between the high and low groups—for each of the five portraits—was sizable. These results are remarkably similar to those found in the induced-complementarity condition (DMHL = 17.5%), which led to the biggest overall difference between the high and low base rates. Clearly, base rates are integrated more as the cues to complementarity become more acute.

We then carried out a traditional analysis (tests of significance) on the DMHL data. For the standard condition, the mean DMHL of 6.8% was significantly different from predictions based on chance [$t(38) = 1.73$, $p < .05$], indicating that our participants generally did use base rates. However, none of the portraits, treated separately, yielded a DMHL that was significant at the .05 level. Two portraits (Anne and Jacques) yielded differences that were marginally significant (t tests yielding $p < .1$) and the other three portraits (Françoise, Paul, and Raphaël) yielded results that were more decidedly not significant. In the suggested-complementarity condition, the average DMHL proved highly significant [$t(38) = 2.84$, $p < .005$]. Two portraits (Anne and Jacques) yielded significant differences [$t(38) = 2.2$, $p < .025$; $t(38) = 1.79$, $p < .05$, respectively]; the other three were marginally significant. We then analyzed the results of the subset of participants ($n = 29$) who had complied with the complementarity constraint in the suggested-complementarity condition. The overall DMHL of 16.4% was significant [$t(27) = 5.16$, $p < .0005$], and significant differences were found between the high and low groups for each of the five portraits [$t(27) = 2.42$, $p < .01$ for Anne; $t(27) = 1.83$, $p < .05$ for Françoise; $t(27) = 2.11$, $p < .025$ for Jacques; $t(27) = 2.65$, $p < .01$ for Paul; $t(27) = 1.93$, $p < .05$ for Raphaël]. The induced-complementarity condition led to the biggest overall difference between the high- and low-base-rate groups [DMHL = 17.5%, $t(38) = 4.58$, $p < .0005$]. Each of the five portraits therein revealed significant differences with respect to DMHL [$t(38) = 2.67$, $p < .01$ for Anne; $t(38) = 2.55$, $p < .01$ for Françoise; $t(38) = 2.97$, $p < .005$ for Jacques; $t(38) = 2.34$, $p < .025$ for Paul; $t(38) = 2.09$, $p < .025$ for Raphaël].

To verify the above findings across conditions, we carried out a one-way between-subjects analysis of variance (ANOVA) in which *degree of cues to complementarity* was the independent variable and the DMHL for each of the portraits was treated as a dependent measure. The result was significant [$F(2,12) = 34.038$, $p < .001$]. Scheffé tests showed significant differences both between the standard- and the suggested-complementarity conditions ($p < .05$) and between the suggested- and induced-complementarity conditions ($p < .005$). There was also a significant difference between the DMHL for those ($n = 29$) who respected complementarity in the suggested-complementarity set and those ($n = 11$) who did not [$t(8) = 2.20$, $p < .05$].

We then analyzed the differences between the estimates provided by our participants and idealized Bayesian estimates worked out from Equation 1a. We followed Wells and Harvey's (1978) approach and calculated, for each participant, the probability estimate that a participant would have given if he/she had been in the other base-

Table 2
Means of Participants' Estimates (Chance in Percent That the Portrait Is That of a Math Teacher) and Scores Based on Bayesian-Derived Estimates (Last Six Columns) in the Suggested-Complementarity Condition After Having Removed Those Participants Who Violated the Complementarity Constraint ($N = 29$)

Portraits	P_{Low}	P_{High}	DMHL	$BE-AE_{Low}$	$BE-AE_{High}$	$BE-AE_{Mean}$	b_{Low}	b_{High}	b_{Mean}
Anne	20.8	37.5	16.7	8.4	16.2	12.3	67	51	59
Françoise	38.1	51.3	13.2	19.4	18.8	19.1	40	41	41
Jacques	58.5	75.6	17.2	16.1	7.9	12.0	52	69	60
Paul	41.9	61.6	19.6	13.0	14.2	13.6	60	58	59
Raphaël	34.6	49.7	15.1	14.7	20.5	17.6	51	42	47
Mean	38.8	55.1	16.4	14.3	15.5	14.9	53	52	53

rate condition. For example, the estimates given by a participant in the low-base-rate condition can be transformed to determine what his/her responses would have been if he/she had been assigned to the high-base-rate group. Conversely, the same can be done for the other group. For statistical assessment, the means of the Bayesian-derived estimates for each portrait were then compared against the mean probability estimates of the actual group. We call this difference the *Bayesian estimate minus the actual estimate* (BE−AE). To distinguish the BE−AE for the low-base-rate condition from that for the high-base-rate condition, we call the former $BE-AE_{Low}$ and the latter $BE-AE_{High}$. $BE-AE_{Low}$ and $BE-AE_{High}$ are presented in the fourth and fifth columns of Table 1. The sixth column of Table 1, called $BE-AE_{Mean}$, gives the mean of these two differences. If the mean of the differences is null or low, that would indicate normative responding. In contrast, a relatively high mean of the differences would indicate non-Bayesian inference making.

