Frequency vs. Probability Formats: Framing the Three Doors Problem

Eric Aaron (aaron@cs.cornell.edu)
Department of Computer Science; Cornell University
Ithaca, NY 14853

Michael Spivey-Knowlton (spivey-knowlton@cornell.edu)
Department of Psychology; Cornell University
Ithaca, NY 14853

Abstract

Instead of subscribing to the view that people are unable to perform Bayesian probabilistic inference, recent research suggests that the algorithms people naturally use to perform Bayesian inference are better adapted for information presented in a natural frequency format than in the common probability format. We tested this hypothesis on the notoriously difficult three doors problem, inducing subjects to consider the likelihoods involved in terms of natural frequencies or in terms of probabilities. We then examined their ability to perform the mathematics underlying the problem, a stronger indication of Bayesian inferential performance than merely whether they gave the correct answer to the problem. With a robustness that may surprise people unfamiliar with the effects of information formats, the natural frequency group demonstrated dramatically greater normative mathematical performance than the probability group. This supports the importance of information formats in a more complex context than in previous studies.

Introduction

Undeniably, reasoning with probabilities can be difficult for people. This is not because such reasoning is unstudied or impossible; there is a well-established mathematics of correct probabilistic reasoning, and we use the term *Bayesian inference* (named after Thomas Bayes (1702(?)-1761)) to refer to a normative inference that agrees with this framework. Nonetheless, cognitive science is rich with demonstrations of people's failures to reason according to Bayesian norms when presented with many kinds of non-trivial probabilistic reasoning problems. In an influential summation, Kahneman & Tversky (1972) take a pessimistic stance, going so far as to conclude that "In his evaluation of evidence, man is apparently not a conservative Bayesian: he is not a Bayesian at all."

Recently, however, Gigerenzer (in press; Gigerenzer & Hoffrage, 1995; Hoffrage & Gigerenzer. 1996) has suggested an explanation for human performance on these tasks without claiming that people lack the ability to function as Bayesian agents. People do have methods or *algorithms* for reasoning about probabilities, but as humans evolved over the ages, the algorithms for Bayesian reasoning were not exposed to information expressed as probabilities. Instead, people gathered information as it came to them, one event at a time and not with the collective information about a set of events that probabilities would give. Thus, Gigerenzer suggests in his framework of *Ecological Intelligence* that people's Bayesian algorithms are adapted for *natural frequencies* (e.g., "out of 160 coin tosses, 80 landed heads") as opposed to the probabilities (e.g., "50% of the coin tosses landed heads") in which information is traditionally presented in studies that produce anti-normative evidence. This difference between *information formats* may not seem dramatic, but in some contexts it can have important effects.

An example taken from Gigerenzer (in press) demonstrates the power of presenting information in a natural frequency format as opposed to a probability format. Consider a physician who just discovered that a symptom-free woman between 40 and 50 years old has had a positive mammogram in a routine breast cancer screening. He needs to advise his patient about the bad news and what to do next, and a first step is estimating the likelihood of the patient actually having breast cancer. Fortunately, he knows all the relevant information (presented here as in Gigerenzer (in press)):

> The probability that a woman has breast cancer is 1% if she is in the same risk group as this patient.
> If a woman has breast cancer, the probability is 80% that she will have a positive mammogram.
> If a woman does <u>not</u> have breast cancer, the probability is 10% that she will still have a positive mammogram.

He can then ask himself:

> Imagine a woman (aged 40 to 50, no symptoms) who has a positive mammogram in your breast cancer screening. What is the probability that she actually has breast cancer? _____%

Unfortunately, a study of 24 physicians done by Gigerenzer & Hoffrage (Gigerenzer, in press; Hoffrage & Gigerenzer, 1996) showed that, under these conditions, physicians frequently mis-estimate the probability of a patient having cancer by nearly a full order of magnitude. The median estimate of actual breast cancer after a positive mammogram was 70%. However, the correct Bayesian estimate is 7.7%, and only 2 out of the 24 physicians (8%) gave that response.

