

In press, Organizational Behavior and Human Decision Processes

Frequency Illusions and Other Fallacies

Steven A. Sloman

Brown University

David Over

University of Sunderland

Lila Slovak

Brown University

Please address correspondence to:

Steven Sloman  
Cognitive and Linguistic Sciences  
Brown University, Box 1978  
Providence, RI 02912  
Email: [Steven.Sloman@brown.edu](mailto:Steven.Sloman@brown.edu).  
Phone: 401-863-7595  
Fax: 401-863-2255

## Abstract

Cosmides and Tooby (1996) increased performance using a frequency rather than probability frame on a problem known to elicit base-rate neglect. Analogously, Gigerenzer (1994) claimed that the conjunction fallacy disappears when formulated in terms of frequency rather than the more usual single-event probability. These authors conclude that a module or algorithm of mind exists that is able to compute with frequencies but not probabilities. The studies reported here found that base-rate neglect could also be reduced using a clearly stated single-event probability frame and by using a diagram that clarified the critical nested-set relations of the problem; that the frequency advantage could be eliminated in the conjunction fallacy by separating the critical statements so that their nested relation was opaque; and that the large effect of frequency framing on the two problems studied is not stable. Facilitation via frequency is a result of clarifying the probabilistic interpretation of the problem and inducing a representation in terms of instances, a form that makes the nested-set relations amongst the problem components transparent.

The greatest influence on the study of human judgment under conditions of uncertainty has been the “heuristics and biases” program, initiated by the work of Kahneman and Tversky in the early 1970s (cf. Kahneman, Slovic, & Tversky, 1983). This program has focused on judgmental error in order to reveal the heuristics and basic principles that govern human reasoning. Recently, a revisionist opinion has developed, arguing that the heuristics and biases program is deeply flawed because it fails to understand behavior in its ecological context. The reason, on this view, that the program has uncovered so much error is because it has primarily asked people to make judgments of single-event probabilities; i.e., the probability of one-time occurrences. This is inappropriate, detractors say, because people did not evolve to make single-event probability judgments; they evolved to make judgments about natural frequencies. Ask people to judge the frequency of events and many errors can disappear.

Proponents of some form of the natural frequency hypothesis include Gigerenzer and Hoffrage (1995) who claim, “An evolutionary point of view suggests that the mind is tuned to frequency formats, which is the information format humans encountered long before the advent of probability theory” (p. 697). Gigerenzer (1998) states, “If there are mental algorithms that perform Bayesian-type inferences from data to hypotheses, these are designed for natural frequencies acquired by natural sampling, and not for probabilities or percentages” (p. 14). The view is echoed by Cosmides and Tooby (1996): “[Humans] evolved mechanisms that took frequencies as input, maintained such information as frequentist representations, and used these frequentist representations as a database for effective inductive reasoning” (p. 16).

Cosmides and Tooby (1996) make the evolutionary argument for the hypothesis most clearly. Our hominid ancestors in the Pleistocene, they say, were able to remember and share specific events that they had encountered and, indeed, this was all they had available to make judgments under uncertainty. They didn't evolve probability estimators because the 'probability' of a single event is intrinsically unobservable. Hence, what evolved, according to Cosmides and Tooby, was an algorithm for computing ratios of counts of specific events.

Cosmides and Tooby's (1996) argument suggests that probability judgments are likely to be accurate when they concern frequencies, but not necessarily when they concern one-time events. More recently, Brase, Cosmides, and Tooby (1998) have followed Kleiter (1994) in making the much weaker claim that people are able to solve probability word problems that are in a very specific format. This format is called "natural frequency via natural sampling" by Gigerenzer and Hoffrage (1999). A central property of natural frequencies is that they are not normalized; instead, they combine information about effect and sample size. Gigerenzer and Hoffrage (1995), also following Kleiter (1994), argue that a critical virtue of natural frequency representations of numerical information is that correct conclusions can be reached with fewer computational steps than they can with relative frequencies or probabilities. The greater computational simplicity afforded by natural frequencies for the specific problems studied by Gigerenzer and Hoffrage and for those studied in the first part of this paper is an important point that we doubt anybody would disagree with. The issues that we will address in this paper are whether this computational simplicity is limited to natural

frequencies and, consequently, whether the computational simplicity has the claimed evolutionary source.

Some of the evidence favored by natural frequency proponents is encapsulated in the claim that certain cognitive illusions disappear when frequency judgments are substituted for single-event probability judgments. Gigerenzer (1994), for example, claims that the illusion of control, the conjunction fallacy, and base-rate neglect all disappear when questions are asked concerning frequencies rather than probabilities. One reason that natural frequency formats reduce the incidence of illusions, according to Hertwig and Gigerenzer (1999), is that the word “probability” is polysemous whereas the natural language sense of “frequency” is primarily mathematical. Gigerenzer, Hoffrage, and Kleinbölting (1991) and Juslin, Winman, and Olsson (2000) argue that overconfidence can be manipulated by varying the representativeness of the sample of questions that are used. These claims have been disputed (e.g., Brenner, Koehler, Liberman, & Tversky, 1996; Griffin & Buehler, 1999). A general argument against facilitation by frequency relative to probability judgment was made by Kahneman and Tversky (1996), who pointed out that biases have been demonstrated with frequency judgments since the onset of the heuristics and biases program.

The current paper focuses on two cognitive illusions that have been reported to show some of the largest effects of frequency versus single-event probability frames, base-rate neglect and the conjunction fallacy. We compare, both empirically and conceptually, the natural frequency hypothesis to the “nested-sets” hypothesis, that the effect of frequency is due to greater transparency in critical set relations induced by framing problems in terms of instances rather than properties.

## Base-rate neglect

Consider the following problem first posed by Casscells, Schoenberger, and Grayboys (1978) to 60 students and staff at Harvard Medical School:

If a test to detect a disease whose prevalence is 1/1000 has a false positive rate of 5%, what is the chance that a person found to have a positive result actually has the disease, assuming you know nothing about the person's symptoms or signs? Assuming that the probability of a positive result given the disease is 1, the answer to this problem is approximately 2%. Casscells et al. found that only 18% of participants gave this answer. The modal response was 95%, presumably on the supposition that, because an error rate of the test is 5%, it must get 95% of results correct.

Cosmides and Tooby (1996) tested the natural frequency hypothesis using several instantiations of a formally identical problem. One manipulation used the following wording that states the problem in terms of relative frequencies and asks for a response in terms of frequencies:

Frequency version with transparent nested-sets relations:

1 out of every 1000 Americans has disease X. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease. Imagine that we have assembled a random sample of 1000 Americans. They were selected by lottery. Those who conducted the lottery had no information about the health status of any of these people.

Given the information above, on average, how many people who test positive for the disease will *actually* have the disease? \_\_\_\_\_ out of \_\_\_\_\_.

In this frequency version, Cosmides and Tooby found that most people gave about the right answer (of the 50 or 51 people who test positive, only 1 has the disease). 72% did so in their Experiment 2, Condition 1 and 80% in their Experiment 3, Condition 2. They interpret this high performance as evidence that people are adapted for frequency.

