ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

COGNITION

Cognition xxx (2006) xxx–xxx

www.elsevier.com/locate/COGNIT

Waiting for the bus: When base-rates refuse to be neglected ☆

Karl Halvor Teigen [a,*], Gideon Keren [b]

[a] Department of Psychology, University of Oslo, P.O. Box 1094, Blindern, NO-0492 Oslo, Norway

[b] Faculty of Technology Management, Eindhoven University of Technology, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Received 8 September 2005; revised 24 March 2006; accepted 31 March 2006

Abstract

The paper reports the results from 16 versions of a simple probability estimation task, where probability estimates derived from base-rate information have to be modified by case knowledge. In the bus problem [adapted from Falk, R., Lipson, A., & Konold, C. (1994). The ups and downs of the hope function in a fruitless search. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 353–377). Chichester, UK: Wiley], a passenger waits for a bus that departs before schedule in 10% of the cases, and is more than 10 min delayed in another 10%. What are Fred's chances of catching the bus on a day when he arrives on time and waits for 10 min? Most respondents think his probability is 10%, or 90%, instead of 50%, which is the correct answer. The experiments demonstrate the difficulties people have in replacing the original three-category 1/8/1 partitioning with a normalized, binary partitioning, where the middle category is discarded. In contrast with typical studies of "base-rate neglect", or under-weighing of base-rates, this task demonstrates a reversed base-rate fallacy, where frequentistic information is overextended and case information ignored. Possible explanations for this robust phenomenon are briefly discussed.

☆ This research was founded by internal grants from the University of Oslo to K.H.T., and from the Eindhoven University of Technology to G.K. Thanks are due to Janneke P.C. Staaks, Eindhoven, and Sverre K.L. Eldøy, Oslo, for valuable contributions.

* Corresponding author. Tel.: +47 22 84 51 87; fax: +47 22 84 51 75.
E-mail address: k.h.teigen@psykologi.uio.no (K.H. Teigen).

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2006.03.007

Keywords: Base-rates; Probability judgments; Frequentistic probabilities; Partitioning; Probability problem

1. Introduction

Research on subjective probabilities over the last 40 years has uncovered a number of problems people encounter when asked to produce probability estimates, both in response to laboratory tasks and in real life (for reviews, see Baron, 2002; Gilovich, Griffin, & Kahneman, 2002; Kahneman & Tversky, 1982; Pohl, 2004; Wright & Ayton, 1994). In many cases, estimates are consistently too high or too low compared to normative standards. Such systematic deviations are commonly called biases.

One type of biased probability assessments has been demonstrated in situations where individual case information is available, which might, at least in principle, make a particular outcome more or less likely than indicated by the frequencies associated with its category membership. In extreme cases, this can lead to an almost complete "base-rate neglect". Kahneman and Tversky (1973) found that when people were asked to estimate the probability that a person is an engineer or a lawyer, they relied primarily on characteristics of the individual descriptions and seemed to ignore the relative frequencies of lawyers and engineers in the population from which the case was drawn. Other studies show that base-rates are some times considered relevant and taken into account (for an overview, see Koehler, 1996), yet are usually given less weight than dictated by normative, Bayesian, considerations.

In situations where individualizing information is not available, or where the statistical nature of the task is emphasized, people may rely more strongly upon extensional information (Nisbett, Krantz, Jepson, & Kunda, 1983). For instance, teachers usually explain basic principles of probability calculus by chance devices like coin tosses, draws from an urn, spinners, and lotteries, indicating that an intuitive grasp of probabilities in these areas is within the reach of most students. In general, studies of how people understand and use base-rates range from demonstrations of almost complete base-rate neglect, to instances where base-rates are successfully incorporated in the assessment of conditional probabilities (Koehler, 1996).

The present study differs from previous base-rate research by presenting a simple probability estimation task, where people stick to the base-rates instead of modifying them in the light of individual case information. Participants in the present studies seem to (completely) ignore the case information and base their probability estimates solely on statistical information, leading to a reversed "base-rate fallacy". Paraphrasing Keren and Lewis (1994), who described two opposite beliefs about random sequences as gambler's fallacy of Type I and Type II, we may contrast the traditional Base-rate fallacy Type I, where people under-weigh or ignore base-rate information, with a Base-rate fallacy Type II, where base-rates are overweighed, and case information is ignored.

> Consider the following situation:
> Fred goes to work by a bus that departs only once every hour. Fred has observed that the bus arrives before schedule in about 10% of the cases, 0–10 min after schedule in 80% of the cases, and is more than 10 min late in 10% of the cases.
> One day Fred arrives at the bus stop exactly on time (T_1). What is the probability that he will catch the bus?
> Another day he arrives 10 min late (T_2). What is the probability that he will catch the bus?

Most people will easily calculate from the base-rates that Fred has a 90% chance on the first day, and only a 10% chance on the second day. But imagine that on the third day, Fred arrives exactly on time, and waits for 10 min without seeing any bus. What is the probability that he will still catch the bus?

When this question was posed to a group of psychology students at the University of Oslo, they got into trouble. Some claimed that he would still have a 90% chance, since he had arrived on time. Others insisted that he has only a 10% chance, since the bus is rarely more than 10 min late. Who were correct?

Neither party got it right. Fred's chances of catching the bus after 10 min waiting time are lower than they were at T_1, but not so low as they would have been if he had arrived at T_2. Actually, as can easily be shown, he has a fifty/fifty chance.

Readers with a penchant for Bayesian statistics will recognise the situation as one where an initial hypothesis has to be modified by evidence. We may distinguish between the two complementary hypotheses H1 (that the bus has already departed) and H2 (that the bus will still arrive). According to Bayes theorem, the relative strength of these two hypotheses after waiting (the posterior odds) can be expressed as the prior odds multiplied by the likelihood ratio $[p(\mathrm{H2}|\mathrm{data})/p(\mathrm{H1}|\mathrm{data}) = p(\mathrm{H2})/p(\mathrm{H1}) \times p(\mathrm{data}|\mathrm{H2})/p(\mathrm{data}|\mathrm{H1})]$. At T_1 the prior probability for H2 is 9/10, whereas the complementary prior probability for H1 is 1/10, yielding prior odds for H2:H1 = 9:1. Fred has subsequently collected evidence by waiting and observing that the bus did in fact not show up in the 0–10 min interval. This observation is unlikely under H2, $p = 1/9$, but extremely likely under H1, $p = 1.0$, yielding a likelihood ratio of the conditional probabilities of 1:9. When the prior odds and the likelihood ratio are multiplied we get 9:9, which means that both hypotheses are equally likely. Thus, they have both a 50% chance of being true.