In the same study, 24 other physicians were given the same estimation task, but they were given the information in a frequency format, as shown here:

Ten out of every 1,000 women have breast cancer.

Of these 10 women with breast cancer, 8 will have a positive mammogram.

Of the remaining 990 women without breast cancer, 99 will still have a positive mammogram.

Imagine a sample of women (aged 40 to 50, no symptoms) who have positive mammograms in your breast cancer screening. How many of these women do actually have breast cancer? _____ out of _____

With this frequency format presentation, 11 out of 24 physicians (46%) gave the correct estimate, a dramatic improvement over the 8% of physicians in the probability format case. It does seem that information format can affect doctors' ability to perform Bayesian inference about issues of great importance to themselves and their patients.

In general, cognitive algorithms for Bayesian probabilistic reasoning may be tuned for a natural frequency format rather than the more common probability format. Several studies such as the one presented above seem to support this hypothesis (Gigerenzer, in press). We further tested it by considering a difficult puzzle, the notorious three doors problem, a task not necessarily native to any particular profession. In that context, we investigated whether a difference in information format affected the ability of subjects to perform a correct analysis of the likelihoods underlying the problem. Could this seemingly subtle change actually make a significant difference on this popular brainteaser?

The Three Doors Problem

The three doors problem (also called the Monty Hall problem) (Granberg & Brown, 1995; Selvin, 1975a; Selvin 1975b) has been known as a difficult task since its introduction as the mathematically equivalent three prisoners problem (Gardner, 1959a; Gardner 1959b), a task on which people typically fail to behave as normative Bayesians. In fact, people choose the incorrect answer to the problem so frequently that it has been used as a scenario in which one can study regret over making a losing decision (Gilovich, Medvec & Chen 1995). It recently enjoyed a resurgence in popular interest due to a series of columns in Parade Magazine (vos Savant, 1990a, 1990b, 1991a, 1991b) and related stories in publications including the New York Times (Tierney, 1991). It can be concisely presented as a multi-part game involving a player, a host, and three rooms behind closed doors. In one room is a valuable prize, a car; in the other two, something nearly valueless, a penny. In part 1 of the game, the player selects a door, which stays closed. In part 2, the host opens one of the other two doors to reveal a penny behind it. In part 3, the player is offered the chance to stay with the initially chosen door or switch to the remaining unopened door, and the player keeps whatever is behind the door finally selected in part 3. The problem: In part 3, should the player stay or switch?

It has been consistently presumed that people will use the following two-stage analysis: When the player makes the choice in stage one, he has a 33% chance of picking the car; then, in stage two, after seeing the open door and penny behind it, the two remaining doors each have a 50% chance of hiding the car. People generally choose to stay with their initial choice under these circumstances (Granberg & Brown, 1995) but, no matter how counterintuitive it may seem, the player should switch. When two doors remain unopened, the initially chosen room has a 33% chance of containing the car, and the other unopened room therefore has a 67% chance. One way to understand this is that the likelihood that the car is behind the initially chosen door is not affected by opening another door, it remains 33%. Since the total probability that the car is behind one of the doors remains 100%, the probability associated with the other closed door is 67%. Other more detailed and varied solutions to this problem are available in many forms, including a calculation from Bayes' Theorem. Granberg & Brown (1995) provide an excellent and informative starting point for more information on the problem.

Typically, in presentations of the three doors problem, the facts necessary for the mathematics involved are presented in a probability format. People rarely give the correct answer and, one would presume, even more rarely are able to arrive at a correct Bayesian analysis of the game. Here, we explore whether Gigerenzer's hypothesis that Bayesian reasoning algorithms are better adapted for natural frequencies than probabilities holds for the three doors problem.