Several authors have proposed an alternative hypothesis to explain this effect, the nested-sets hypothesis (Ayton & Wright, 1994; Evans, Handley, Perham, Over, & Thompson, 2000; Girotto & Gonzales, 2001; Johnson-Laird, Legrenzi, Girotto, Legrenzi, & Caverni, 1999; Kahneman & Tversky, 1982; Mellers & McGraw, 1999). Our version of this alternative hypothesis can be broken down into four assertions:

1. Frequency descriptions induce a representation of category instances (an outside view), rather than category properties (an inside view).

People normally represent a category, e.g. patients who have been tested for a disease, from a perspective that gives access to the category's internal structure – relations amongst its features or properties – perhaps by considering a prototypical instance. Such a representation is useful for purposes like similarity judgment and categorization but does not always afford coherent probability judgment. The frequency frame induces a different perspective by asking participants to think about multiple instances of the category and so about the set or class corresponding to the category.

2. Representing instances can reveal the set structure of a problem.

This is a fairly direct consequence of representing the instances of the categories in a problem. These instances make up the sets or classes that correspond (in an outside

view) to the categories. Most representational schemes that identify instances and the categories they belong to will automatically also specify the set structures relating the categories (e.g., Euler circles, mental models, taxonomic hierarchies, see Figure 1).

3. Revealing set structure can make nested-set relations cognitively transparent for problem solving.

This assertion is the most psychologically contentful, and – for that very reason – the most ill-specified. The idea is that a representation of a problem that exposes certain relations thereby draws attention to them and makes them potentially relevant to a problem solver. Sometimes, however, elementary set operations, such as taking the complement of a set or partitioning it, or taking the intersection of two sets or their union, are necessary before a nested-set structure can be achieved. Cognitive resources are of course bounded, so it is not always easy to use even the elementary operations to get a nested-set structure. Presumably, those relations and operations that require the least processing from the initial representation are the most likely to be perceived and used. The easiest relations to extract are the most elementary: set membership and set inclusion. These determine the subset relations in nested-set structures. How much of the set structure of a problem is revealed depends on the complexity of the problem, details of its form, and the working memory capacity of the problem solver. The account of sets of mental models in Johnson-Laird et al. (1999) provides a more complete specification of one approach that is consistent with this hypothesis. Indeed, Gigerenzer and Hoffrage (1999) allude to a relation between frequencies and mental models. The relation we see is that frequency formats are one way to enable mental model-like set representations.

4. Arithmetic operations that follow from transparent nested-set relations are easy to perform generally and not just in frequency problems.

For example, suppose you are asked for the probability that your car will start if the throttle is open. Imagine that you believe the probability that the car starts in general is .2 and the probability that the throttle is open is .8. Because the throttle must be open for the car to start, the car starting can be represented as a subevent of the throttle-being-open event, which makes it easy to see that the answer is  $.2/.8 = 1/4$ . Even more generally, the ability to think about sets and subsets, their relations, and their relative sizes is necessary for many problems. These include problems that would have been important under primitive conditions, such as the dividing up or sharing of resources.

To explain facilitation on the medical diagnosis problem, the nested-sets hypothesis assumes that an effective representation of the three relevant categories is isomorphic to the Euler circles of Figure 1. In this relatively easy case, the representation makes explicit, without any additional set operations, that the set of possibilities in which a person tests positive is a subset of all possibilities and the set corresponding to having the disease is a subset of the positive tests. Once these nested-set relations are understood, the answer can easily be seen to reflect the degree to which the subset corresponding to having the disease covers the set corresponding to a positive test, approximately represented by 1 out of 50 or 51. Our hypothesis is that frequency frames can increase performance by eliciting such a representation and that other frames can too.

The discussions in Gigerenzer and Hoffrage (1995) and Gigerenzer (1998) make two claims. One is that the natural sampling of a natural frequency under primitive conditions was adaptive. The second is that problems about natural frequencies are

computationally easy for people. We acknowledge that natural frequencies can increase computational simplicity, but we do not see how the claim of adaptiveness provides any further explanatory power. We agree with these authors that computational simplicity is inversely proportional to the number of mathematical steps required to solve a problem, but this has nothing to do with biological evolution. Moreover, computational simplicity cannot by itself explain why people find these problems so easy. Problems can only be easy for people if represented in a way that allows them to apply a solution procedure. People's ability to get the solution in these cases is far more general than an ability to process certain types of frequency information. It is an ability to perform elementary logical and set theoretic operations until a nested set structure is represented.

We follow Tversky and Kahneman (1983) in holding that some frequency problems are easy because their nested-set structure is transparent. This explanation has nothing to do with natural frequencies as such. Logical syllogisms can become easy when their set structure is represented clearly with Euler circles. (Some psychologists have even proposed that ordinary people naturally use mental Euler circles for easy syllogisms and try to make all syllogisms easy by attempting to represent them in this way; see Evans, Newstead, & Byrne, 1993.) As we shall see, people find some single-case probability problems easy as well when their set structures are made transparent. Elementary relations between and operations on finite sets are as mathematically simple as any relations and operations. (The exception is taking the set of all subsets of a set, which rapidly increases complexity.) These set relations and operations are the very basis of elementary logic and arithmetic. They also lie at the heart of many practical problems and not just those about frequencies, as we illustrated in 4 above.

What does the natural frequency hypothesis actually predict regarding the medical diagnosis problem? Clearly, it predicts relatively few correct responses under a single-event probability format. Gigerenzer and Hoffrage (1995) also say "Relative [as opposed to natural] frequency formats elicit the same (small) proportion of Bayesian algorithms as probability formats" (p. 692). Assuming that most (if not all) correct responses can be reduced to "Bayesian algorithms," then this hypothesis would not predict facilitation even for Cosmides and Tooby's (1996) version shown above, because it uses relative frequencies that are normalized on 1000. This version cannot be put into what Gigerenzer and Hoffrage call the "standard menu" of a natural frequency problem. To do so, the problem would have to state, not only that 1 person in 1000 people has the disease and a positive test, but also that, out of the 999 people without the disease, 49.95 (5% of 999) had a positive test as well. 49.95 can hardly be called a natural frequency. One could easily infer from this that the exact answer is 1 out of 50.95, but not because of natural frequencies per se. Gigerenzer and Hoffrage use an example of an illiterate physician to try to show how easy the natural sampling of a natural frequency would be under primitive conditions, but it is clear that this physician could never come up with frequencies with fractional values. Gigerenzer and Hoffrage do also describe what they call a "short menu" for the natural sampling of a natural frequency. Using that we could ask about a natural sample of 100 people with the disease and a positive test and 5095 with a positive test, but of course that would be a different problem than the one presented. Furthermore, an actual illiterate physician would have a very hard time recalling 100 out of 5095 cases and could not read a word problem about these cases (in the unlikely event that words even exist in her language for 100 and 5095). Gigerenzer

and Hoffrage would seem to be correct in that literate participants in our society would find this word problem utterly trivial. They would only need to grasp that the 100 people with the disease and the positive test are a nested subset of the 5095 with the positive test. So the nested-sets hypothesis has the advantage. It explains why it is easy to get approximately the right answer in Cosmides and Tooby's version, as well as easy to get the exact answer in Gigerenzer and Hoffrage's standard menu problems, when these can be written down, and trivial to get the right answer in their short menu problems.