To be consistent with Gigerenzer et al.'s (1988) study, we included $BE-AE_{Low}$. We also included an index developed from one that they dubbed *b*, which corresponds (in our terms) to

$$\frac{\text{DMHL}}{\text{DMHL}+\text{BE}-\text{AE}_{\text{Low}}}. \quad (2)$$

We point out, however, that we did not assume that the two groups (high and low) behaved symmetrically (Baratgin, 1999). Consequently, we called Equation 2 above b_{Low} and constructed a more complete index, which we called b_{Mean}:

$$b_{\text{Mean}} = 100 \times \frac{1}{2}(b_{\text{Low}} + b_{\text{High}})$$

or

$$b_{\text{Mean}} = 100 \times \frac{1}{2}\left[\left(\frac{\text{DMHL}}{\text{DMHL}+\text{BE}-\text{AE}_{\text{Low}}}\right) + \left(\frac{\text{DMHL}}{\text{DMHL}+\text{BE}-\text{AE}_{\text{High}}}\right)\right]. \quad (3)$$

This index varies from 0 (no indication of base-rate utilization) to 100 (perfect Bayesian treatment of base-rate information). The indices b_{Low}, b_{High}, and b_{Mean} are presented in the last three columns of Table 1.

Although it is obvious that the participants generally did not approach these problems in a thoroughly normative manner, we noticed that the participants' estimates approached normative Bayesian-derived probabilities as the cues to complementarity increased. The overall $BE-AE_{Mean}$ started out high at 22.1% in the standard condition and decreased to 17.6% in the suggested-complementarity condition and to 12.4% in the induced-complementarity condition. The same held for $BE-AE_{Low}$, which steadily declined from 21% to 16.5% and, finally, to 12.1% for the standard, suggested-complementarity, and induced-complementarity conditions, respectively. The index b_{Mean} also revealed increasing Bayesian behavior. In the standard, suggested-complementarity, and induced-complementarity conditions, the index b_{Mean} provided outcomes of 24, 39, and 58, respectively.

To see what kinds of errors the participants made, we then took a closer look at the suggested-complementarity condition. Table 3 shows that 11 participants (27.5%) violated the complementarity constraint 34 times overall (55% of error-prone participants were exclusively subadditive, 9% were exclusively superadditive, and the remaining 36% made errors in both directions). This is in line with the findings of Davidson and Hirtle (1990). Our analysis also showed that errors were evenly distributed across the five portraits.

In an effort to gain more insight into the participants' responses, we considered two possible accounts for the errors in the suggested-complementarity condition. Both stem from the idea that error-prone participants did not properly transform their initial estimates by making them complementary (whereas those in the induced condition did). Each of the two proposals reflects a kind of unrealized transformation; perhaps, once one of these transformations was completed, the indices would reveal greater base-rate integration. Thus, we carried out these transformations on noncomplementary estimates in order to determine their effects on the indices.

One potentially unrealized transformation that we investigated was normalization wherein initial estimates of, say, 60% and 20% retained their respective weights and became 75% and 25%. When we applied this trans-

Table 3
Kind and Frequency of Complementarity Violations Per Portrait in the Suggested-Complementarity Condition

	Low Group		High Group		Subtotal		
Portrait	Subadditive	Superadditive	Subadditive	Superadditive	Subadditive	Superadditive	Total
Anne	4	0	3	1	7	1	8
Françoise	3	2	2	0	5	2	7
Jacques	4	1	2	1	6	2	8
Paul	2	2	2	0	4	2	6
Raphaël	1	1	3	0	4	1	5
Total	14	6	12	2	26	8	34

Note—Errors were committed by 11 participants.

formation we found that the conditions' indices of base-rate integration were not dramatically affected: DMHL = 12, $BE-AE_{Mean} = 15.6$, and $b_{Mean} = 41$. Also, the reported statistical difference between the suggested- and induced-complementarity conditions remained significant. The second potentially unrealized transformation was "splitting the difference," wherein the initial estimates could be converted additively and equally so that the difference from 100% could be distributed equally (in this case, 60% and 20% became 70% and 30%, respectively). When errors were transformed by hand in this manner, the conditions' indices of base-rate integration were again not noticeably different: DMHL = 11, $BE-AE_{Mean} = 19$, and $b_{Mean} = 37.9$. Thus, we concluded that a failure to carry out a singular kind of transformation after arriving at initial estimates did not affect the base-rate integration indices in the suggested-complementarity condition. The significance of these results will be taken up in the Discussion section.

Finally, we analyzed the participants' confidence measures, which led to three findings. First, we found no difference in confidence judgments with respect to the high and low conditions. This is reassuring, because it showed that the relative weight of the base-rate information plays no role in participants' judgments. Second, we found that confidence grew systematically as the cues to complementarity increased. The participants reported higher confidence as the task's complementarity aspect became more compelling: The standard, suggested-complementarity, and induced-complementarity conditions yielded mean confidence ratings of 43.2%, 53.4%, and 61.5%, respectively. Third, the highly diagnosing portraits of Anne and Jacques tended to prompt higher confidence estimates (56.3% and 54.8%, respectively) than did the intermediate ones of Paul and Raphaël (52.4% and 52.2%, respectively). The lowest level of confidence was associated with the portrait of Françoise (47.6%).