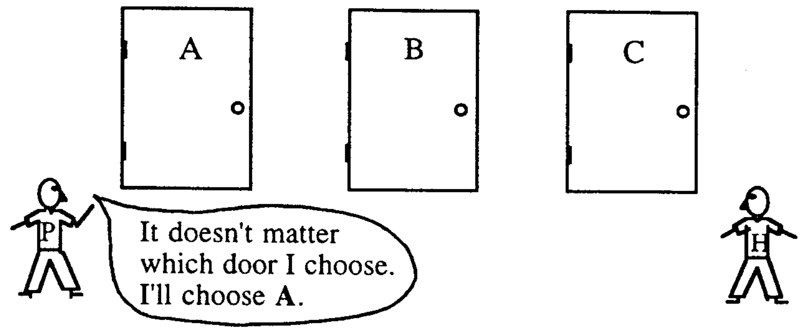
Method

We ran a series of three closely related studies to examine the effect of information format on subjects' ability to correctly perform the mathematics underlying the three doors problem. We focused not only on subjects' answers to the stay/switch question but also on whether inducing subjects to reason with natural frequency formats rather than probability formats improved their normative mathematical performance on the way toward the final stay/switch answer. Every participant in every study was read the same description of the three doors game (corresponding to and accompanied by Figure 1), designed to avoid probability or frequency specific terminology as much as possible[1] and present the game as unambiguously as possible. We stated explicitly, for instance, that in every run of the game, the host opens a door to a room containing a penny and then gives the player a chance to switch, and that the host's choices, within the rules of the game, are made fairly when applicable. Omissions of such details as these can lead to questions extraneous to our study (see, e.g., (Falk, 1992) and (Granberg & Brown, 1995)). In general, we made an effort to maximize the number of subjects who understood the game and its rules without giving information that was not obvious from the traditional description. If subjects failed to understand the rules and workings of the game, they would certainly be

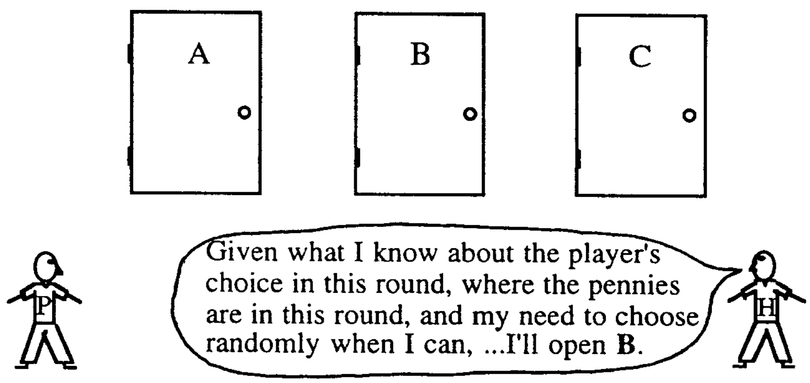
[1] Our presentation agrees with the traditional one in using a single-game (single-event) description in matters such as likelihoods of the initial placement of the car. While this may tend to bias subjects against a natural frequency interpretation (Gigerenzer & Hoffrage, 1995), it affects all subjects equally and should not artificially enhance any advantageous effects a frequency format might have.

The 3-DOORS GAME

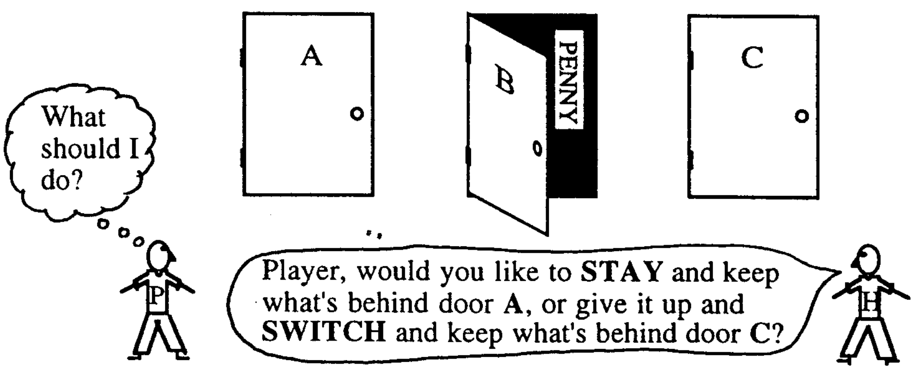
- A car has been placed randomly in one of three rooms and will not be moved during this round of the game. The other two rooms each contain only a penny.
- All doors are closed.
- The round begins with the player (P) choosing a door.