Gigerenzer and Hoffrage's (1995) account of word problem facilitation shares core features with the nested-sets hypothesis. They use tree diagrams that are essentially a method like Euler circles to bring out set structure. These are 'trees' in the technical set-theoretic sense and also in the logical sense and so the formal equivalent of mental models (Jeffrey, 1981; Johnson-Laird et al., 1999).

In sum, Gigerenzer and Hoffrage (1995, 1999) would appear not to predict relative frequency facilitation in the medical diagnosis problem as reported by Cosmides and Tooby (1996), except inasmuch as their hypothesis is indistinguishable from the nested-sets hypothesis, an hypothesis offered by Tversky and Kahneman in 1983. The first 4 experiments we report examine Cosmides and Tooby's (1996) claims by examining the differences amongst various probability and relative frequency formats of the medical diagnosis problem. For the sake of consistency, we refer to these claims as the natural frequency hypothesis.

### Experiment 1

To test the hypothesis that the cause of facilitation in Cosmides and Tooby's (1996) frequency conditions is that the text of their problems makes the critical nested-set

relations transparent, we used a problem that had all the attributes of theirs except that it concerned single-event probabilities instead of frequencies. Experiment 1 attempts to both replicate their frequency effect and to look for facilitation using the clarified probability problem. The experiment therefore compares 3 problems, one involving unclarified probability, one frequency, and one probability with transparent nested-set relations. The problems all express uncertainty values as fractions (e.g., 1/1000). The natural frequency hypothesis does not state that facilitation is merely a result of simpler arithmetic, so it is important to control the difficulty of the arithmetic demanded by the various problems. A fraction is itself neither a frequency or a probability; it is the interpretation of the value that classifies it. The nested-sets hypothesis predicts that both the frequency and the transparent probability problem should facilitate performance over and above the unclarified probability problem by equal amounts.

### Method

Materials. The probability problem used was similar to Casscells et al.'s (1978) original, except that probabilities were presented in ratio form to ease calculations and the assumption was made explicit that the probability of testing positive is 1 if the individual does have the disease (cf. Cosmides & Tooby, 1996, Experiment 5):

#### Probability problem without clear nested-sets relations:

Consider a test to detect a disease that a given American has a 1/1000 chance of getting. An individual that does not have the disease has a 50/1000 chance of testing positive. An individual who does have the disease will definitely test positive. What is the chance that a person found to have a positive result actually

has the disease, assuming that you know nothing about the person's symptoms or signs? \_\_\_\_\_%

The frequency problem used is shown above, taken verbatim from Cosmides and Tooby.

We also used a probability version from their paper that seemed to make the nested-sets relation just as transparent as the frequency version (Experiment 6, Condition 1).

However, their version, like Casscells et al., said “5% of all people who are...” to describe the false positive rate. To ease calculations and to maintain focus on a single individual, we instead said, “the chance is 50/1000 that someone who is...”:

Probability version with transparent nested-sets relations:

The prevalence of disease X among Americans is 1/1000. A test has been developed to detect when a person has disease X. Every time the test is given to a person who has the disease, the test comes out positive. But sometimes the test also comes out positive when it is given to a person who is completely healthy. Specifically, the chance is 50/1000 that someone who is perfectly healthy would test positive for the disease.

Imagine that we have given the test to a random sample of Americans. They were selected by lottery. Those who conducted the lottery had no information about the health status of any of these people.

What is the chance that a person found to have a positive result actually has the disease? \_\_\_\_\_%

Participants. 25, 45, and 48 Brown University undergraduates ranging in age from 18-24 were tested in the probability, frequency, and probability plus nested sets conditions, respectively.

Procedure. The procedures for this and subsequent studies were identical. Participants were tested following introductory psychology, economics, and cognitive science courses. Questionnaires were distributed containing one problem from one randomly chosen condition in addition to a few unrelated items. Each problem was followed by the question, "How confident are you that your decision is correct?" Responses were collected on a 1 (not at all confident) to 7 (extremely confident) scale. Participants were asked to take their time and hand in their booklet once they were finished. Participants were also asked whether they had encountered a similar problem before. The responses of all those who said "yes" were not considered and are omitted from this report.

### Results and Discussion

All responses between 1.8% and 2.2% (written in any format) were scored as correct. Proportions correct for all 3 conditions are shown in the first row of Table 1. Like previous studies, relatively few participants gave the correct response to the original problem (20%), even though it expressed uncertainties as fractions. Unlike Cosmides and Tooby's (1996) Stanford University students, 72% of whom were correct under a frequency frame, only 51% of our Brown students did. Nevertheless, significantly more gave the correct answer in the frequency than in the probability only conditions,  $\chi^2(1) = 5.25$ ;  $p < .05$ . However, significantly more also gave it in the nested-sets probability condition (48%),  $\chi^2(1) = 4.30$ ;  $p < .05$ , and the difference between the two nested-sets conditions was not significant,  $\chi^2(1) < 1$ . Apparently, the determinant of facilitation is not the use of a frequency format per se, but the use of any format that clarifies how the events relate to one another.

Confidence judgments for correct responses were consistently higher than for incorrect responses in every relevant condition in this paper. Otherwise, no systematic differences in confidence obtained and therefore the data will not be reported.

### Experiment 1B

We were surprised that a smaller proportion of our students were correct in the frequency condition of Experiment 1 than Cosmides and Tooby's (1996) students using the same problem. Both studies were run at highly selective undergraduate institutions. To examine the robustness of this result, we tried to replicate it.

#### Method

A different group of 48 Brown students was tested on the frequency problem of Experiment 1. Otherwise, the methodologies were the same.

#### Results and Discussion

In this experiment, only 31% of participants gave the correct answer (Table 1), not significantly more than in the probability only condition of Experiment 1,  $\chi^2(1) < 1$ . Responses in this condition did differ from all other conditions reported in this paper in that a large proportion of participants (27%) gave a response of 1 out of 1000. Instead of neglecting base rates, these participants relied exclusively on base rates and failed to consider the case data. Evans et al. (2000) and Cosmides and Tooby (1996) report parallel results. This may have occurred because the frequency formulation gives base rates more emphasis than other versions by giving it pride of place at the start of the text: "1 out of every 1000 Americans has disease X." The finding is not highly robust though; only 11% of participants gave this response in the frequency condition of Experiment 1

(only 2 participants did otherwise). In sum, our studies found a smaller and less robust effect of frequency format on performance than reported by Cosmides and Tooby.