Students should be excused for not being intuitive Bayesians, as it is debatable how well this formula captures the thinking process in humans (for a review, see Birnbaum, 2004). There is, however, another, much simpler way of solving the problem: After Fred has waited for 10 min, he can eliminate the 80% chance of a bus in the 0–10 min period, because he *knows* it did not arrive in this interval. Only two possibilities remain: Either the bus arrived ahead of schedule, or it will arrive more than 10 min late. Both outcomes are rare, but being equally rare and forming an exhaustive set, p(he has missed the bus) = p(he will catch the bus) = .5. The situation is formally similar to an urn problem, where we start with 10 white, 80 blue, and 10 red chips. Originally there is a 90% chance of drawing a colored (blue or red) chip, but after all blue chips have been removed, the chance of drawing a colored chip is reduced to 50% (whereas the chance of drawing a white chip has increased from 10% to 50%).

Variants of the bus problem have previously been studied by Falk, Lipson, and Konold (1994), who claim that this problem is representative of a large class of waiting and search situations that most of us are familiar with from daily life. Following these authors, probabilities in such situations can be described as more or less intense hopes. It is possible to distinguish between the long-term hope (that the bus will finally arrive) and the short-term hope (that the bus will arrive within the next few minutes). One can accordingly observe two competing hope functions: As we wait, short-term hope will rise, whereas long-term hope will tend to decline. Falk et al. observed that the general shape of both these functions were in line with the normative model, whereas the specific probability estimates were rarely correct. In these studies, probability estimates during waiting were compared to values derived from Bayesian calculations, but they also introduced a simplified version that could be solved by the simpler strategy described above. This led to a substantial increase in the number of correct solutions for one of their problems, whereas the effects on the bus problem were not tested. Falk et al. observed than many subjects made ''conservative'' estimates that seemed to be determined chiefly by the initial probabilities, and not duly changed in the light of the evidence, running counter to the ''base-rate fallacy'' observed in other studies.

Besides being a base-rate problem, the bus problem can also be viewed as a partitioning task (e.g., Fox & Rottenstreich, 2003). The problem initially introduces a threefold partition of the outcome space, based on arrivals ahead of schedule, 0–10 min after schedule, and more than 10 min behind schedule. When the frequencies associated with these outcomes are incorporated, we get a 1/8/1 possibility space. The 10 min waiting time without the bus arriving indicates that the middle category is empty and can be eliminated, leaving a binary 1/1 partition. Under this analysis, three steps are required to solve the problem:

(1) Elimination of one (initially the most likely) possibility.
(2) Reconstruction of a new possibility space, where the two remaining, equally likely events can be compared.
(3) Translate the newly reconstructed possibility space in terms of probabilities. Consequently, in our example, the probabilities should be upgraded from 10% to 50%.

These steps correspond to the remove step and the re-standardize step discussed in theories of Bayesian conditionalization (Kleiter et al., 1997).

Fox and Rottenstreich (2003) have shown how response inconsistencies can arise in problems where the set of possibilities can be partitioned in different ways. People who are asked what is the probability that ''next week, the hottest day of the week will be Sunday'', are encouraged to form a sevenfold class partition: [Sunday hottest, Monday hottest, ...], yielding a probability of Sunday equal to 1/7. However, the same question can also be formulated so as to facilitate a case partition: ''What is the probability that Sunday will be hotter than every other day next week?'' This question implies two rather than seven alternatives, as Sunday can be hotter or not hotter than the other days. Participants that are disposed to follow the "principle of insufficient reasons", where cases are treated as equivalent under conditions of equal ignorance, will suggest a prior probability of 1/2 rather than 1/7 in response to the case partition question (Fox & Rottenstreich, 2003). Similar difficulties have been reported with combinatorial problems, where solutions differ according to how finely grained people structure the sample space (Keren, 1984).

The bus problem differs from previous partition problems in that it requires people to give up the original threefold class partitioning, and the base-rates corresponding to these partitions, in favour of a new binary case partitioning (the bus has gone/the bus will still arrive). Fox and Rottenstreich's (2003) results suggest that binary case partitions can unduly influence probability estimates, whereas the bus problem suggests that binary case partitions can also be inappropriately resisted.

In a pilot study, we gave a paper and pencil version of the bus problem to 222 students at the University of Oslo. Half were simply asked about a passenger's probability of catching (or missing) the bus on a day when he arrives at the bus stop exactly on time, and on another day when he arrives at the bus stop 10 min late (Group A). For the other half, the second question was replaced by a question of his chances on a day the passenger arrives on time, but waits for 10 min without seeing any bus (Group B). In both conditions, the probability questions were phrased in four different ways, each given to a different subgroup of participants. Two were framed in negative terms (what is the probability that the bus has left/that he has missed the bus) and two in positive terms (what is the probability that the bus will come/that he will catch the bus). The first question, in both these pairs, focuses on the probability of the bus, the second on the passenger's probability. The general patterns of answers were in all cases the same. Regardless of frame and person focus, a bimodal pattern of answers emerged: After waiting for 10 min, 47% of participants in Group B estimated his probability of catching the bus as low ($p < .25$), 44% thought it was still high ($p > .75$), whereas only 10% suggested a medium probability ($.25 < p < .75$). A similar number of 50% responses occurred also in Group A, where they were inappropriate, perhaps indicating that some 50/50 answers simply express uncertainty or ignorance (Bruine de Bruin, Fischhoff, Millstein, & Halpern-Felsher, 2000), instead of reflecting an explicit comparison of the base-rates involved.

An interesting aspect of the bus problem is that it initially encourages a representation in terms of "external uncertainty" (Kahneman & Tversky, 1982), based on frequentistic probabilities, which has usually been found to facilitate correct extensional reasoning (Lagnado & Sloman, 2004). However, in the present case, a reliance on frequencies seems to block the correct problem representation, which requires that the original threefold partition is replaced by a binary partition in two equiprobable, mutually exclusive and exhaustive categories, after the middle category of 0–10 min delays is eliminated. In their original study, Falk, Lipson, and Konold (1994) speculated that an external attribution of uncertainty may account for the "conservatism" shown by many of their subjects. It is apparently difficult to revise a probability based on external frequencies with knowledge based on personal and individual information.

The remainder of this paper explores several variants of the bus problem, along with some other tasks having the same logical structure, to investigate the robustness of the basic phenomenon, and to examine factors that might facilitate the correct response. It will be seen that people experience severe difficulties in integrating case information (about Fred's waiting time) with base-rate information (about the frequency of bus departures). Most frequently, the case information is simply ignored. Apparently, participants do not distinguish between unconditional probabilities (where no case information is presented) and the conditional probability of catching the bus after a period where no bus arrivals have been observed.

2. The present studies

2.1. Method

The studies to be reported below are based on responses from altogether 896 Dutch students, recruited from the Universities of Nijmegen, Tilburg, and Utrecht, who volunteered to take part in six different experiments. Within each wave of data collection, participants were randomly allocated to one (and only one) condition, and received 5 Euro for completing this and several other, unrelated tasks.