- The host (H) must open a door to show the player a room with a penny in it.
- In doing this, he must: 1) Not open the door the player chose.
  2) Choose randomly when, after these constraints, he still has a choice.



- The host gives the player a chance to switch.



- After the player decides to STAY or SWITCH, the round ends, and the player keeps whatever is in the room he chose.

Figure 1. Diagram accompanying spoken explanation of the three doors game.

unable to perform the desired Bayesian analysis accurately in this experimental context, regardless of whether information is presented to them in natural frequency or probability format. We were not concerned that our detailed description of the game might result in a slightly higher rate of correct answers to the stay/switch question than the traditional presentation, because our focus was on the mathematics and not simply the stay/switch answers, and our comparisons were between the two subject groups (i.e., those given frequency format questions and those given probability format questions).

After hearing the description of the game, participants answered questions on written questionnaires. There were three different questionnaire types, each determining a different study. For each of the three types, there were frequency and probability versions, designed to induce subjects to consider the game and related information in terms of either natural frequencies or probabilities. (Figure 2 shows the frequency format questionnaire for Experiment 2; its correct answers are given in the Appendix. Figure 3 shows the probability format questionnaire for Experiment 2.) Every questionnaire had certain common features, such as an introduction, the stay/switch question, a number of questions phrased symmetrically (i.e., if a question is asked about one door, it is asked about all three) with the correct answers given to at least the first of the questions, and a question asking subjects if they were previously familiar with the three doors problem. For our results, we considered only subjects for whom the problem was novel.

Jumping to the expected interpretation, that when only two doors remain closed in the game the chances of the prize being behind either one are equal, does not require any deep Bayesian analysis of the sort that a frequency/probability format variation is likely to influence. Therefore, we needed to induce subjects to actually compute the likelihoods after the host opens a door from the likelihoods beforehand, and asking a series of computation-based questions leading to the stay/switch question was our device for eliciting this analysis. To promote frequency format analyses, we provided an introductory section and questions like those shown in Figure 2. Similarly, to promote probability format analyses from subjects, the probability versions contained a corresponding introduction and questions like those shown in Figure 3. Within each study, questionnaires were matched so that the questions on the two versions were in minimal pairs; as much as possible, corresponding questions asked for the same information (or clearly isomorphic information) to try to ensure that the questions themselves would not predispose one group to more correct answers than the other group. In general, differences between the frequency and probability versions were minimized, except for the differing information formats. In particular, most of the introduction and the final stay/switch question were identical on all questionnaires, across experiments and information format versions.

On each questionnaire, there were three questions immediately before the stay/switch question that were about the likelihoods of potential car placement after the host opens a door to a car-less room; we consider these the "math" questions for determining whether subjects performed the mathematical analysis correctly. These questions were identical across experiments. In this paper, we give the percentages of subjects who answered the stay/switch question correctly, regardless of their performance on the math questions, and of subjects who answered all the math questions correctly, regardless of their performance on the stay/switch question. (Only one person answered all the math questions correctly and missed the stay/switch question.) In all three experiments, participation was during class time and no further incentive to participate was given.

Experiment 1

The first experiment attempted to elicit analysis without too much coaching, using 8 questions on each questionnaire and giving the correct answer to only the first question. This is the presentation we intended to use as our closest analogy to the three doors problem in its traditional form (unaccompanied by outside questions).

Subjects

112 Cornell University students from introductory courses in psychology, cognitive science, German, and computer science were participants. 58 subjects had frequency format questionnaires; 54 subjects had probability format questionnaires.

Results

21% of the subjects in the frequency version gave the correct stay/switch answer, while 18% gave the correct answer in the probability version; $t(110) = .29$, $p>.1$. On the math questions, 7% of the subjects in the frequency version gave correct responses, while 0% of the subjects in the probability version gave those correct answers; $t(110) = 1.98$, $p < .05$.