### Experiment 1C

One possible explanation for poor performance with the original Casscell's et al. (1978) problem and with our clarified version of it in Experiment 1 is that the problems have no unambiguous probabilistic interpretation. Both problems refer to the probability of an individual "getting" a disease, without specifying a time period over which the disease might be "gotten," so whether it applies to the event at hand is questionable. In contrast, the problems that do show facilitation unambiguously specify the relevant time period as the current moment (e.g., the number of Americans who *have* the disease). Moreover, the problems that do show facilitation specify that the test was given to a random sample of Americans, making the sample data a valid base rate; the problems that don't show facilitation do not clearly specify that the sample is random. So the advantage of the problems with clear nested sets might have little to do with nested sets per se. Facilitation might result from their unambiguous probabilistic interpretation. To examine this possibility, we tested a version of the problem that clarified the two points just mentioned.

We used this opportunity to examine the nested-sets hypothesis directly by asking participants to draw a diagram that reflected how they thought about the problem. The nested-sets hypothesis predicts that those who draw a nested-sets diagram are more likely to produce a correct answer to the probability question than those who don't draw a nested-sets diagram.

### Method

A different group of 28 Brown students was tested on the following problem. It differs from the probability problem without clear nested-sets relations in that it describes the base rate unambiguously and it clarifies that the individual being judged was randomly chosen:

The chance that a randomly chosen person has a particular disease is 1/1000.

Consider such a randomly chosen person. If this individual does not have the disease, the person's chance of testing positive is 50/1000. An individual who does have the disease will definitely test positive. What is the chance that, if a randomly chosen person is found to have a positive result, this person actually has the disease? \_\_\_\_\_%

After answering the question, participants turned the page where they were asked to draw a diagram depicting their thought process. The instructions were:

Without changing your previous answer, please draw a picture in the space below that reflects how you think about the problem on the previous page; any kind of diagram representing how you imagine the relations amongst the critical parts of the problem. You are encouraged to refer back to the problem on the previous page. However, please leave your original answer unchanged.

In other respects, the methodology was identical to previous studies.

### Results and Discussion

In this experiment, 39% of participants gave the correct answer (Table 1), more, but not significantly more than in the probability only condition of Experiment 1,  $\chi^2(1) = 1.51$ , n.s.. This result suggests that some facilitation may arise merely by clarifying the mapping of the terms of the problem into probabilities. Although less facilitation was

observed than was obtained by clarifying the set relations of the problem (probability with clear nested sets condition), this difference was also not significant,  $\chi^2(1) < 1$ .

Therefore, although the possibility remains that nested-sets representations must be made explicit for some people to solve the problem, these data do not directly support such a conclusion.

Two research assistants rated the diagrams that participants drew according to i. Whether the diagrams represented nested-set relations (e.g., a Venn diagram, Euler circles, a hierarchical tree), and ii. Whether the nested-set relations depicted were faithful to the problem. The research assistants agreed on over 95% of cases, the remaining cases were resolved through discussion. The results show a strong correlation between depicting nested-set relations and solving the problem. Of 8 participants who drew an accurate nested-sets diagram, all 8 gave a correct answer. Of the remaining 20, only 3 gave a correct answer,  $z = 4.16$ ;  $p < .0001$ . Results are in the same direction if performance is conditioned on only the first criterion. Together, the results suggest that participants can construct a nested-sets representation if the problem is stated clearly enough and that doing so is strongly correlated with getting the answer right.

## Experiment 2

The nested-sets hypothesis predicts that any manipulation that increases the transparency of the nested-sets relation should increase correct responding. In this experiment, we increased transparency by providing an Euler circle diagram that makes the nested-set relations explicit, a manipulation also used by Cosmides and Tooby (1996, Experiment 4). The same 3 problems as Experiment 1 were used. The presence of the diagram should boost performance in all 3 conditions relative to Experiment 1.

## Method

The method was identical to Experiment 1 except that each problem was presented with the instruction to use the accompanying diagram (see Figure 1) to help reason through the problem. Twenty-five participants were tested with the probability only problem, 38 frequency format, and 48 nested-sets probability.

## Results and Discussion

Proportions of correct responses are shown in the third row of Table 1. The mean proportion was 46% with no significant differences between conditions,  $\chi^2(1) < 1$  for every pairwise comparison. The diagram facilitated performance in the probability only condition relative to Experiment 1 marginally significantly,  $\chi^2(1) = 3.21$ ,  $p = 0.07$ , but had no systematic effect otherwise. Making the nested-set relations transparent facilitated performance if and only if they were not already clear; the diagrams provided no added benefit once the set structure had already been exposed. Similar results are reported by Cosmides and Tooby (1996). In condition 2 of their Experiment 4, they found that 76% of participants were correct, a percentage comparable to that in their standard frequency conditions. In our study, the problems remained difficult for just over half the participants for reasons of calculation or conceptual mapping despite transparent nested-sets structure.

### Experiment 3

Although our transparent probability problem considers a single individual and the probability that that individual has the disease, the problem uses the term “prevalence,” and the phrase “every time,” and asks participants to imagine a random sample of Americans. One might argue that any or all of these aspects of the problem cause people

to represent the problem in terms of frequency, rather than probability. That is, a proponent of the natural frequency hypothesis could claim that, even though Cosmides and Tooby (1996) themselves characterized essentially the same question as a probability problem, it might really induce a frequency set, or induce one in the group of people we tested. To make sure that our results hold even with a problem that is not susceptible to these concerns, we tested a version of a probability problem with transparent nested sets that in every sense concerns the probability that a single individual has the disease, and not the frequency of the disease in a group.

### Method

The method was identical to previous experiments except that a different group of 42 participants was tested on a single problem:

#### Probability version B with transparent nested-sets relations:

The probability that an average American has disease X is 1/1000. A test has been developed to detect if that person has disease X. If the test is given and the person has the disease, the test comes out positive. But the test can come out positive even if the person is completely healthy. Specifically, the chance is 50/1000 that someone who is perfectly healthy would test positive for the disease.

Consider an average American. Assume you know nothing about the health status of this person.

What is the probability that if this person is tested and found to have a positive result, the person would actually have the disease? \_\_\_\_\_

### Results and Discussion

40% of respondents gave the correct response (shown in the fourth row of Table 1). This is not significantly different than the proportion correct in the Probability+nested sets condition (48%;  $z < 1$ ) or the Frequency + nested sets condition (51%;  $z = 1.01$ ; n.s.) of Experiment 1. Thus, this version of the problem, that concerns probability and not frequency in every respect, leads to just as much correct responding as the problem framed in frequency terms merely by stating the problem clearly and making the structure of the situation transparent.