Discussion

Evidently, normative responding was facilitated when a task pointed to the two alternatives, and was further facilitated when it compelled the participants to respect the complementarity constraint. We point to our three main findings. First, the participants were more likely to employ base rates in both the suggested- and induced-complementarity conditions than in the standard condition. This difference was evident for each of the five portraits.

Second, there was a difference between the induced- and suggested-complementarity conditions. This is especially important, because it revealed a facilitative role for the cue when it was more strictly related to complementarity. There is no conceivable way to induce complementarity without first requesting two estimates, as was done in the suggested-complementarity condition. Thus, we supposed that the additional cue in the induced-complementarity condition (to make sure that the two estimates added up to 100%) served to promote probabilistic reasoning and, in turn, the integration of base rates. Note that this was accomplished without cues to the base rates. An alternative hypothesis would be that the difference between the participants in the suggested- and induced-complementarity conditions was an artifact due to the participants' compliance with the instructions. That is, a participant in the induced-complementarity condition, after first generating intuitive estimates that were not complementary, applies a transformation that made the estimates sum to 100%. According to this hypothesis, it would be the transformation that accounted for the changes in the indices and not the cue. Our analyses of errors ruled out this alternative. When the complementarity violations of the suggested-complementarity condition were normalized or else adjusted by "splitting the difference," indices in that condition remained low relative to those in the induced-complementarity condition.

Third, confidence level increased as complementarity was emphasized. This indicated that an appreciation of probabilistic aspects of the task was important to the participants' confidence levels. There may be another interpretation for the increasing confidence measures across task conditions: Higher confidence measures may have been a by-product of increasing task demands. When more effort was required, the participants had more confidence in their answers. Nevertheless, it is clear that the participants were increasingly more successful and confident as they were encouraged to integrate probabilistic information.

EXPERIMENT 2

In Experiment 2, we examined whether Gigerenzer et al.'s (1988) Urn problem, a version of the task that highlights its probabilistic features and typically demonstrates integration of base-rate information, would provide cues to probabilistic reasoning that engaged respect for complementarity. If our hypothesis that normative performance on the task is linked to probabilistic reasoning viewed more generally is correct we should find that Gigerenzer et al.'s (1988) experimental paradigm leads to low rates of complementarity violation. However, if our indices show base-rate integration while producing high rates of complementarity violations, it would indicate that complementarity and base-rate integration are independent processes.

In Experiment 2, we presented Gigerenzer et al.'s (1988) Urn task and assigned subjects to the high- and low-base-rate conditions with instructions to determine estimates with respect to each of the two professional groups. The two estimates served as dependent measures of respect for complementarity.

Method

Participants. Forty students from the same population of students as those in Experiment 1 took part in this study.

Procedure and Materials. Each participant was individually interviewed. He or she sat facing the interviewer, who was seated at

a desk. The first four sentences of the instructions were identical to those used in the suggested-complementarity condition of Experiment 1. The rest is presented in the excerpt below, which came from the low-base-rate condition:

> Ten (10) of these portraits are presented in front of you. Among them you will find three portraits of math teachers indicated with an M and the seven remaining ones are portraits of teachers of French literature; they are marked with an F. In the five urns in front of you there are copies of the same ten portraits. You are asked to draw at random one portrait from each urn. For each portrait, please indicate the chances (in percent) that the person described is one of the 30 math teachers and the chances (in percent) that the person described is one of the 70 teachers of French literature. You are asked to give your answer on a scale from 0% to 100%.

The participants gave their two estimates for each portrait after each draw. The order of presentation of the urns was systematically shuffled to prevent a spurious order effect. We ensured that we had 10 copies of the same portrait in each urn, but we let the participants believe they were drawing at random. The participants were entitled to ask any questions until they began drawing from the first urn. After that, questions were forbidden. The high-base-rate condition, of course, inverted the base rates provided in the above instruction.

The participants were randomly assigned to one of the two conditions (low and high base rates of math teachers). The task modeled the instructions after those in the suggested-complementarity condition of Experiment 1, but implanted the random draw described above, which was based on Gigerenzer et al.'s (1988) study.

Results

We will refer to the present results as outcomes of the Urn condition. We analyzed the participants' responses in a manner similar to the one used in Experiment 1. The mean difference (DMHL) between the high and low groups across the five portraits in the suggested-complementarity Urn condition was well above zero (DMHL = 14.7%), indicating that the participants largely processed base rates. Other indices also point to Bayesian responding ($BE-AE_{Mean}$ = 12.87% and b_{Mean} = 53). These findings largely confirmed those of Gigerenzer et al. (1988). The highly diagnosing portraits of Anne and Jacques again yielded results closest to Bayesian estimates. The results are presented in Table 4.