All participants were initially presented with a scenario on a computer screen, which in most cases described Fred who is travelling to work with a bus, along with the base-rates of early and late bus departures. After receiving information about the time of Fred's arrival and his waiting time, participants were asked to estimate the probability that the bus will still arrive. This formulation was used in all conditions, as the pilot experiment had shown that different ways of framing the outcome yielded essentially the same responses. Answers were given by checking a number on a probability scale from 0% to 100%.

As all participants belonged to the same student population, and were recruited and tested in the same way, we have, for the sake of simplicity, listed the different conditions with consecutive numbers from 1 to 16, instead of describing them as parts of six separate experiments. In doing so, we present them in a logical rather than a strict chronological order. Statistical tests are, however, only used to compare conditions where participants have been drawn randomly from the same subject pool.

2.2. Conditional versus unconditional probabilities

As mentioned already, solving the bus problem may be difficult due to the complexity in realizing the conditional nature of the problem. To give participants a chance to compare conditional and unconditional probabilities, the basic bus problem was given to four groups. Participants in Condition 1 (base line) were asked to make only one probability estimate, referring to a passengers' chance of catching the bus after a waiting period. The instructions were as follows:

> Fred travels every day to work by a bus that departs regularly on the hour (i.e., 6:00, 7:00, and 8:00 a.m.) from the station next to his house.
> Based on his long experience he noticed that, on the average, in one out of 10 cases the bus departs before schedule, in eight out of 10 cases it departs 0–10 min late, and in one out of 10 cases it departs more than 10 min late.
> Suppose that Fred arrives at the bus stop exactly on time and waits for 10 min *without the bus arriving*. What is the probability (chance) that the bus will still arrive?

The results are displayed in the first row of Table 1. The distribution of answers is clearly bimodal, with a majority of participants (63%) thinking that Fred has only a 10% chance of catching the bus, whereas a smaller, but distinct group (26%) thinks that the bus will certainly, or almost certainly arrive. Only one participant selected the correct 50% answer.

Table 1
Frequencies of conditional probability estimates, after one option has been eliminated from consideration

Condition	Probability estimates											*N*
	0	10	20	30	40	**50**[a]	60	70	80	90	100%	
Conditional vs. unconditional probabilities												
1. Conditional only (baseline)	–	24	3	–	–	1	–	–	–	7	3	38
2. Cond. after T_1	–	32	1	–	–	1	–	–	–	–	–	34
3. Cond. after T_2	–	26	2	1	–	1	–	–	1	5	–	36
4. Cond. after both	4	30	–	–	–	2	–	–	–	3	1	40
Fred and John												
5. Fred (early)	3	28	2	1	2	1	1	1	1	6	3	49
6. John (late)	1	16	1	1	–	3	2	1	1	9	15	50
Only two options remain												
7. Control (no hint)	–	36	–	1	–	4	–	–	5	11	5	62
8. With hint	1	33	2	–	1	16	1	–	4	3	1	62
Comparing the options												
9. With middle category												
a. Bus is gone	–	37	1	2	–	11	–	–	–	1	–	52
b. Bus will arrive	–	27	5	1	–	11	1	1	2	4	–	52
10. Without middle category												
a. Bus is gone	–	26	2	1	–	17	1	1	–	1	–	49[b]
b. Bus will arrive	1	21	1	3	1	15	1	1	–	4	1	50
Reorganizing options												
11. 2/6/2 bus	–	2	27	5	2	2	5	1	4	–	3	51
12. 6/2/2 bus	–	1	12	1	3	22	1	–	8	–	1	49
13. 2/6/2 diagnosis	–	3	7	1	–	50	–	–	–	–	1	62
Causally relevant partitions												
14. Three drivers	–	29	3	1	–	6	1	–	4	6	2	52
15. Three companies	–	7	1	–	–	40	1	–	–	–	–	49

[a] Correct answer = 50% in all conditions.
[b] One participant failed to answer this question.

Participants in the next three conditions answered the same question after first being asked to estimate unconditional probabilities. In Condition 2, they were asked first to estimate Fred's probability of catching the bus on a day when he arrives exactly on time. In Condition 3, participants were asked to estimate Fred's probability on a day when he arrives 10 min late. Finally, participants in Condition 4 had to answer both these questions before estimating Fred's chances the day when he had waited for 10 min. The unconditional estimates were introduced to highlight the difference between base-rate probabilities and conditional probabilities. For instance, in Condition 3 the correct answers would be a 10% chance the day Fred arrives 10 min late, against a 50% chance the day he arrived on time and waited for 10 min.

A majority of participants, in all conditions, answered the first questions concerning unconditional probabilities correctly, but failed on the conditional probability question. For instance, most participants (81%) in Condition 3 agreed that Fred's chances are only 10% when he arrives 10 min late, but 72% also claimed that he had a 10% chance when he arrives on time and waits for 10 min. Distributions of answers (Conditions 2–4) to the last question are presented in Table 1. They show, unequivocally, that initial questions about unconditional probabilities were completely ineffective in "debiasing" responses to the conditional probability question.

2.3. John meets Fred – highlighting timing

Perhaps the most parsimonious explanation of the preceding results is that people leave their probabilities unchanged because they think they are asked about unconditional prior probabilities throughout the experiment. In other words, they fail to incorporate timing (i.e., before and after waiting for 10 min). To make the contrast between the conditional and unconditional probabilities more compelling, we introduced a scenario with *two* passengers, one arriving early, with an initial 90% chance of catching the bus, the other very late, thinking that he has only a 10% chance. What will happen to their probabilities when they meet? It is difficult to believe that they will adopt completely different probabilities about their chances of catching the same bus. The waiting passenger's initial high expectations should be reduced as a function of waiting time, whereas the latecomer should pick up some hope from meeting the waiting passenger. This is a situation many of us will recognize from daily life: You approach a station fearing to be too late for a bus or a train, but are comforted from seeing other passengers still waiting.

Participants in Conditions 5 and 6 were accordingly introduced to two passengers, Fred and John, living in the same neighborhood, but not acquainted with each other. Both groups were asked to imagine that Fred arrives one day at the bus stop exactly at 8:00. What is his probability of catching the bus? They were then told that on the same day, John leaves home rather late, realizing on his way that he will arrive at the bus stop at 8:10. What will John think about his probability of catching the bus?

Participants in Condition 5 were finally told that at 8:10, when John arrives, Fred is still waiting. What are now, according to Fred, his chances of catching the bus?

Participants in Condition 6 were instead asked to take John's perspective. They were told that when John arrives at 8:10, he meets Fred, who tells him that he has been waiting for 10 min without the bus arriving. What are now John's chances of catching the bus?