#### Experiment 4

One implication of the nested-sets hypothesis is that a problem should be more difficult if the relevant relations are not nested. If the facilitation observed on this problem is due to transparent nested-sets relations, then a slightly modified problem whose set representation is not nested should show less facilitation under both a probability and a frequency frame. To create such a problem, we changed the false negative rate from 0 to 1/1000. This is equivalent to changing the hit rate from 1 to 999/1000. The critical set relations are no longer nested because the chance of having the disease is not nested within the chance of testing positive; a small chance of having the disease obtains even without testing positive. The nested-sets hypothesis predicts that few participants will arrive at the solution to this problem. Cosmides and Tooby's (1996) hypothesis that frequency formats facilitate judgment predicts that more people should get the Bayesian answer under a frequency than under a probability frame. Of course, as discussed above, Gigerenzer and Hoffrage (1995) predict no facilitation with probabilities or relative frequencies, so they turn out to predict no facilitation on these or any of our problems.

### Method

The method was identical to previous experiments except that two different groups of 33 participants were tested on new problems. One problem, the Probability version with positive false negative rate, was identical to the probability version with transparent nested-sets relations of Experiment 1 except that the sentence “Every time the test is given to a person who has the disease, the test comes out positive.” was replaced with “The test is almost sure to come out positive for a person who has the disease. Specifically, the chance is 999/1000 that someone who has the disease will test positive.” To create a Frequency version with positive false negative rate, the corresponding sentence was replaced with, “The test almost always comes out positive for someone who has the disease. Specifically, out of every 1000 people who have the disease, the test comes out positive 999 times.” The answer to this problem is almost identical to all previous problems, about 2%.

### Results and Discussion

15% of those given the Probability version with positive false negative rate gave the correct response. 21% of those given the Frequency version with positive false negative rate were correct. These do not differ from each other or from the proportions getting the original problem correct in Experiment 1 (all  $z$ 's  $< 1$ ). We conclude that the critical variable for facilitation with the medical diagnosis problem is not whether the problem is framed in terms of frequency rather than probability, but whether the problem can be represented using nested sets and the nested-set relations are made transparent by the statement of the problem.

### Experiment 4B

Our interpretation of Experiment 4 is that the addition of a false negative rate increased the difficulty of the problem by preventing participants from relying on simple inclusion relations to solve it. But an alternative interpretation is that the false negative rate made the arithmetic required to solve the problem more difficult by confusing people with too many numbers. That is, the false negative rate could have interfered with the construction of a response rather than a representation of the problem.

To address this issue, we developed problems that, like Experiment 4, provided additional numerical ratios for participants to consider, but, unlike Experiment 4, those ratios were irrelevant to the problem. If performance was low in Experiment 4 because of the presence of too many numbers, then performance should be equally low in this experiment. In contrast, if performance was low because participants saw that the false negative rates were relevant, incorporated them into their nested-sets representations, but then were unable to pick out the relations necessary to solve the problem, then irrelevant ratios should not affect performance. Participants should simply ignore them and show as much facilitation as in the clear nested-sets problems of Experiment 1.

As we did in Experiment 1C, we used this opportunity to examine the nested-sets hypothesis directly by asking participants to draw a diagram depicting their representation of the problem. The nested-sets hypothesis predicts that those who draw a nested-sets diagram are more likely to produce a correct answer to the probability question than those who don't draw a nested-sets diagram.

#### Method

The method was again identical to previous experiments except that two different groups of 30 were tested on new problems. One problem, the Probability version with irrelevant ratios, was identical to the probability version with transparent nested-sets relations of Experiment 1 except that the irrelevant sentence “The test was done in a modern hospital where people are up and about within 12 hours after surgery at a rate of 999/1000” was appended to the first paragraph. A Frequency version with irrelevant ratios was constructed by appending the sentence “The test was done in a modern hospital where 999 people out of 1000 are up and about within 12 hours after surgery.”

After responding to the problem, participants were asked to draw diagrams using the instructions of Experiment 1C. In other respects, the methodology was identical to previous studies.

### Results and Discussion

23% of participants tested with the probability version and 50% with the frequency version gave the correct answer (Table 1), an almost statistically significant difference,  $\chi^2(1) = 3.52$ ;  $p = .06$ . These results were not predicted by either hypothesis and is the only case we have observed suggesting facilitation with a frequency format over a comparable probability format. The lack of facilitation in the probability version relative to the original problem suggests that the addition of any numbers -- relevant or not -- to the problem confuses people and renders them unable to do the necessary calculations to answer the question correctly. But the facilitation observed in the frequency condition (performance was significantly higher than in the frequency condition of Experiment 4,  $\chi^2(1) = 4.53$ ;  $p < .05$ ) suggests that only relevant numbers inhibited performance, participants were able to screen out the irrelevant ones, and this is consistent with the

hypothesis that low performance in Experiment 4 occurred because of failure to pick out the relevant relations from a complicated representation. The apparent advantage of frequency over probability suggests that a frequency format somehow makes it easier for people to distinguish relevant from irrelevant ratios.

As in Experiment 1C, two research assistants rated the participants' diagrams according to whether they depicted nested-set relations and whether the relations depicted were faithful to the problem. Again, agreement was over 95% and the remaining cases were resolved through discussion. The frequency condition elicited only a few more nested-sets diagrams than the probability condition (14 vs. 12), but those in the frequency condition tended to be more faithful to the problem (10 vs. 5). Like the probability judgments, these results suggest that participants were a bit better able to pick out the relevant information in the frequency than probability condition and also to represent it correctly. The results again show a strong correlation between depicting nested-set relations and solving the problem. Of 15 participants who drew an accurate nested-sets diagram, 11 gave the correct answer. Of the remaining 45 participants, only 11 gave a correct answer,  $z = 3.40$ ;  $p < .001$ . The pattern holds for both probability and frequency conditions and also if performance is conditioned only on whether participants drew a nested-sets diagram.

### Conjunction fallacy

A second cognitive illusion of probability judgment that has been claimed to disappear under a frequency frame (e.g., Gigerenzer, 1994) is the conjunction fallacy of Tversky and Kahneman (1983). Their most famous example begins with the description:

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

and asks participants to judge the relative likelihood of the following two statements:

Linda is a bank teller. (A)

Linda is a bank teller and is active in the feminist movement. (A&B)

Because Statement A is entailed by Statement A&B (i.e., if Linda is a feminist bank teller, then she is a bank teller), the probability of A is necessarily greater than or equal to the probability of A&B (the conjunction rule of probability). In this case, the rule follows from common sense. No commitment to Bayesianism or any other theory of the foundations of probability is required to see that, if one has to make a guess about Linda, choosing the more general statement is sensible. But, as Tversky and Kahneman (1983) showed and many others have replicated (e.g., Bar-Hillel & Neter, 1993; Johnson, Hershey, Meszaros, & Kunreuther, 1993), people fail to do so consistently. Rather, most people judge Linda more likely to be a feminist bank teller because she sounds like a feminist. That is, people reason from the inside, in terms of properties, not in accordance with the fact that the set of bank tellers includes the set of feminist bank tellers.

Evidence that probability judgments can respect nested-set relations (greater probability assigned to more inclusive sets) is provided by Fiedler (1988). He asked people either to evaluate the relative probability that Linda was a bank teller versus a feminist bank teller (he asked them to rank order the statements according to their probability) or about relative frequency ("To how many out of 100 people who are like Linda do the statements apply?"). Fiedler found that 91% of participants violated the

conjunction rule in the probability condition but only 22% violated it in the frequency condition.