Only 1 out of the 40 participants violated the complementarity constraint in the Urn condition. This participant violated the constraint on all five estimates, belonged to the high-base-rate group, and erred in the direction of subadditivity. Once the results for this participant were removed from the analysis, DMHL increased to 15.77% and $BE-AE_{Mean}$ decreased to 11.88 (b_{Mean} = 57). Table 4 shows how this 1 participant affected the indices of base-rate integration, albeit minimally. Interestingly, 8 participants asked the interviewer during the experiment whether the sums of the probabilities were required to be 100%; however, their questions were not answered.

Although statistical comparisons between the Urn task, on the one hand, and the suggested- and induced-complementarity versions of Experiment 1, on the other, are not advisable because of procedural differences, a preliminary comparison revealed some noteworthy findings. With respect to DMHL and complementarity, the participants' responses in the Urn condition were comparable to those in the induced-complementarity condition of Experiment 1 and were more normative than those in the suggested-complementarity condition. The same held true for their confidence ratings (mean of 60.22% in the present experiment).

Discussion

It was apparent that Gigerenzer et al.'s (1988) Urn problem increased not only the integration of base-rate information but participants' respect for the complementarity constraint as well. Thus, our hypothesis was supported: As the task appeared to be more probabilistic in

Table 4
Means of Participants' Estimates (Chance in Percent That the Portrait Is That of a Math Teacher) and Scores Based on Bayesian-Derived Estimates (Last Six Columns), Experiment 2 (N = 40)

Portraits	P_{Low}	P_{High}	DMHL	$BE-AE_{Low}$	$BE-AE_{High}$	$BE-AE_{Mean}$	b_{Low}	b_{High}	b_{Mean}
					Urn Condition				
Anne	19.5	35	15.5	7.4	11.7	9.5	68	57	62
Françoise	39	49.5	10.5	13.8	19.4	16.6	43	35	39
Jacques	58.3	74.8	16.5	12.5	8.2	10.3	57	67	62
Paul	44	60.3	16.3	16.3	11.9	14.1	50	58	54
Raphaël	37	51.8	14.8	11.1	16.7	13.9	57	47	52
Mean	39.6	54.3	14.7	12.2	13.6	12.9	55	52	53
					Urn Condition (Refined)				
Anne	19.5	35.8	16.3	7	10.9	8.9	70	60	65
Françoise	39	50	11	13	18.9	15.9	46	37	41
Jacques	58.3	75.8	17.5	11	7.1	9.1	61	71	66
Paul	44	62.9	18.9	14.9	9.3	12.1	56	67	62
Raphaël	37	52.1	15.1	10.4	16.4	13.4	59	48	54
Mean	39.6	55.3	15.8	11.3	12.5	11.9	58	56	57

Note—All instructions requested complementary estimates. The refined summary removes the single participant who violated the complementarity constraint. For other explanations, see Table 1.

nature, consideration of relevant probabilistic principles increased. When one counteracts the neglect of base-rate information on the Engineer–Lawyer problem, one encourages respect for complementarity as well.

Although our results from the Urn condition were largely similar to those found in Gigerenzer et al. (1988), the participants in their study gave estimates that were, on the whole, closer to those predicted by the Bayesian norm. For example, b_{Low} in Experiment 2 of the present study provided a mean of 55, whereas the equivalent index in their study produced a mean of 65. This variation may have stemmed from two factors. One factor may have been related to the portraits themselves. Gigerenzer et al. (1988) portrayed engineers and lawyers, whereas we portrayed two kinds of teachers. The other factor may have been that Gigerenzer et al. (1988) presented six portraits and we presented five. With six portraits, base rates could have been applied more intuitively; that is, one could have concluded that, with six portraits and a distribution of 70% engineers and 30% lawyers, a distribution of four engineers and two lawyers would have been likely. With five portraits (and the same 70/30 distribution) two configurations are plausible—three engineers and two lawyers or four engineers and one lawyer. Thus, six portraits may have increased Bayesian responding by allowing participants who appreciated base-rate information to constrain the problem (Baratgin, 1999).

How can one explain the increased respect for complementarity in the Urn problem? Some might argue that a "frequentist" presentation of the contents of the urn (showing that seven elements came from one profession and that three came from the other) aided the participants in estimating probabilities normatively (see Cosmides & Tooby, 1996; Gigerenzer, 1993; Gigerenzer & Hoffrage, 1995; Jones, Jones, & Frisch, 1995). We agree that the Urn version prompted respect for complementarity, but not necessarily for the reasons offered by frequentists. We see two limitations to the frequentist proposal. First, a frequentist presentation appears sufficient but not necessary for increased normative responding. The induced-complementarity condition in Experiment 1 prompted estimates indicating base-rate integration at levels comparable to those found on the Urn problem without a frequency format. Second, the presentation of random sampling does not explain what mechanism is triggered (see Griffin & Dukeshire, 1992, for a critique of Gigerenzer et al.'s, 1988, argument). It is proposed here that the random sampling in the Urn condition drew attention to base rates and that this increased base rate use as well as complementarity compliance.