Discussion

This questionnaire was constructed for minimal coaching to the correct stay/switch answer and without much help leading to the correct math, lacking three questions present in Experiments 2 and 3. The results show no significant effect of the variation of information format on people's performance on the stay/switch question. However, they show a small but significant effect suggesting that the frequency format facilitates Bayesian performance on the *mathematics leading to the stay/switch response*.

Experiment 2

In Experiment 2, the questionnaires had 11 questions, which explicitly asked for all the components needed for a solution according to Bayes' Theorem, trying to elicit information from subjects without explicitly giving any added facts. The added three questions would, we presume, encourage deeper

3-Doors Problem Questionnaire

Imagine that you've seen 30,000 rounds of the game played in which the player chooses door A in part 1 of the round. We will be asking you questions only about those rounds. We are not trying to trick you with any of the questions, just making sure you understand the game. We will even give you the first answer to get you started.

1) Of these 30,000 rounds in which the player chooses door A in part 1 of the round, in how many is the car actually behind door A? _10,000__

2) Of these 30,000 rounds in which the player chooses door A in part 1 of the round, in how many is the car actually behind door B? _________

3) Of these 30,000 rounds in which the player chooses door A in part 1 of the round, in how many is the car actually behind door C? _________

The next three questions all ask about the host opening door B in part 2 of a round.

4) Of the rounds in the answer to question 1), the rounds in which the player chooses A in part 1 and the car is actually behind door A, in how many of those rounds will the host open door B in part 2 of the round? _________

5) Of the rounds in the answer to question 2), the rounds in which the player chooses A in part 1 and the car is actually behind door B, in how many of those rounds will the host open door B in part 2 of the round? _________

6) Of the rounds in the answer to question 3), the rounds in which the player chooses A in part 1 and the car is actually behind door C, in how many of those rounds will the host open door B in part 2 of the round? _________

Keeping in mind your answers to 4), 5), and 6) above...

7) In how many of these 30,000 rounds of the game in which the player picks door A in part 1 of the round does the host open door B in part 2 of the round? _________

Keeping in mind your previous answers....

8) In how many of those rounds from question 7), the ones in which the player picks A in part 1 of the round and the host picks B in part 2 of the round, is the car actually behind door A? ________

9) In how many of those rounds from question 7), the ones in which the player picks A in part 1 of the round and the host picks B in part 2 of the round, is the car actually behind door B? ________

10) In how many of those rounds from question 7), the ones in which the player picks A in part 1 of the round and the host picks B in part 2 of the round, is the car actually behind door C? ______

Last: Based on your knowledge, in a round of the game in which the player chooses door A in part 1 of the round and the host opens door B in part 2 of the round, what should the player do in part 3 of the round? Should the player STAY with door A or SWITCH to door C? _________

FOR OUR INFORMATION: Had you heard of this game before today? _________

Figure 2. Questionnaire for Experiment 2, Frequency format.

3-Doors Problem Questionnaire

We will be asking you questions only about rounds in which the player chooses door A in part 1 of a round of the game. We are not trying to trick you with any of the questions, just making sure you understand the game. We will even give you the first answer to get you started.

1) In a round in which the player chooses door A in part 1, what is the probability that the car is actually behind door A? _1/3_ (or _33.3%_, whichever you prefer)

2) In a round in which the player chooses door A in part 1, what is the probability that the car is actually behind door B? _________

3) In a round in which the player chooses door A in part 1, what is the probability that the car is actually behind door C? _________

The next three questions all ask about the host opening door B in part 2 of a round.

4) In a round as in question 1), in which the player chooses door A in part 1 and the car is actually behind door A, what is the probability that the host will open door B in part 2 of the round? _________

5) In a round as in question 2), in which the player chooses door A in part 1 and the car is actually behind door B, what is the probability that the host will open door B in part 2 of the round? _________

6) In a round as in question 3), in which the player chooses door A in part 1 and the car is actually behind door C, what is the probability that the host will open door B in part 2 of the round? _________

Keeping in mind your answers to 4), 5), and 6) above...