Participants in both conditions agreed that Fred initially estimates his probability to be around 90%, while John, approaching the bus stop 10 min later, will think he has only a 10% chance. At the time of John's arrival, most participants in Condition 5 realized that Fred's chances have diminished. Yet, in line with previous conditions, they adjusted his probability all the way down to 10%. John, who in contrast had a 10% probability before meeting Fred, had his probability upgraded. As a result, Fred and John end up with different probability distributions, as shown in Table 1, rows 5 and 6. When the answers are grouped in low probabilities, medium, and high probabilities, the conditions differ significantly, $\chi^2(2, N = 95) = 10.94$, $p < .004$. In Condition 5, 67% of the participants assess Fred's probability as low ($p = 0$–.2), 8% as medium ($p = .4$–.6), and 20% as high ($p = .8$–1.0), compared with Condition 6 where 36% of the participants assess John's probability as low, 10% as medium, and 50% as high. (Four estimates, two of $p = .3$ and two of $p = .7$, fell between the categories and were excluded from the analysis.)

Thus, even if a majority thinks that probabilities should change during waiting time, they have only vague and mainly incorrect ideas about the amount of change that is required, leading to inconsistent answers. If we were to take these answers literally, we would expect an asymmetry in optimism between "newcomers" and "veterans" at the bus stop. Newcomers should feel encouraged by meeting "veterans" that have waited for a long time already, whereas veterans should become increasingly discouraged as time goes by. The fact that only four participants, in the two conditions combined, got the correct answer clearly suggests that the current manipulation was ineffective.

2.4. Only two options remain

Merely contrasting conditional with unconditional probabilities, or arranging a meeting between Fred and John, was clearly not sufficient to promote a redefinition of the appropriate outcome space. In another attempt to encourage participants to eliminate the middle category in the 1/8/1 partition, we reminded participants that after the waiting period, Fred has only two options left. This explicit hint suggests a transition from the original three-partition outcome space to a binary space, where the options are that the bus left early, or the bus will still come. Participants were also asked to provide confidence ratings of their estimates. If people experience the task to be difficult, and are considering several possible answers, this should be reflected in a high degree of uncertainty about what is the correct answer and thus in low confidence ratings. Alternatively, they may be convinced that their answer is the correct one, despite the disagreement between different respondents, in which case high confidence ratings should be expected.

Participants in Condition 7 (control condition) received the basic bus scenario with questions about Fred's probability of catching the bus after waiting for 10 min (as in Condition 1 – the base line). Subsequently, they were asked to indicate how certain they were about their answers, on an 11-point rating scale ranging from 0 (pure guess), to 10 (completely certain), and, in addition, to briefly explain how they got their answer (open question). Participants in Condition 8 received exactly the same vignette and the same questions, with the addition of one single statement: "Fred concludes that either the bus must have arrived before schedule or, else, it must be delayed with more than 10 min". This statement was inserted immediately before the probability estimate, with the purpose of facilitating a comparison between the two remaining, equally likely alternatives.

The distribution of answers, displayed in Table 1, shows that only four out of 62 participants in Condition 7 arrived at the correct answer (50%). It is evident from their explanations that these four participants had compared the two remaining alternatives, and found them equally likely. They were also quite confident in their answers (ratings of 9 and 10), indicating that this was not simply a "fifty/fifty"-answer based on ignorance (Bruine de Bruin et al., 2000). Most other participants believed that the probability had to be 10%, or else around 90%.

Fred's rather trivial conclusion in Condition 8, that "either the bus must have arrived before schedule, or else it must be delayed with more than 10 min", boosted the number of correct answers from 4 (Condition 7) to 16 (Condition 8). The difference in proportion correct answers between the two conditions was statistically significant (Fisher exact $p < .01$). Notwithstanding, the modal answer in Condition 8 was still 10%.

Participants in both conditions were generally quite confident about their 10% answers, median confidence = 8.0, whereas median confidence for those who answered 90% was 7.0, slightly (but not significantly) lower.

2.5. Comparing the remaining options

A correct solution of the bus problem requires that the probability associated with the waiting period is eliminated, the remaining possibilities are compared, and their probabilities are recomputed. Condition 8 suggested that instructions that promote the elimination of the waiting period, and encourage explicit comparisons between the remaining possibilities facilitate, but fall short of ensuring the correct solution. Two new conditions were designed to highlight the comparison between the two remaining alternatives even more directly, by asking participants explicitly which of these two alternatives is more likely. This "qualitative" judgment was followed by a "quantitative" question about numeric probabilities. The main issue is whether those who conclude that both alternatives are equally likely will recompute the probabilities so as to arrive at the correct 50/50 solution.

Participants in Condition 9 received the bus scenario with the same text as in Condition 1, whereas in Condition 10, the passage describing the middle category ("in eight out of 10 cases it departs 0–10 min late") was omitted. This was done to facilitate the comparison between the two remaining alternatives, which were both described as occurring "in 1 out of 10 cases".

All participants were asked the following questions:

1. Suppose that Fred arrives at the bus stop exactly on time and waits for 10 min *without the bus arriving*. What is more likely (choose one option):
   (a) The bus arrived too early
   (b) The bus will still arrive
   (c) Both options are equally likely.
2. Respond to the following questions on a scale from 0% to 100%.
   (a) What is the probability that the bus arrived too early?
   (b) What is the probability that the bus will still arrive?

In response to Question 1, a majority (74.3%) of the participants, combined over the two conditions, correctly chose (c): Both options are equally likely. In Condition 9, four participants believed that the bus arrived too early, seven that the bus will still arrive, and 41 that both options were equally likely. In Condition 10, the corresponding frequencies were seven, eight, and 34, respectively. The difference in proportion correct answers between the two conditions is statistically not significant.

Given these results, one could conclude that the probabilities for the two remaining alternatives would be 50/50. The distribution of answers to Question 2, presented in rows 9 and 10 of Table 1, shows that this was not the case. In both conditions, a majority thought that there is still a 10% chance that the bus has gone, and also a 10% chance that the bus has not arrived. Correct 50% answers were endorsed by 21–35% of the participants, slightly (but not significantly) more in Condition 10 where only these two alternatives were explicitly mentioned in the instructions. A closer look at those who responded ‘‘equally likely’’ to Question 1 showed that 67 of 75 responses were consistent in the sense that they actually produced the same probability estimates to both alternatives of Question 2. Of these, 22 gave 50% responses to both questions (there is a 50% probability that the bus has arrived early, and another 50% probability that it will arrive late), whereas 41 gave 10% responses both for ‘‘early’’ and ‘‘late’’ estimates.

The present experiment showed that even after realizing that the two remaining alternatives are equally likely, most responders continue to believe that the correct probability is only 10%. This adds up to a total probability of 20% rather than 100% for an exhaustive set of outcomes. The results also suggest that the major difficulty in solving the problem may lie in translating the new possibility space in probabilistic terms (the third phase in the three steps conjectural process proposed in the introduction).