Tversky and Kahneman (1983) were the first to show fewer conjunction fallacies when problems are framed in terms of frequency. They interpreted this result in terms of the nested-sets hypothesis: Presenting the options as concrete classes made the inclusion relation between the two sets more transparent. Specifically, probability frames tend to encourage people to think about events in terms of properties; one naturally considers the match between the properties of Linda and the properties of bank tellers. Probability frames elicit a focus on intensional structure, a perspective on events that causes people to rely on the representativeness of outcomes and obscures the set relations between instances. In contrast, frequency frames induce a representation in terms of multiple instances, so that problem solvers can “see” the embedding of one set of instances inside another. When the correct logical relation between categories (all feminist bank tellers are bank tellers) is evident, participants are more likely to generate judgments of probability that respect it. Hence, the incidence of the conjunction fallacy is reduced. Agnoli and Krantz (1989) provide evidence consistent with this interpretation. They found a marked decrease in the incidence of the conjunction fallacy when they used Euler circles to train participants to interpret categorical relations as nested sets. They found that the effect of such training was largely restricted to probability judgment; it had little influence on similarity judgment. For untrained participants, probability and similarity judgments were highly correlated. Participants trained with nested sets showed a much weaker correlation.

Both Fiedler (1988) and Tversky and Kahneman (1983) found significant effects for frequency frames; at most 20-25% of participants violated the conjunction rule in both frequency conditions. However, Tversky and Kahneman found markedly different results than Fiedler in the probability condition: Tversky and Kahneman found that only 65% of participants committed the conjunction fallacy; Fiedler found that 91% did. The difference may be a result of different tasks. Fiedler's probability question asked participants to rank order each alternative. This forced a choice between events by not offering the option to judge both alternatives as equally probable. Without this option, some people may have committed the conjunction fallacy despite the (perfectly reasonable) conviction that the two options were equally likely. Tversky and Kahneman asked for probability ratings -- not rankings -- and obtained far fewer conjunction fallacies, possibly because ratings allowed people to assign equal probabilities to the two events. Whereas the question was held constant in Tversky and Kahneman's study, Fiedler used a ranking question in the probability condition and a rating question for frequency. Hertwig and Chase (1998) showed that ranking does lead to more conjunction fallacies than rating. The confound can also be found in Hertwig and Gigerenzer (1999) who extensively discuss the issue. They attribute about half of the difference between their probability and frequency conditions to the difference between rating and ranking.

### Experiment 5

Experiment 3 was designed to distinguish the natural frequency and nested-sets hypotheses by determining whether there would be an effect of frequency in the absence of a transparent nested-set relation. In contrast to the natural frequency hypothesis, we predicted that a frequency format would not reduce the conjunction fallacy when the

relation between the critical statements was opaque. We hid the nested-sets relation by inserting statements between the conjunction and its constituent. We expected that spacing the critical statements would obscure their inclusion relation so that they would be judged independently. We attempted to reconcile Tversky and Kahneman's (1983) and Fiedler's (1988) findings by using both ranking and rating tasks. We expected that ranking would lead to more conjunction errors, because it forces a choice between the critical statements. This effect could be attenuated because ranking also forces a comparison of statements, a process that could increase the likelihood of discovering the critical set inclusion relation. However, ratings are usually also made in relation to other statements and so could involve just as much comparison, at least when statements with similar terms are in close proximity. In sum, we predict that performance in the probability and frequency frames should not differ for either task.

### Method

Between 31 and 50 students returned their questionnaires in each condition. Two problems were tested: "Linda" and "Bill the accountant" (see Tversky & Kahneman, 1982). All conditions in the Linda problem used the description above. All participants were asked to rate or rank-order the following set:

is a teacher

is a bank teller (A)

reads for a hobby

enjoys attending parties

is a doctor who has two children

is a psychologist

smokes cigarettes

is an accountant who writes for a hobby

is a lawyer

is a bank teller who is active in the feminist movement (A&B)

rollerblades for a hobby

where the critical statements were separated by seven unrelated items. The format of the Bill problem was identical. The order of the critical statements was counterbalanced.

Questions were asked in both probability and frequency frames. In the probability version, participants were asked to either "Please rank order the following items from most to least likely" (Ranking), or "Using the information above, estimate the probability that:" (Rating). Frequency frames for Linda were: "Suppose there are 1,000 women who fit this description. Please rank order the following items from most to least frequent in this set of 1,000 women:" (Ranking); or "Suppose there are 1,000 women who fit this description. Estimate how many of them:" (Rating). The tasks for the Bill problem were identical except that "women" was changed to "men."

### Results and Discussion

We eliminated all respondents who had heard of the problem previously, gave the same answer for every option, or failed to follow the instructions. This left between 22 and 42 participants per condition. Table 2 shows the percentages of correct responses from the remaining participants for frequency and probability judgments in the rating and ranking conditions for each problem as well as percentages of ratings that assigned equal values to conjunction and constituent. As predicted, probability and frequency frames did not differ in either the rating or ranking conditions for either problem ( $\chi^2(1) < 1$  for

every comparison). These results suggest that frequency frames do not improve performance over probability frames unless the nested-sets structure of the problem is exposed. By making it opaque in the frequency condition, we eliminated the enhancement typically seen in relation to single-event probability judgments. Notice that more than 2/3 of ranking participants (80% for the Bill problem) produced conjunction fallacies, even with a frequency format.

Conjunction fallacies were less common when participants were not forced to rank order statements. Rating and ranking differed significantly for the Linda problem,  $\chi^2(1, N = 47) = 6.14, p < .05$ , for probability, and  $\chi^2(1, N = 68) = 4.10, p < .05$ , for frequency, and marginally significantly for the probability Bill problem,  $\chi^2(1, N = 51) = 3.53, p = .06$  and significantly for frequency  $\chi^2(1, N = 53) = 5.03, p < .05$ . Statements can be rated individually but ranking requires comparison among statements. To the extent that perceiving the nested-set relation between Statements A and A&B requires comparison, one might expect fewer conjunction fallacies with the ranking than with the rating task. Apparently, if there is any such effect, it is overwhelmed by the availability of response options in the rating task.

As shown in the third column of Table 2, a substantial number of participants chose to assign equal values to the conjunction and its constituent when rating. These responses largely account for differences between rating and ranking tasks. Hertwig and Chase (1998) argue that the difference between the tasks emanates from different strategies: They claim that people rate by making independent evaluations of evidential support and integrate across cues using rules. In contrast, people rank by making pairwise comparisons of evidential support using single cues. Such a theory cannot be

ruled out by these data. Our suggestion is simpler, but the important point here is that ranking produces more conjunction fallacies than rating.