EXPERIMENT 3

The presentation of the Urn condition in Experiment 2 encouraged the participants both to integrate base rates and to respect complementarity, but we did not directly present a control condition. If one were to treat the suggested-complementarity condition of Experiment 1 as a control (because it offered the same task instructions but lacked the Urn presentation), results would indicate that the Urn context provided strong cues to normative responding (see Tables 1 and 4). In Experiment 3, we determined whether this observation was replicable in one overarching procedure that licensed empirical claims.

In fact, two features of the Urn condition presented in Experiment 2 distinguished it from the suggested-complementarity condition presented in Experiment 1. One was the presence or absence of the random-sampling (Urn) format; the other was the experimenter–participant interaction (the suggested-complementarity task of Experiment 1 was presented to participants as a group and the Urn task of Experiment 2 individually). Thus, in Experiment 3, we investigated the importance of these two task features as we compared performance on the Urn task and on the suggested-complementarity task. For the present experiment, we called the suggested-complementarity condition the no-Urn condition. The participants in all four conditions were presented the instructions requesting estimates with respect to both groups of teachers. We anticipated that the presence of the Urn alone would be critical to increasing normative responding as measured by base-rate integration and by rates of complementarity violation.

Method

Participants. One hundred twenty-eight students, who came from the same population of students as did those in the earlier experiments, took part.

Procedure and Materials. The only novel tasks (with respect to conditions that have not been described earlier) were (1) the suggested-complementarity (*no-Urn*) version presented *individually* and (2) the *collective* presentation of the *Urn* condition. The individual-no-Urn task simply presented each participant with the problem set of the suggested-complementarity condition (of Experiment 1) individually in the presence of the experimenter. The collective-Urn problem presented the materials used in Experiment 2 to groups of 8 participants; 1 member of each group publicly drew out the portraits from the urn. As in Experiment 2, efforts were made to give the impression that the draws from the five Urns were random. However, each urn contained 10 identical portraits; 10 piles of portraits were then prepared and displayed so that, once the experimenter publicly ascertained which portrait was drawn, he could distribute the copies to each participant containing the same task instructions as those in the individual condition. (The collective-no-Urn task was essentially a replication of the suggested-complementarity condition of Experiment 1 and the individual-Urn task was essentially a replication of Experiment 2.) Given the limited number of groups, one of two random orders of presentation was used throughout this experiment.

Design. As in the earlier experiments, the participants were randomly assigned to either low or high base rates of math teachers. The design was identical to that of the earlier experiments, except that two variables were manipulated: the *presence* or *absence* of the Urn format and the *collective* or *individual* manner of experimental presentation. Thus, this was a 2 (high vs. low base rate of mathematics teachers) × 2 (Urn vs. no-Urn) × 2 (individual vs. collec-

Table 5
Means of Participants' Estimates (Chance in Percent That the Portrait Is That of a Math Teacher) and Scores Based on Bayesian-Derived Estimates (Last Six Columns) in Experiment 3 ($N = 128$)

Portraits	P_{Low}	P_{High}	DMHL	BE−AE$_{Low}$	BE−AE$_{High}$	BE−AE$_{Mean}$	b_{Low}	b_{High}	b_{Mean}
Urn-Collective Condition									
Anne	19.1	37.8	18.8	6.8	14.3	10.5	73	57	65
Françoise	35.6	52.5	16.9	15.3	20.2	17.7	53	46	49
Jacques	57.5	76.3	18.8	14.3	9.5	11.9	57	66	62
Paul	40.9	60.3	19.4	15.3	14.9	15.1	56	56	56
Raphaël	35.6	53.8	18.1	13.9	17.9	15.9	57	50	54
Mean	37.8	56.1	18.4	13.1	15.4	14.2	58	54	56
Urn-Individual Condition									
Anne	20.0	36.9	16.9	7.4	12.5	10.0	69	57	63
Françoise	36.3	50.6	14.4	13.4	15.1	14.2	52	49	50
Jacques	58.8	76.9	18.1	13.0	6.0	9.5	58	75	67
Paul	43.8	61.6	17.8	12.6	9.1	10.8	59	66	62
Raphaël	37.5	52.8	15.3	13.0	19.4	16.2	54	44	49
Mean	39.3	55.8	16.5	11.9	12.4	12.2	58	57	58
No-Urn-Collective Condition									
Anne	19.8	33.3	13.6	9.0	19.2	14.1	60	41	51
Françoise	36.9	46.1	9.1	21.7	26.5	24.1	30	26	28
Jacques	59.7	74.1	14.4	22.4	11.2	16.8	39	56	48
Paul	48.3	57.8	9.6	26.4	22.5	24.5	27	30	28
Raphaël	40.4	53.8	13.4	20.5	21.4	20.9	39	39	39
Mean	41.0	53.0	12.0	20.0	20.1	20.1	37	37	37
No-Urn-Individual Condition									
Anne	20.3	31.3	11.0	9.4	19.8	14.6	54	36	45
Françoise	37.7	45.3	7.6	21.6	23.2	22.4	26	25	25
Jacques	61.1	73.2	12.1	15.8	7.6	11.7	43	61	52
Paul	49.3	59.2	9.9	22.3	19.8	21.0	31	33	32
Raphaël	36.8	49.2	12.4	13.1	21.2	17.2	49	37	43
Mean	41.0	51.6	10.6	16.4	18.3	17.4	39	37	38

Note—The Urn-individual condition is equivalent to the Urn condition of Experiment 2. The no-Urn-collective condition is equivalent to the suggested-complementarity condition of Experiment 1. All conditions requested two estimates. For other explanations, see Table 1.

tive presentation) between-subjects design. There were 16 participants in each condition.