2.6. Reorganizing options

The original 1/8/1 distribution of bus departures appears to have formed a strong ‘‘gestalt’’, resistant to restructuring. Two factors may contribute to this robust finding: (1) The first set of departures are ahead of schedule, and might have been discounted as exceptions, partly because people know, from experience, that such departures are generally rare (most bus lines have explicit rules against early departures, which, unlike delays, can be controlled by the bus driver). (2) The original problem requires the solver to compare two alternatives that are placed on either side of a middle category (the 0–10 min interval). Difficulties may arise because the alternatives to be compared are not adjacent, and the part of the prior distribution that has to be eliminated is the major, middle part.

In Conditions 11 and 12 we introduce two manipulations to investigate the effects of these factors. First, we introduce a 2/6/2 distribution where all frequencies refer to departures *after* schedule (0–5, 5–10, and 10–15 min, respectively). Second, this distribution is manipulated by changing the frequencies in one condition to 6/2/2. The correct solution requires in this case a comparison between two adjacent, equal categories, after the first category has been eliminated.

Condition 11 (2/6/2). Participants in this condition received the following scenario:

> "Fred goes regularly to work using a bus that departs on the quarter past the hour (i.e., 6:15, 7:15, and 8:15 a.m., and so forth). Based on his long experience he noticed that the bus is *never* too early. Further, on average, in two out of 10 cases the bus departs within the first 5 min, that is, between 0 and 5 min from the scheduled time. In six out of 10 cases it departs between 5 and 10 min from the scheduled time. Finally, in two out of 10 cases the bus departs with a delay of between 10 and 15 min (the bus has never a delay of more than 15 min).
> Suppose that Fred arrives at the bus stop at 8:20, that is, 5 min too late. He then waits for 5 min *without the bus arriving*. What is the probability that the bus will arrive within the next 5 min (i.e., with a delay of 10–15 min)?"

Condition 12 (6/2/2). Participants in this condition received an identical scenario, except that the base-rates of bus departures were six out of 10 cases in the 0–5 min interval, two out of 10 cases in the 5–10 min interval, and two out of 10 in the 10–15 min interval. Fred arrives at the bus stop at 8:15, that is, exactly on time. He then waits for 5 min without the bus arriving.What is the probability that the bus will arrive within the next 5 min (i.e., with a delay of 10–15 min)?

Most participants (53%) in the 2/6/2 condition believed that Fred had a 20% probability of catching the bus. Only two participants gave the correct 50% answer. Thus, the 2/6/2 partition, where none of the departures is before schedule, did not produce more correct responses than the 1/8/1 partitions, with one departure ahead of schedule, as was the case in the previous experiments.

The 6/2/2 partition produced many correct responses, 50% probability being the most frequent answer. The proportion of correct answers in Condition 12 was significantly larger than in Condition 11 ($p < .001$, Fisher exact probability test). Evidently, it is easier to arrive at the "correct" 50% response when the alternative to be eliminated comes first, and the two remaining alternatives are adjacent in time and thus easily comparable. Solution of the 6/2/2 problem may also be facilitated by the fact that both remaining possibilities belong to the future. In contrast, the 2/6/2 problem requires a comparison to be made between a past and a future possibility. Still, even in the 6/2/2 condition, many participants (approximately 25%) stuck to the prior 20% probability.

2.7. Change of context

In the following condition, the bus problem was substituted with a medical problem having a similar logical structure: There are three possible diagnoses with prior frequencies of 2/6/2 (same frequencies as in Condition 11), until a blood test proves that the most likely diagnosis can be ruled out. We predicted that the probabilities of the remaining two diagnoses would be easy to recalculate, for two reasons: (1) Ruling out a diagnosis effectively eliminates this option from the set of possibilities, by suggesting a change of the event space on the class level (2) There is no temporal dimension or other natural order of events that prevents the remaining diagnoses from being directly compared.

Condition 13 (2/6/2 diagnosis problem). The scenario given to participants in this condition was the following:

> "Fred feels sick for several days and finally decides to see a physician.Given Fred's symptoms, the physician concludes that Fred may suffer from either disease A, B, or C. Based on his long experience the physician notes that, on average, in two out of 10 cases Fred's symptoms entail disease A. Further, in six out of 10 cases the symptoms entail disease B, and in two out of 10 cases they entail disease C. As a first step, the physician refers Fred to take a blood test that is only testing whether the patient has disease B. Evidently, the results of the blood tests are negative. Therefore, the possibility that Fred has disease B is ruled out with certainty".

One half of the participants was then asked what is the probability that Fred has disease C. The other half was asked the same question about disease A. As there was no difference between these answers, the results were pooled.

In this condition, most participants realized that they had to eliminate the most likely alternative and then recompute the appropriate probabilities. More than 80% (50 out of 62) estimated the probability for either of the two remaining, previously unlikely, diagnoses to be around 50%. The difference in proportion correct answers between Conditions 12 and 13 is highly significant ($p < .001$). The results from Conditions 11 to 13 indicate that the difficulties people experience with the bus problem cannot simply be attributed to the "before schedule" category, used in the previous conditions, but must be partly due to difficulties in eliminating the empty set and comparing the remaining two alternatives. When they are juxtaposed, either by directing attention to the fact that these are the only two alternatives left (as in Condition 8), or by asking for an explicit comparison (as in Conditions 9 and 10), or by placing them side by side on a temporal dimension, as in Condition 12, 50%-answers become more frequent. Still, many people seem to be reluctant to give up the prior probabilities. Only in Condition 13, in which the content (but not the structure!) has been fundamentally changed, a high proportion of correct answers was observed. Eliminated diagnoses, as opposed to missing buses, imply a change in the parent distribution that makes it obvious that the probabilities should also change.

2.8. Causally relevant partitions

The results from Conditions 9 and 10 showed that one possibility can be eliminated, and the remaining possibilities compared, without their probabilities being recomputed. In contrast, in the medical diagnosis problem (Condition 13), the two unlikely diagnoses became 50% probable after the most likely diagnosis had been eliminated. It is possible that the latter context is more concrete and seems to offer a causal pattern, which induced participants to entirely remove the eliminated option. The next two conditions, constitute an attempt to make the 1/8/1 partition of bus departures more concrete and causally relevant, by being based either on three different categories of bus drivers (Condition 14), or on buses belonging to three different companies (Condition 15).

Participants in Condition 14 were told that 10 drivers operated the line. Of these, one driver is known to be constantly in a hurry and thus always arrives at the bus stop before schedule. Another eight drivers arrive 0–10 min late. Finally, one of the 10 drivers is particularly slow, and his bus always arrives more than 10 min late.

Suppose that Fred arrives at the bus stop exactly on time and waits for 10 min without the bus arriving. What is the probability that the bus will still arrive?