7) In a round of the game in which the player picks door A in part 1 of the round, what is the probability that the host will open door B in part 2 of the round? _________

Keeping in mind your previous answers....

8) In a round as in question 7), a round in which the player picks A in part 1 of the round and the host picks B in part 2 of the round, what is the probability that the car is actually behind door A? ________

9) In a round as in question 7), a round in which the player picks A in part 1 of the round and the host picks B in part 2 of the round, what is the probability that the car is actually behind door B? ________

10) In a round as in question 7), a round in which the player picks A in part 1 of the round and the host picks B in part 2 of the round, what is the probability that the car is actually behind door C? ______

Last: Based on your knowledge, in a round of the game in which the player chooses door A in part 1 of the round and the host opens door B in part 2 of the round, what should the player do in part 3 of the round? Should the player STAY with door A or SWITCH to door C? _________

FOR OUR INFORMATION: Had you heard of this game before today? _________

Figure 3. Questionnaire for Experiment 2, Probability format.

analysis than the Experiment 1 questionnaire. As in Experiment 1, we gave the correct answer to only the first question.

Subjects

68 Cornell students from introductory courses in psychology, cognitive science, and German were participants. 34 had frequency format questionnaires, and 34 had probability format questionnaires.

Results

29% of the subjects in the frequency version gave the correct stay/switch answer, while 12% gave the correct answer in the probability version; $t(66) = 1.82$, $p<.1$. Considering only the math questions, 21% of the subjects in the frequency version gave correct responses, while 0% of the subjects in the probability version responded correctly; $t(66)=2.93$, $p<.01$.

Discussion

We saw a marginally significant effect of information format on people’s performance when considering only the stay/switch question. However, we saw a robust effect on performance on the math questions, supporting the notion that frequency format facilitates Bayesian inference when figuring out the game's underlying mathematics.

Experiment 3

For Experiment 3, we used the same 11 questions as in Experiment 2, but we gave the correct answers to the first 6 — nearly all but the math and stay/switch questions. Our goal was to inhibit any floor effect or paralysis over math discomfort by giving subjects a head start and providing evidence, upon their self-checking, that they either successfully understood the game or did not understand it and needed to analyze it further.

Subjects

75 Cornell students from introductory courses in psychology, cognitive science, and German were participants. 38 had frequency format questionnaires, and 37 had probability format questionnaires.

Results

37% of the subjects in the frequency version gave the correct stay/switch answer, while 27% gave the correct answer in the probability version; $t(73) = .9$, $p>.1$. On the math questions, 26% of the subjects in the frequency version gave correct responses, while 0% of the subjects in the probability version responded correctly; $t(73)=3.59$, $p<.001$.

Discussion

We found a similar effect to that in Experiment 2. There was some concern unique to Experiment 3, however, that some subjects might have merely copied the answers from questions 4-6 into the answer slots for questions 8-10, which would coincidentally have been the correct math answers in the frequency version. Possibly-copied answers appeared with negligible frequency in the probability version, however, and the overall effect in Experiment 3 was not very different than in Experiment 2, so we consider this possible copying to be a highly unlikely influence on our results.

General Discussion

The results compellingly demonstrate that information presented and manipulated in a frequency format facilitated Bayesian competence in understanding the mathematics underlying the three doors problem. Given a presentation of the problem in frequency format, rates of correctness on the math questions ranged from 7% to 26%, depending on the experiment. Given a presentation in probability format, correctness on the math questions was a flat 0% in all experiments. This supports Gigerenzer's hypothesis about the importance of information format to normative Bayesian performance on inference tasks.