Results and Discussion

We analyzed the participants' responses with respect to the Urn and no-Urn problems as presented individually and collectively. Not surprisingly, the mean difference (DMHL) between the high and low groups across the five portraits in each of the four conditions was well above zero. The other indices also pointed to Bayesian responding. These results, presented in Table 5, largely confirmed those found in the two earlier experiments.

A 2 (format: Urn vs. no-Urn) × 2 (task administration: collective vs. individual) between-subjects ANOVA was carried out with the DMHL serving as a dependent measure. The ANOVA showed a main effect of format [$F(1,16) = 56.365, p < .001$], a marginal, nonsignificant effect of task administration [$F(1,16) = 4.12, p = .06$], and no significant interaction ($p = .78$). Apparently, the group presentation slightly (and nonsignificantly) facilitated normative performance. Similar analyses with other dependent measures (BE−AE$_{Mean}$ and b_{Mean}) confirmed a main effect of format, and a nonsignificant effect of task administration, across the experiment. Thus, we can confidently conclude that the Urn context significantly facilitated normative responding.

The rates of complementarity violations were of particular interest. There were no violations of complementarity in the two Urn conditions, and 15 participants (23%) committed such violations across the two no-Urn conditions. The latter is comparable to the rate of such errors in the suggested-complementarity condition of Experiment 1 (and to the rate found by Davidson & Hirtle, 1990). The 8 participants who committed such errors were in the collective-condition group and 7 were in the individual-condition group. As before, errors were distributed across the portraits. There were no such errors in the two Urn conditions.

In sum, Experiment 3 confirmed the findings of the two earlier experiments. The Urn condition prompted high rates of base-rate integration and no violations of complementarity. In contrast, the (control) no-Urn condition (which was equivalent to the suggested-complementarity condition of Experiment 1) was less likely to prompt

base-rate integration or respect for complementarity. The task administration did not lead to robust effects. Although collective administration tended to influence normative reasoning (according to DMHL), this factor did not affect the rate of complementarity errors nor did it affect the other, finer measures of base-rate integration.

These findings support two points made earlier. On the one hand, all four of these conditions requested two estimates which, according to our hypothesis, prompted some degree of normative reasoning in comparison with that required by the standard problem. On the other hand, we did not find thoroughly normative responding in any of the conditions of the present experiments. Even in those problems that strongly underlined probabilistic considerations, the participants appeared to aim for an integration of the portraits with probabilistic information.

GENERAL DISCUSSION

The literature offers two kinds of explanations concerning the neglect to respect complementarity. One explanation is that participants provide estimates of a given hypothesis independently of the alternative hypothesis; that is, participants adopt a "non-distribution conception of probability" (Teigen, 1983; see also Van Wallendael & Hastie, 1990). This suggests that reasoners lack the competence to consider the fundamental probabilistic principle. Another class of explanations can be drawn from work done on Bayesian reasoning. Many researchers (e.g., Braine, Connell, Freitag, & O'Brien, 1990; Macchi, 1994, 1995) have proposed that participants, in considering base rates, become confused because they invert the given conditional probability and thus make the structure of the problem ambiguous; that is, the estimate $P(\text{not-H} \mid \text{D})$ is mistaken for $P(\text{D} \mid \text{not-H})$. It is plausible to suppose that this kind of error occurs when complementary estimates are requested as well. This kind of explanation suggests that difficulties arise while one is performing the task but that competence is still available.

We assume that competence is available and have argued that facilitative Engineer–Lawyer problems prompt higher rates of correct responses because they draw the participants' attention to sources relevant to normative responding. Whereas a standard version of the Engineer–Laywer problem showed how the participants may not have mechanically drawn out probabilistic information, they were clearly capable of applying normative solutions once such information was made relevant to them. More effort was applied to probabilistic information as it became more relevant. Cuing participants to complementarity (as in Experiment 1) was one way to point to the importance of probabilistic reasoning, and the Urn problem used in Experiments 2 and 3 was another. Our proposal, which was motivated by relevance theory (Sperber & Wilson, 1986/1995), described why the neglect to respect complementarity is linked to base-rate neglect: Participants' *general* failure to accord effort to the probabilistic information in the standard Engineer–Lawyer problem led to low levels of normative performance.