Participants in Condition 15 were told that Fred can go to work by buses belonging to three different companies, A, B, and C. Based on his experience, Fred knows that one out of 10 buses belongs to company A, eight belong to company B, and one to company C. One day he arrives at the bus stop and waits for some time without the bus arriving. He is then told by another passenger that the drivers of Company B are today on strike. What is the probability that the first bus that comes, belongs to Company C?

As can be seen from the distribution of answers in Table 1, introducing three types of drivers had little facilitating effect. Although the waiting time has eliminated the "normal" drivers, only six participants (correctly) think that the probability of a bus arrival is 50%; the majority (56%) believes there is still only a 10% chance.

In contrast, the announcement of a strike in company B (Condition 15) made it easy to see that the probability of a bus from company C is now 50%. Eliminations due to strike is evidently more potent in transforming the situation than eliminations due to waiting time. After all, a strike decisively changes the prior distribution, whereas waiting time appears to leave it intact.

2.9. When will Fred change his opinion?

When will Fred start to feel that the bus has already gone? In all the previous conditions, participants were asked to estimate the probability corresponding to a specific waiting time. In Condition 16, they were asked instead to estimate the waiting time corresponding to a specific probability, namely how long will it take a passenger to believe that $p(\text{bus is gone}) = p(\text{bus will still come}) = .5$.

Participants in this condition were given the basic scenario about Fred who arrives on time and waits, for 10 min, without the bus arriving. When will Fred change his belief from being more than 50% certain that the bus will come, to becoming more than 50% certain that the bus has already gone? One half of the participants received the question in a negative-outcome frame: "At what point in time will Fred become more than 50% certain that the bus has already gone (i.e., that he missed the bus)?" The other half received the same question in a positive-outcome frame: "At what point in time will Fred become less than 50% certain that the bus will still arrive (i.e., that he will catch the bus)?" Participants had to complete the following sentence: "I think Fred will become less than 50% certain that the bus will still arrive [more than 50% certain that the bus has already gone] after . . .. . . minutes".

Both frames led to very similar distributions of answers, so responses from the two subgroups were pooled.

The question format in this condition highlights the fact that Fred's probability actually changes by waiting, a suggestion seemingly overlooked by those who claimed, in the other conditions, that even after 10 min, his probability would remain constant at 90%. To be consistent, these respondents should maintain that Fred would wait forever, without losing hope. The majority, who proposed that Fred's initial 90% probability has sunk to 10% after 10 min, could be expected to think that the turning point from belief to doubt takes place in the middle of the 10 min waiting period. The distribution of answers, displayed in Fig. 1, shows that a substantial number of participants (37%) indeed believed that the turning point comes after a waiting time of about 5 min. Yet, an even greater number of respondents (46%) placed the 50% point around 10 min, in agreement with the normative, Bayesian account of probability revision. Very few thought it would take *more* than 10–11 min to realize that the bus has probably gone. Thus, when the task is changed from probability estimation to time estimation, the frequency of correct responses increases dramatically. From being the most infrequent answer it has become the most popular one.

The question format used in Condition 16 might have facilitated the correct response in several ways. First, by suggesting that Fred's probabilities actually

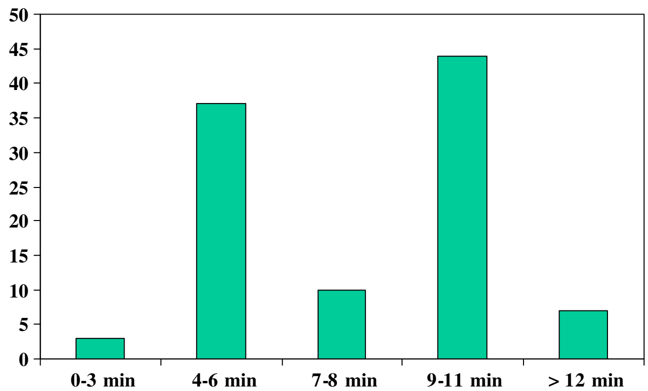

Fig. 1. When will Fred start to think that the bus has probably gone? Percentages of answers, Condition 16.

change, and that they will at some point reach a level of 50% probability. This *p* value had not been explicitly mentioned in the other conditions, and had to be inferred from the novel 1/1 partition (which most participants failed to do). These suggestions, however, do not explain why the 50/50 point should be placed around 10 min and not earlier (or later). A plausible, albeit speculative, explanation would be that the present scenario focuses more strongly on Fred's situation and how his beliefs develop over time, favoring an internal rather than an external interpretation of probabilities. While base-rates are clearly external, their role in the present scenario is to inform Fred's initial beliefs, which also can be modified by other, case dependent pieces of information, like the duration of his waiting period.

3. General discussion

Laypeople and experts have often been criticised for relying too strongly on single observations and not taking base-rates sufficiently into account. In contrast, the present results appear to demonstrate a bias in the opposite direction. In several experiments, participants estimated probabilities for bus arrivals by an exclusive reliance on base-rates and a corresponding neglect of what they had observed. In many conditions, waiting time was treated as an irrelevant piece of information, with no bearing on the probability estimation task. Under these circumstances, statistical information is not overruled by case information, as in the classical demonstrations of the base-rate fallacy. On the contrary, the case information appeared to be overruled by statistics, creating a kind of "base-rate perseverance" or base-rate fallacy Type II.

The different modifications of the basic bus scenario can be regarded as "debiasing" attempts, aiming to cue participants toward the correct response. Many of these manipulations proved to be ineffective, or were only partly successful, demonstrating the robustness of the base-rate perseverance phenomenon. For instance, simply contrasting probabilities conditional on waiting time with unconditional probabilities, where waiting time was not involved, did not increase the number of correct answers. Focus on the remaining options, along with manipulations that facilitate (or emphasize) the elimination of one of the original partitions, were somewhat more successful. Yet the eliminated category did not appear to be completely discounted. Although Fred can conclude that as far as his observations go, there was no bus arrival in the 0–10 period, most participants seem to overlook the effect upon the probabilities of late and early bus arrivals. They can both still be 10%. Many participants were evidently still thinking in terms of the original, external statistics, instead of reorganizing the possibility space and recalculating probabilities accordingly.

In their discussion of probabilistic versus statistical problems, Girotto and Gonzalez (2001) observe that Bayesian frequency problems are often considered "solved" when people fill in the correct numbers in a statement of the form: ...... out of ...... (e.g., Gigerenzer & Hoffrage, 1995). In other words, they are asked to provide the two terms of a ratio, without having to compute the ratio itself. In probabilistic versions of the same problem, they are asked for a single numerical probability (percentage), requiring this computation to be carried out. Thus, the two tasks differ not only in information format (frequencies versus probabilities), but also in question format (asking for two non-normalized terms versus one normalized ratio). This is not a trivial difference, as can be seen from Conditions 9 and 10, where the problem first seem to be "solved", when people noticed that the odds for catching the bus and missing the bus were the same. Still they fail to make use of this knowledge when they were asked to compute the probabilities.