### General Discussion

The studies reported support the nested-sets hypothesis over the natural frequency hypothesis. They found that, first, the facilitation observed with a frequency frame on the medical diagnosis task was also induced by a clearly stated single-event probability frame (Experiments 1, 2, and 3). Second, comparable facilitation was obtained with a diagram that clarified the critical nested-set relations of the problem (Experiment 2). Third, the effect of frequency framing was not robust (Experiment 1B). It was eliminated by adding complexity to the problem (Experiment 4), though not completely when the complexity was irrelevant (Experiment 4B). Fourth, individuals who drew an appropriate nested-sets diagram were more likely to answer the probability or frequency question correctly than those who didn't (Experiments 1C and 4B). Fifth, the effect of a frequency over a probability frame on the conjunction fallacy, as reported by Fiedler (1988) and Hertwig and Gigerenzer (1999), was exaggerated by a confound with rating versus ranking. Finally, the benefit of frequency in reducing the conjunction fallacy was eliminated by making the critical nested-set relation opaque by spacing the conjunction and its constituent.

Taken together, the data suggest that facilitation in probabilistic responding on word problems is not a direct consequence of presenting problems in a frequency format, but rather of making the probabilistic interpretation of the problem unambiguous and of making the set structure of the problem apparent by inducing participants to represent it in a way that highlights how relevant instances relate to one another. Our claim is

decidedly not that frequency formats never help to make nested-set relations clear. Indeed, they can very effectively by cueing a representation that makes subset relations transparent. Moreover, we found facilitation from a frequency format in Experiment 4B. But clarifying nested-sets relations is not a panacea. The lesson of Experiment 4 is that transparent set structure is useful only when the critical relations are simple enough to be extracted and used. More importantly, not all situations can meaningfully be construed in terms of multiple events. For instance, the probability of a defendant's guilt cannot meaningfully be framed in terms of frequency.

Probability theory associates a measure of uncertainty with a set. Any manipulation that causes a correct representation of set structure has the potential to improve probability judgments. Frequency formats can sometimes do that but they are neither necessary, because extensional set-theoretic relations can be made transparent in other ways (Experiments 1, 2, and 3; Ajzen, 1977; Evans et al., 2000; Mellers & McGraw, 1999), nor sufficient, because other criteria must be satisfied to ensure a veridical representation, like representative sampling, unambiguous interpretation of variables, and accurate working and long-term memory retrieval (Experiments 4 and 5, Bar-Hillel & Neter, 1993; Gigerenzer, Hoffrage, & Kleinbölting, 1991; Gluck & Bower, 1988; Lewis & Keren, 1999; Tversky & Kahneman, 1974). Hence, the success of some frequency formats provides no support for the psychological reality of a cognitive module or algorithm specially tuned to frequencies. Perhaps the strongest evidence for this point is a thought experiment: If the Linda problem were given along with Euler circles that exposed its set structure, hardly anyone would commit the conjunction fallacy even in a single-event probability context (we found that only 2 of 39 people did).

The nested-sets hypothesis is the general claim that making nested-set relations transparent will increase the coherence of probability judgment. Making set relations transparent is not the same as inducing a frequency frame, though frequency frames can clarify set relations. Transparency may be a usual consequence of representing a problem's structure in terms of multiple instances, but set relations can represent possibilities even for a single instance. Nested-set relations are more general than frequency representations. For one, they may be purely ordinal. This is most obvious with the Linda problem. The nested sets need only represent the subset relation between the conjunction and constituent, not the actual frequency of each. Also, nested-set representations can be normalized, revealing such properties as independence.

A complete understanding of how people made their judgments in our experiments requires knowledge of the task that participants thought they were performing. People may not always have been judging the axiomatized sense of probability, but might have construed the problems otherwise, perhaps they thought they were judging the strength of evidence in favor of an hypothesis (cf. Briggs & Krantz, 1992). Judgments of evidential strength, and many other measures, need not satisfy the axioms of probability. The normative issue of whether or not participants' judgments should have conformed to the axioms of probability is an interesting one, but not really relevant here (see Vranas, 2000, for an enlightening normative analysis). A question that might be more pertinent is whether our manipulations changed the task that participants assigned themselves. In particular, manipulations that facilitate performance may operate by replacing a non-extensional task interpretation, like evidence strength, with an extensional one (Hertwig & Gigerenzer, 1999). Note that such a construal of the effects we have shown just

reframes the questions that our studies address: Under what conditions are people's judgments extensional, and what aspects of their judgments correspond to the precepts of probability theory?

The natural frequency via natural sampling hypothesis suffers from several inadequacies (Over, 2000a,b; Sloman & Over, in press). Suffice it to say here that proponents of the natural frequency hypothesis claim that their theory can be grounded in evolutionary theory. This claim has two components. One is that people have the ability to solve word problems in certain formats because the natural sampling of a natural frequency was adaptive under primitive conditions. The second is that people do not have the ability to understand single-case probability problems because single-case judgments would not have been adaptive under primitive conditions. In our view, solving the computationally easy word problems depends on a much more general ability to understand elementary logical and set operations and relations. Why this general ability evolved by natural selection is a question that we cannot pursue here (but see Over, in press, and Stanovich & West, in press). We also think that the claim about single-case probability is less credible than its antithesis, an evolutionary story of why people should be good at making some single-case judgments. Most events that people have to think about are one-offs. Different battles, interpersonal relationships, intellectual enterprises, and many other endeavors have unique features that may be relevant to judgments made about them. Human environments produce varied, complex interactions that we must respond to intelligently to survive. It could not have served primitive people's reproductive success to have relied naively only on counts of previous experiences that they recalled, perhaps inaccurately, to deal with such novelty and complexity. They also

had to rely on their theories and causal and explanatory models of previous experience. Such theories and models would have allowed them to make judgments and predictions given their understanding of the unfolding structure of a situation, rather than solely on the basis of their limited memories of possibly biased sample frequencies. In this way, they could come to reliable degrees of belief about uncertain singular propositions. This was as much true for our evolutionary ancestors when they confronted challenging new environments as it is for us today.

## References

- Ajzen, I. (1977). Intuitive theories of events and the effects of base-rate information on prediction. Journal of Personality and Social Psychology, *35*, 303-314.
- Ayton, P. & Wright, G. (1994). Subjective probability: What should we believe? In G. Wright and P. Ayton (Eds.) Subjective probability (pp. 163-183). John Wiley & Sons: Chichester, UK.
- Bar-Hillel, M. & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. Journal of Personality and Social Psychology, *65*, 1119-1131.
- Brase, G.L., Cosmides, L. & Tooby, J. (1998). Individuals, Counting , and statistical Inference: The role of Frequency and whole-object representations in judgment under uncertainty. Journal of Experimental Psychology: General, *127*, 3-21.
- Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in Probability and Frequency Judgments: A Critical Examination. Organizational Behavior and Human Decision Processes, *65*, 212-219.
- Briggs, L. & Krantz, D. (1992). Judging the strength of designated evidence. Journal of Behavioral Decision Making, *5*, 77-106.
- Casscells, W., Schoenberger, A., & Grayboys, T. (1978). Interpretation by physicians of clinical laboratory results. New England Journal of Medicine, *299*, 999-1000.
- Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. Cognition, *58*, 1-73.