Although base rates became increasingly important for estimates concerning all the portraits as conditions increased the cues to complementarity, the highly diagnosable portraits of Anne and Jacques were consistently the most likely to prompt differences between high- and low-base-rate groups and to prompt estimates revealing normative behavior. This leads one to doubt the claim that participants rely exclusively on the *representativeness* heuristic, because these informative (stereotyped) portraits provoked the highest rates of normative responses. It appeared then that the participants sought out information that was coherent and relevant, particularly with respect to two factors: (1) the clarity of information in the personality portraits and (2) the probabilistic features of the problem.

This work adds to mounting evidence that shows that reasoning, viewed from the point of view of discourse, can prompt normative responses under conditions that make certain aspects of a task relevant (see Noveck, 1997; Sperber, Cara, & Girotto, 1995). Work done with respect to tasks developed by Kahneman and Tversky has been especially illustrative in this regard. Normative performance on the Linda problem increased when its conversational peculiarities were avoided (Hertwig & Gigerenzer, 1999; Politzer & Noveck, 1991) and normative performance on the Taxi-cab problem increased with the useful presentation of information and carefully worded questions (Macchi, 1995). If one makes clear to participants the speaker's (the experimenter's) communicative intent, normative performance is more likely because the participants have a clearer idea of what the experimenter wants, and participants, being relatively good listeners, try to oblige to the best of their abilities. Viewed in this light, one can see that the rate of normative responding varies as a function of the communicative intent in a problem's protocol.

To summarize, we argue that the neglect to respect complementarity is not only common on the standard Engineer–Lawyer problem but closely linked to the well-known base-rate neglect. We were thus motivated to demonstrate that the two kinds of neglect rise and fall together. We showed that two levels of cue to complementarity prompted two corresponding levels of base-rate integration and that the Urn task—a version known for facilitating base-rate integration—prompted the participants to respect complementarity. Given the nature of the facilitative problems, we take our findings to mean that normative performance on tasks such as these is more likely when a participant is invited to perceive a problem probabilistically.

REFERENCES

Alberoni, F. (1962). Contribution to the study of subjective probability: Part I. *Journal of General Psychology*, **66**, 241-264.

Asare, S. K., & Wright, A. (1995). Normative and substantive exper-

tise in multiple hypotheses evaluation. *Organizational Behavior & Human Decision Processes*, **3**, 171-184.

Baratgin, J. (1999). *Psychologie du raisonnement probabiliste dynamique: Recherches Historiques, Pragmatiques et Expérimentales sur la règle de Bayes* [Psychology of dynamic probabilistic reasoning: Historical, pragmatic, and experimental aspects of Bayes' rule]. Doctoral dissertation, École Polytechnique.

Braine, M. D. S., Connell, J., Freitag, J., & O'Brien, D. P. (1990). Is the base rate fallacy an instance of asserting the consequent? In K. L. Gilhooly, M. T. G. Leane, R. H. Logie, & G. Erdos (Eds.), *Lines of thinking: Reflections on the psychology of thought* (Vol. 1, pp. 165-180). New York: Wiley.

Carroll, G. S., & Siegler, R. S. (1977). Strategies for the use of base rate information. *Organizational Behavior & Human Performance*, **19**, 392-402.

Cohen, J., Dearnaley, E. J., & Hansel, C. E. (1956). The addition of subjective probabilities. *Acta Psychologica*, **12**, 371-380.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgement under uncertainty. *Cognition*, **58**, 1-73.

Davidson, D., & Hirtle, S. C. (1990). Effects of nondiscrepant and discrepant information on the use of base rates. *American Journal of Psychology*, **103**, 343-357.

Einhorn, H. J., & Hogarth, R. M. (1985). Ambiguity and uncertainty in probabilistic inference. *Psychological Review*, **92**, 433-461.

Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. *Memory & Cognition*, **12**, 402-410.

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

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality* (pp. 284-313). London: Routledge.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 129-162). New York: Wiley.

Gigerenzer, G. (1996). On narrow norm and vague heuristics: A reply to Kahneman and Tversky. *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 & 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., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Kruger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge: Cambridge University Press.

Ginosar, 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.

Girotto, V. (1994). *Il Ragiomento*. Bologna: Il Mulino.

Griffin, D., & Dukeshire, S. (1992). *On the use and neglect of base rates: A test of the visual random sampling hypothesis*. Unpublished manuscript, University of Waterloo, Department of Psychology.

Hertwig, R., & Gigerenzer, G. (1999). The "conjunction fallacy" revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, **12**, 275-305.

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

Jones, S. K., Jones, K. T., & Frisch, D. (1995). Biases of probability assessment: A comparison of frequency and single case judgments. *Organizational Behavior & Decision Processes*, **2**, 109-122.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, **3**, 430-453.

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

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral & 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 & Social Psychology*, **59**, 1140-1152.

Macchi, L. (1994). On the communication and comprehension of probabilistic information. *Psycoloquy* [On-line serial], **5**(11), base-rate. 11.macchi.

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

Marks, D. F., & Clarkson, J. K. (1972). An explanation of conservatism in the bookbag-and-pokerchips situations. *Acta Psychologica*, **36**, 145-160.