In our initial analysis of the bus problem, we observed that a correct solution requires three steps: (1) elimination of one category, corresponding to the waiting period; (2) comparison of the two remaining possibilities; and (3) recomputation of the probabilities involved. Participants who claim that Fred's probability, after waiting, is 10% or 90% may have failed to perform all, one or two of these steps. For instance, a 90% response could indicate that the participant realizes that the probabilities have to be recomputed to add up to 100%, but fails to eliminate the 80% probability associated with the waiting period. A 10% response could also reflect elimination or comparison failures, but since a large number of 10% responses persisted even in conditions where most participants admitted that only two (equally probable alternatives) remain, the main obstacle to a correct response seems to reside in the recomputation phase. However, this obstacle was easily overcome in two conditions where a causal category (a medical diagnosis or a bus company) was eliminated. Thus, "eliminations" seem to take place at two levels: On a causal, class level, and on an observational, singular case level. Eliminations on the class level lead to recomputation of probabilities, whereas eliminations on the case level appear to be inconsequential for probability estimates.

This idea is portrayed in the following table:

1. Class level (frequentistic)	Early bus departures 1 of 10	Intermediate departures 8 of 10	Late departures 1 of 10
2. Case level (singular)	Missed bus $p = ?$	Waiting period no bus	Bus still to come $p = ?$

When people are given frequentistic probabilities, as in our cover stories, the problem is formulated on a class level, where a "parent distribution" is introduced. However, these parent probabilities can readily be applied to the individual case, guiding a passenger's unconditional expectations about early and late buses. Thus, Fred is correctly assumed to have a 90% probability of catching the bus when he is on time, and 10% when he is 10 min late.

Probabilities that are introduced on the class level, in terms of 'objective' (frequentistic) values, will not be changed unless the parent distributions themselves undergo a change. This is what happens in the diagnosis problem or with the bus strike, where a whole class of events is eliminated from the distribution. Information indicating such changes on the class level is seen to have consequences for expectancies on the case level as well; thus the occurrence of a strike in one bus company indicates that only buses in the first or third category are left. With a new, binary possibility space, both can be expected with a probability of 50%.

Case expectancies may also change in the course of an individual situation, for instance as a function of waiting time. This was evident in Condition 16, where participants were asked to estimate the duration of waiting time corresponding to $p = .5$. However, when asked to report numeric probabilities, as was the case in most other conditions, they went back to the class level and reported values corresponding to the parent distribution. When this distribution was still intact, they reported unchanged p values (10% or 90%).

The distinction between a class and a case level can, under some circumstances, be conceptualized as a cause–effect relationship; the parent distribution of bus departures can be seen as "causing" individual expectations, while expectations that change across a concrete waiting episode are not granted the same power to "cause" a revision of chances at the class level. The two levels may also be related to the distinction between "external" and "internal" probabilities (Kahneman & Tversky, 1982). The original, threefold partitioning of the outcome space clearly refers to a distribution of objective, external events, whereas Fred's expectancies of catching or not catching the bus is more easily conceived as reflecting his internal state of knowledge. As the problem is initially introduced in terms of external probabilities, most participants will naturally search for an answer on that level. Only a strong focus on individual expectancies (as in Condition 16) transforms the problem into a question of internal probabilities, and makes it possible to construct a new, binary possibility space on this level.

One additional source of difficulty may arise from the fact that the category to be eliminated in the present case is a span of time, where nothing happens. The diagnosis problem and the case of three bus companies differ from the standard bus problem not only by changing the parent distributions, but also by being based on a natural division into discrete categories, rather than reflecting arbitrary divisions on a time dimension. However, Falk et al. (1994) found in their original study that people also experience difficulties in the "desk problem", which implies a search for a target object in a successive series of drawers. Drawers constitute discrete categories on a spatial rather than a temporal dimension. Still, many of their subjects made similar errors on this task, as they did on the bus problem.

Regardless of interpretation, our results show that people experience difficulties combining probabilistic information coming from different sources. In most of our debiasing attempts, we failed in "priming" the binary case partitioning, supposedly because the class partitioning was already so strongly "primed" by the background information about prior frequencies.

The problem of combining information is complicated by the fact that the base-rates are given in a numeric format that is easily transformed into probabilities, whereas the case information describe waiting times and no probabilities. Thus, the numeric probabilities of .10 and .90 (or .20 and .80) are highly available, whereas the correct probability of .50 is not presented as an option, except on the response scale, where it is listed along with 10 other numbers. In informal classroom exercises, we have made .50 a more prominent option by informing the participants that 10%, 50%, and 90% are the most common answers, and asking them which of these three answers they endorse. This was given to triads of students, who were asked to come to an agreement about the correct answer, after a group discussion (the three options were distributed among the three group members, trying to find convincing arguments in favour of "their" option). By this procedure, some groups will come to the conclusion that 50% must be the correct answer, but in the exercises we have performed so far, the majority still feel 10% and 90% to be a more convincing conclusion.

In addition to highlighting the 50% option, one can reduce the impact of the 10% (or 90%) options by using verbal rather than quantitative descriptions. Students working in computer rooms at the University of Oslo were asked to participate in an internet-based experiment. They were informed that the task involved a question about probability estimation. Those who logged on and agreed to go through the task were randomly allocated to different conditions. Participants in one condition ($n = 43$) were given the basic bus scenario, further simplified by offering only three response alternatives: 10% (22), 50% (7), and 90% (14). The response distribution (in parentheses) shows that only seven participants (16.3%) chose the correct response. Participants in another condition ($n = 37$) were given the same scenario, but without numeric base-rates. Instead they were informed that the bus came "rarely too early", "usually 0–10 min late", and "rarely more than 10 min late". Fred arrives on time and waits for 10 min without the bus arriving. What is Fred's probability of catching the bus? The response alternatives were: "low probability" (6), "about 50%" (24), and "high probability" (7). By this manipulation, the correct answer was given by 64.9% participants. This is a highly significant improvement, $\chi^2(1, N = 80) = 19.77$, $p < .001$, above Condition 1, indicating that when the base-rates are not given in terms of exact numbers, they lose much of their ability to selectively dictate the solution. Further research is in order to replicate this effect, and to decide how much is due to the verbal base-rates, the verbal response alternatives, or the prominence of the 50% alternative as the only remaining number.