- Evans, J. St. B. T., Handley, S. H., Perham, N., Over, D. E., & Thompson, V. A. (2000). Frequency versus probability formats in statistical word problems. Cognition, *77*, 197-213.
- Evans, J.St.B.T., Newstead, S.E., & Byrne, R.M.J. (1993). Human reasoning: The psychology of deduction. Hove, UK: Lawrence Erlbaum Associates Ltd.
- Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright and P. Ayton (Eds.) Subjective probability (pp. 129-162). John Wiley & Sons: Chichester, UK.
- Gigerenzer, G. (1998). Ecological intelligence: An adaptation for frequencies. In D. Dellarosa Cummins & C. Allen (Eds.), The evolution of mind. New York: Oxford University Press.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. Psychological Review, *102*, 684-704.
- Gigerenzer, G., & Hoffrage, U. (1999). Overcoming difficulties in Bayesian reasoning: A reply to Lewis and Keren (1999) and Mellers and McGraw (1999). Psychological Review, *106*, 425-430.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. Psychological Review, *98*, 506 - 528.
- Giroto, V., & Gonzalez, M. (2001). Solving probabilistic and statistical problems: A matter of information structure and question form. Cognition, *78*, 247-276.
- Griffin, D., & Buehler, R. (1999). Frequency, probability, and prediction: easy solutions to cognitive illusions? Cognitive Psychology, *38*, 48-78.

- Hertwig, R. & Chase, V. M. (1998). Many reasons or just one: how response mode affects reasoning in the conjunction problem, Thinking and Reasoning, 4, 319-352.
- Hertwig, R. & Gigerenzer, G. (1999). The “conjunction fallacy” revisited: how intelligent inferences look like reasoning errors. Journal of Behavioral Decision Making, 12, 275-305.
- Jeffrey, R.C. (1981). Formal Logic: Its scope and Limits (2<sup>nd</sup> edition). New York: McGraw-Hill.
- Johnson, E.J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. Journal of Risk and Uncertainty, 7, 35-51.
- Johnson-Laird, P.N., Legrenzi, P., Girotto, V., Legrenzi, M., & Caverni, J-P. (1999). Naive probability: a mental model theory of extensional reasoning. Psychological Review, 106, 62-88.
- Juslin, P., Winman, A., & Olsson, H. (2000). Naive empiricism and dogmatism in confidence research: A critical examination of the hard—easy effect. Psychological Review, 107, 384-396.
- Kahneman, D., Slovic, P., & Tversky, A. (1983) (Eds.), Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 509-520). Cambridge: Cambridge University Press.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. Psychological Review, 103, 3, 582-591.

- Agnoli, F. & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. Cognitive Psychology, *21*, 515-550.
- Lewis, C., & Keren, G. (1999). On the difficulties underlying Bayesian reasoning: a comment on Gigerenzer and Hoffrage (1995). Psychological Review, *106*, 2, 411-416.
- Mellers, B.A., & McGraw, P.A. (1999). How to improve Bayesian reasoning: Comment on Gigerenzer and Hoffrage (1995). Psychological Review, *106*, 2, 417-424.
- Over, D.E. (2000a). Ecological rationality and its heuristics. Thinking and Reasoning, *6*, 182-192.
- Over, D.E. (2000b). Ecological issues: A Reply to Todd, Fiddick, & Krause. Thinking and Reasoning, *6*, 385-388.
- Over, D.E. (in press). From massive modularity to metarepresentation: The evolution of higher cognition. In D.E. Over (Ed.) Evolution and the psychology of thinking: The debate. Hove, UK: Psychology Press.
- Sloman, S., & Over, D.E. (in press). Probability judgment from the inside out. In D.E. Over (Ed.) Evolution and the psychology of thinking: The debate. Hove, UK: Psychology Press.
- Stanovich, K.E., & West, R.F. (in press). Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality. In D.E. Over (Ed.) Evolution and the psychology of thinking: The debate. Hove, UK: Psychology Press.
- Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

- Tversky, A. & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.) Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. Psychological Review, 90, 293-315.
- Vranas, P. B. M. (2000). Gigerenzer's normative critique of Kahneman and Tversky, Cognition, 76, 179-193.

### Acknowledgements

We would like to thank David Krantz and Ralph Hertwig for helpful comments and Ian Lyons, Peter Desrochers, and Jeff Stibel for helping to run the studies. This work was supported by NIMH grant MH51271-01A3 and NASA grant NCC2-1217.

Table 1. Percentages correct for the medical diagnosis problem in Experiments 1-4  
(sample sizes in parentheses).

	FRAME		
	<u>Probability</u>	<u>Frequency</u>	<u>Probability</u>
	<u>only</u>	<u>+ nested sets</u>	<u>+ nested sets</u>
Experiment 1	20 (25)	51 (45)	48 (48)
Experiment 1B		31 (48)	
Experiment 1C	39 (28)		
Experiment 2	48 (25)	45 (38)	46 (48)
(with Venn diagram)			
Experiment 3			40 (42)
(pure probability)			
Experiment 4		21 (33)	15 (33)
(with false negatives)			
Experiment 4B		50 (30)	23 (30)
(irrelevant ratios)			

Table 2

Percentages of correct responses for frequency and probability frames in the rating and ranking conditions of Experiment 5 and percentages of responses that assigned equal values to conjunction and constituent in rating condition. The equal ratings constitute a subset of the correct responses in the rating condition.

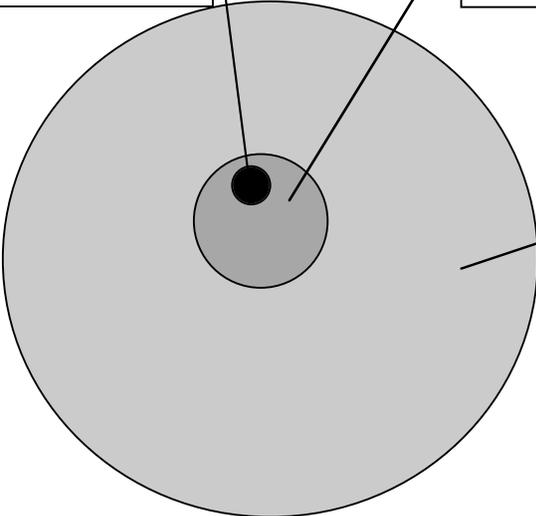
	Ranking	Rating	Ratings that were equal
Linda: Probability	25.0	65.2	34.8
Linda: Frequency	30.1	66.7	19.0
Bill: Probability	17.2	45.5	18.2
Bill: Frequency	20.1	54.2	29.2

Figure legend

Figure 1. Euler circles used in Experiment 2. **Bold text** appeared in probability conditions, *italicized text* in frequency condition.

**Chance that a person has the disease**  
*Number of people who have the disease*

**Chance that a person tests positive for the disease**  
*Number of people that test positive for the disease*



**All Possibilities**  
*All People*