Noveck, I. A. (1997). Deductive competence need not be problematic: Some comments on Evan and Over's proposal concerning human rationality. *Current Psychology of Cognition*, **17**, 162-172.

Piatelli-Palmarini, M. (1993). *L'illusione di sapere*. Milan: Arnoldo Mondadori.

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

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**, 67-83.

Slovic, P., Lichtenstein, S., & Fischhoff, B. (1988). Decision making. In R. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens's Handbook of experimental psychology, learning and cognition* (Vol. 2, pp. 673-737). Cambridge: Cambridge University Press.

Sperber, D., Cara, F., & Girotto, V. (1995). Relevance theory explains the selection task. *Cognition*, **57**, 31-95.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication & cognition* (2nd ed.). Oxford: Blackwell. (1st ed. Cambridge, MA: Harvard University Press, 1986)

Swieringa, R., Gibbins, M., Larsson, L., & Sweeney, J. L. (1976). Experiments in the heuristics of human information processing. *Journal of Accounting Research*, **14** (Suppl.), 159-187.

Teigen, K. H. (1974). Overestimation of subjective probabilities. *Scandinavian Journal of Psychology*, **15**, 56-62.

Teigen, K. H. (1983). Studies in subjective probability III: The unimportance of alternatives. *Scandinavian Journal of Psychology*, **24**, 97-105.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 153-160). Cambridge: Cambridge University Press.

Van Wallendael, L. R., & Hastie, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & Cognition*, **18**, 240-250.

Von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. Cambridge: Cambridge University Press.

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

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

NOTE

1. Bayes formula (1) reads as follows:

$$P(\mathrm{H}\,|\,\mathrm{D}) = \frac{P(\mathrm{D}\,|\,\mathrm{H})P(\mathrm{H})}{P(\mathrm{D})},$$

where

$$P(\mathrm{D}) = P(\mathrm{D}\,|\,\mathrm{H})P(\mathrm{H}) + P(\mathrm{D}\,|\,\text{not-H}) + P(\mathrm{D}\,|\,\text{not-H})P(\text{not-H}) \neq 0.$$

In the Engineer–Lawyer problem, the two alternative hypotheses are H = E(Engineers) or Not-H = L(Lawyers). From a normative point of view, Bayes's rule (Equation 1) gives a relation between the probabilities $P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{Low}}$ and $P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{High}}$, with *Low* and *High* representing the

two kinds of base rates. In terms of ratio relations, Equation 1 yields the following: Suppose that $P(\mathrm{E})_{\mathrm{Low}} = P(\mathrm{L})_{\mathrm{High}} = P(\mathrm{E})$; then

$$\frac{P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{Low}}}{P(\mathrm{L}\,|\,\mathrm{D})_{\mathrm{Low}}} = \frac{P(\mathrm{D}\,|\,\mathrm{E})_{\mathrm{Low}}P(\mathrm{E})}{P(\mathrm{D}\,|\,\mathrm{L})_{\mathrm{Low}}P(\mathrm{L})},$$

and

$$\frac{P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{High}}}{P(\mathrm{L}\,|\,\mathrm{D})_{\mathrm{High}}} = \frac{P(\mathrm{D}\,|\,\mathrm{E})_{\mathrm{High}}P(\mathrm{L})}{P(\mathrm{D}\,|\,\mathrm{L})_{\mathrm{High}}P(\mathrm{E})}.$$

The likelihood principal ensures equality between likelihood ratios:

$$\frac{P(\mathrm{D}\,|\,\mathrm{E})_{\mathrm{Low}}}{P(\mathrm{D}\,|\,\mathrm{notE})_{\mathrm{Low}}} = \frac{P(\mathrm{D}\,|\,\mathrm{E})_{\mathrm{High}}}{P(\mathrm{D}\,|\,\mathrm{notE})_{\mathrm{High}}} = \text{Constant}.$$

Using the complementarity constraint $P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{Low}}+P(\mathrm{L}\,|\,\mathrm{D})_{\mathrm{Low}} = P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{High}}+P(\mathrm{L}\,|\,\mathrm{D})_{\mathrm{High}} = 1$, and dividing the former ratios, one gets

$$P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{High}} = \frac{O^2P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{Low}}}{1+P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{Low}}\,(O^2-1)}, \text{ where } O = \frac{P(\mathrm{E})}{P(\mathrm{L})}. \quad (1\mathrm{A})$$

When the complementarity constraint is not complied with one gets

$$P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{Low}}+P(\mathrm{L}\,|\,\mathrm{D})_{\mathrm{Low}} = 1+\alpha$$

and

$$P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{High}}+P(\mathrm{L}\,|\,\mathrm{D})_{\mathrm{High}} = 1+\beta,$$

and, hence, the following equation:

$$P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{High}} = \frac{O^2P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{Low}}(1+\beta)}{1+\alpha+P(\mathrm{E}\,|\,\mathrm{D})_{\mathrm{Low}}(O^2-1)}.$$

Equations 1A and 1B do not yield the same curves.

(Manuscript received July 8, 1997;
revision accepted for publication October 22, 1998.)