The bus problem bears some similarity to other probabilistic "teasers", where the probability of one option is changed due to the elimination of another option. For example, in the celebrated Monty Hall problem, participants are told that a prize is hidden behind one of three doors. After the player has chosen one door, the game host opens one of the two remaining doors showing that there is no prize behind. Should you stick to your original choice or should you switch to the door the host chose not to open? Most people (including many statisticians) see no reason for switching, finding it difficult to accept the fact that the probability of the single remaining, non-chosen and non-opened door has now increased, $p = 2/3$, due to the elimination of another door (Granberg & Dorr, 1998; Krauss & Wang, 2003). The Monty Hall problem differs, however, from the bus problem in several ways. In the bus problem, when the central possibility (of 80%) is eliminated, the posterior probabilities stay equal, whereas in Monty's case, ruling out one door renders unequal probabilities to the remaining two doors. The Monty Hall problem also differs by asking for a decision (switch versus stay) rather than a probability estimate. Usually, participants' unwillingness or indifference to switching is believed to signify a 50/50 probability estimate. We found, however, that indifference (equal probabilities) in the bus problem was believed to imply only 10% chances for either option.

Thus, it is possible that at least some participants in the Monty Hall problem regard the probability of finding the prize behind the unopened doors to be 33/33 rather than 50/50. To be undecided between two 50/50 options, two 33/33, or two 10/10 is, of course, logically equivalent. Yet, from a psychological perspective, we might think that a situation with equal probabilities for the remaining options will be perceived as more of a dilemma if the prior probabilities were low, than when they were high. A choice between two implausible alternatives is likely to create more conflict and indecisiveness than a comparable choice between two plausible alternatives. We would also predict that information about either outcome (that the bus came early, or that it will come) would be more surprising in a non-normalized 10/10 situation than in a "normal" 50/50 situation (Teigen & Keren, 2003).

Other studies have shown that when people are willing to revise the probability of one option in an exhaustive set, they often do not see the necessity of adjusting the probabilities of other options. For instance, Robinson and Hastie (1985) presented a murder mystery with several suspects and asked people to estimate the probabilities for each of them. At some point one of the suspects received an alibi, and could thus be eliminated from the set. This had a strong effect on this individual's probability (which went down to zero), but did not raise the probabilities of the other suspects accordingly. Such failures to redistribute and recalculate probabilities after one alternative is eliminated, may be linked to a general belief of outcome probabilities as stable attributes of events, which are allowed to be changed only when causally relevant aspects of the situation are affected. Keren and Teigen (2001) found, in line with this, that people often tend to draw dispositional inferences from probability statements, believing for instance that an earthquake which has a high probability of occurring will be stronger, and occur sooner, than less probable earthquakes.

A dispositional interpretation of probabilities will easily come in conflict with the fundamental convention of probability theory, requiring that p values for a set of mutually exclusive and exhaustive events should add up to 1. Additivity violations have been demonstrated in a number of studies, including tasks where only two, complementary outcomes are assessed, leading in some cases to subadditive (Yamagishi, 2002), and in other cases to superadditive estimates (Mandel, 2005). This research indicates that people see no compelling reason to obey the rule of binary complementarity, unless the task is made transparent, for instance by asking for the probability of a target event and for its complement in close succession. A correct solution of the bus problem requires that participants keep the 100% rule in mind throughout and apply it to the changed situation where only two, equiprobable options remain. If this rule is not salient and given priority, noncomplementary estimates will easily ensue.

In sum, it seems that people stick to the prior probabilities, or base-rates, in the bus problem for four main reasons: (1) they seem to be objectively given (describing the external parent distribution), (2) they are given in terms of numbers, (3) responses are also to be given in terms of numbers (on a numeric probability scale), (4) people tend to think about probabilities as inherent properties of the events in question, and (5) they easily neglect the principle of additive complementarity when the situation changes. The idea of recomputing probabilities will only emerge in transparent tasks where one, or several, of these presuppositions are challenged and put to test.

References

Baron, J. (2002). *Thinking and deciding* (3rd ed.). Cambridge: Cambridge University Press.

Birnbaum, M. H. (2004). Base-rates in Bayesian inference. In R. F. Pohl (Ed.). *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* (pp. 43–60). Hove, UK: Psychology Press.

Bruine de Bruin, W., Fischhoff, B., Millstein, S. G., & Halpern-Felsher, B. L. (2000). Verbal and numerical expressions of probability: "It's a fifty-fifty chance". *Organizational Behavior and Human Decision Processes, 81*, 115–131.

Falk, R., Lipson, A., & Konold, C. (1994). The ups and downs of the hope function in a fruitless search. In G. Wright & P. Ayton (Eds.). *Subjective probability* (pp. 353–377). Chichester, UK: Wiley.

Fox, C. R., & Rottenstreich, Y. (2003). Partitioning priming in judgment under uncertainty. *Psychological Science, 14*, 195–200.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency format. *Psychological Review, 102*, 684–704.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases*. Cambridge: Cambridge University Press.

Girotto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: a matter of information structure and question form. *Cognition, 78*, 247–276.

Granberg, D., & Dorr, N. (1998). Further exploration of two-stage decision making in the Monty Hall dilemma. *American Journal of Psychology, 111*, 561–579.

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

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition, 11*, 143–157.

Keren, G. (1984). On the importance of identifying the correct 'problem space'. *Cognition, 16*, 121–128.

Keren, G., & Lewis, C. (1994). The two fallacies of gamblers: Type I and Type II. *Organizational Behavior and Human Decision Processes, 60*, 75–89.

Keren, G., & Teigen, K. H. (2001). The probability-outcome correspondence principle: a dispositional view of the interpretation of probability statements. *Memory & Cognition, 29*, 1010–1021.

Kleiter, G. D., Krebs, M., Doherty, M. E., Garavan, H., Chadwick, R., & Brake, G. (1997). Do subjects understand base-rates?. *Organizational Behavior and Human Decision Processes 72*, 25–61.

Koehler, J. (1996). The base-rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences, 19*, 1–53.

Krauss, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General, 132*, 3–22.

Lagnado, D. A., & Sloman, S. A. (2004). Inside and outside probability judgment. In D. J. Koehler & N. Harvey (Eds.). *Blackwell handbook of judgment and decision making* (pp. 157–176). Oxford: Blackwell.

Mandel, D. R. (2005). Are risk assessments of a terrorist attack coherent?. *Journal of Experimental Psychology: Applied 11*, 277–288.

Nisbett, R. E., Krantz, D. H., Jepson, D. H., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review, 90*, 339–363.

Pohl, R. F. (Ed.). (2004). *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. Hove, UK: Psychology Press.

Robinson, L. B., & Hastie, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 443–456.

Teigen, K. H., & Keren, G. (2003). Surprises: low probabilities or high contrasts?. *Cognition 87*, 55–71.

Wright, G., & Ayton, P. (Eds.). (1994). *Subjective probability*. Chichester, UK: Wiley.

Yamagishi, K. (2002). Proximity, compatibility, and noncomplementarity in subjective probability. *Organizational Behavior and Human Decision Processes, 87*, 136–155.