Note that the 0% correctness rate on the math questions in the probability versions does not conflict with past results, although it does reflect our using a different measure of performance. Previous studies such as that of Granberg & Brown asked merely for a stay/switch answer, not for evidence of whether responders understood the underlying mathematics or evidence about which information format responders might be considering when computing the relevant likelihoods, so our central result is based upon a different question. A comparison of like measures yields compatible results: We found that between 12% and 27% of our subjects answered the stay/switch question correctly, depending on the experiment; in their initial study, Granberg & Brown (1995) found 13% of their subjects answered the stay/switch question correctly. The disagreement between the high end of our numbers and the percentage reported by Granberg & Brown (1995) may have several explanations, from the presence of a computationally-oriented questionnaire in our study to different subject populations and myriad other factors. This discrepancy, however, is unrelated to the crux of our experiment, the value of one information format over another. Are the algorithms that people use for mathematical reasoning better tuned for natural frequencies than probabilities? On this particular brainteaser, it seems that the change in information format can indeed open doors for people.

Finally, in addition to suggesting that Gigerenzer's hypothesis applies to two-stage decision processes such as the three doors problem as well as simpler contexts already explored (Gigerenzer, in press), we hope our results can add a new aspect to the current three doors problem literature. There are many printed and World Wide Web based explanations of the correct stay/switch answer to the game, but we suggest that an explanation in frequency format — which seems to be a highly non-standard approach, given a quick scan of the current literature — might help some of the many people who are initially skeptical of the Bayesian answer to understand and accept that answer. This use of the explanatory power of frequency format presentations would be consistent with the claim (Gigerenzer, in press) that frequency format explanations render more persuasive and understandable arguments about likelihoods relevant to the O.J. Simpson trial, HIV testing, and other topics of widespread popular interest.

Appendix: Answers to Frequency Format Questionnaire

The correct answers to the frequency format questionnaire in Figure 2 are:

1) 10,000
2) 10,000
3) 10,000 (because the car is placed randomly)
4) 5,000 (because the host chooses fairly when he can)
5) 0
6) 10,000
7) 15,000
8) 5,000
9) 0
10) 10,000

And, of course: SWITCH.

Acknowledgments

We thank: Gerd Gigerenzer for his support and encouragement; David Gries for his helpful comments on drafts of this paper; David Dunning, Hal Perkins, and Marian E. Polhill for allowing us time in their classes; Carl Lagoze for his technical support; Nadia Wong and Elvis Au for helping handle the questionnaires. The first author was supported by NSF grant GER-9454149.

References

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition, 43*, 197-223.

Gardner, M. (1959a). Mathematical games: Problems involving questions of probability and ambiguity. *Scientific American, 201(4)*, 174-182.

Gardner, M. (1959b). Mathematical games: How three mathematicians disproved a celebrated conjecture of Leonard Euler. *Scientific American, 201(5)*, 181-188.

Gigerenzer, G. (in press). Ecological intelligence: An adaptation for frequencies. In D. Cummings & C. Allen (Eds.), *The Evolution of Mind.* New York: Oxford University Press, 9-29.

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

Gilovich, T., Medvec, V. H., & Chen, S. (1995). Commission, omission, and dissonance reduction: Coping with regret in the "Monty Hall" problem. *Personality and Social Psychology Bulletin, 21*, 182-190.

Granberg, D. & Brown, T. A. (1995). The Monty Hall dilemma. *Personality and Social Psychology Bulletin, 7,* 711-723.

Hoffrage, U. & Gigerenzer, G. (1996). The impact of information representation on Bayesian reasoning. *Proceedings of the 18th Annual Conference of the Cognitive Science Society*. pp. 126-130. Mahwah, NJ: Lawrence Erlbaum Associates.

Kahneman, D. & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3*, 430-454.

Selvin, S. (1975a). A problem in probability. *American Statistician, 29*, 67.

Selvin, S. (1975b). On the Monty Hall problem. *American Statistician, 29*, 134.

Tierney, J. (1991, July 21). Behind Monty Hall's doors: Puzzle, debate and answer? *New York Times,* pp. 1,20.

vos Savant, M. (1990a, September 9). Ask Marilyn. *Parade*, p. 15.

vos Savant, M. (1990b, December 2). Ask Marilyn. *Parade,* p. 25.

vos Savant, M. (1991a, February 17). Ask Marilyn. *Parade,* p. 12.

vos Savant, M. (1991b, July 7). Ask Marilyn. *Parade,* p. 26.