

# Overcoming illusory inferences in a probabilistic counterintuitive problem: The role of explicit representations

ELISABET TUBAU

*University of Barcelona, Barcelona, Spain*

and

DIEGO ALONSO

*University of Almería, Almería, Spain*

In the context of conditional probabilities, a good example of the marked discrepancy between intuition and formal reasoning is the Monty Hall dilemma (MHD). We used the MHD to study the effects of practicing the game, making explicit the underlying structure, or enhancing the representation of the different possibilities, on reaching and stating the correct answer. The results of the experiments showed that accumulated experience with the MHD increased the proportion of switching responses but did not change erroneous intuitions (Experiment 1). However, when the dilemma was presented in the form of an adversary game that made the underlying structure more explicit, more participants formed complete mental representations that enabled them to reason correctly (Experiment 2). This result was observed even without any practice with the game if the participants were encouraged to represent possibilities (Experiment 3). Therefore, in this context, correct reasoning seems to depend more on the ability to consider different possibilities than on extensive practice with the game.

A typical finding in probabilistic reasoning research is that subjective judgments frequently violate mathematical theorems. This phenomenon was reported initially by Kahneman and Tversky (1972), and since then many other authors have published data that illustrate this point. Examples of the phenomenon are base-rate neglect, the conjunction fallacy, and the anchor effect. Since Kahneman and Tversky's heuristic approach, a number of authors have sought to explain the mechanisms underlying intuitive probabilistic reasoning and the cause of the discrepancies between intuition and formal reasoning. For example, Gigerenzer and Murray (1987) and Gigerenzer and Hoffrage (1995) argued that phenomena such as base-rate neglect or the conjunction fallacy could be better explained by means of the frequentist hypothesis than by means of the representativeness or similarity heuristic (Tversky & Kahneman, 1983). Frequentist approaches (see also Cosmides & Tooby, 1996) hold that these biases reflect a poor adaptation of the reasoning processes to information rep-

resented in a probabilistic format (from 0 to 1) or in percentages (from 0 to 100), but also that human reasoning can improve significantly when information is given in frequencies instead of probabilities, and when the answer can also be given by means of frequencies. In other words, from the frequentist perspective, performance in probabilistic reasoning improves when input and output are represented by natural frequencies. However, other researchers (e.g., Girotto & Gonzalez, 2001, 2002) have claimed that the improvement found with frequentist formats is not due to the use of frequencies, but depends on the ability to represent the sets or partitions implied by the problem structure.

On the other hand, there are counterintuitive problems in which the cause of the erroneous inferences is not to be found in the problem format but in the explicitness of the underlying assumptions (e.g., the prisoner's paradox or the three-card problem; see Bar-Hillel & Falk, 1982; Piattelli-Palmarini, 1994). In these situations, the numbers involved in the operations are not expressed in terms of probabilities or percentages and, in fact, they can be solved without applying mathematical procedures such as the Bayes theorem (see the Monty Hall dilemma, below). The difficulty of these problems can also be understood in the context of Girotto and Gonzalez's (2001) set representation or from the perspective of the mental model theory (MMT) applied to extensional probabilistic reasoning (Johnson-Laird, Legrenzi, Girotto, Sonino-Legrenzi, & Caverni, 1999). According to the latter theory, probabilistic reasoning is based on mental models that represent the ex-

---

We thank Evan Heit, Mike Oaksford, and two anonymous reviewers for their helpful comments on a previous version of this article. We also thank Joan López-Moliner for his programming assistance and for his clever suggestions. Preparation of this article has been supported in part by the Secretaría General de Universidades e Investigación (Junta de Andalucía) Research Grant LJC/EBA to D.A. during his time at the University of Barcelona. Correspondence concerning this article should be addressed to E. Tubau, Departament de Psicologia Bàsica, Universitat de Barcelona, Pg. Vall d'Hebron, 171, 08035 Barcelona, Spain (e-mail: etubau@psi.ub.es).

pected possibilities (which may be represented by means of either natural frequencies or probabilities). Consequently, reasoning may be erroneous and biases may appear, in conditions in which it is difficult to imagine and represent all the models or possibilities.

### The Monty Hall Dilemma

One example of a very counterintuitive probabilistic problem is the well-known *three-doors problem* or *Monty Hall dilemma* (MHD). The problem, adapted from an American TV game show of the 1960s, involves an interaction between the host (Monty Hall) and a guest. The guest is presented with three closed doors, only one of which hides a valuable prize, and has to choose one of them. After the guest makes the selection, the host, who knows where the prize is, opens one of the nonchosen doors to show that it does not contain the prize. The guest now has the chance of switching to the other closed door or staying with his or her initial selection. That is, the guest is faced with the dilemma of switching or sticking (recent analyses of this problem can be found in Eisenhauer, 2000; Granberg & Brown, 1995; and Piattelli-Palmarini, 1994).

In many lay observations and experimental studies, it has been reported that both novices and expert statisticians have great difficulty in producing the rational decision. For example, after analyzing the behavior of 104 students from different disciplines, Friedman (1998) found that, on average, more than 90% preferred to stay with the chosen alternative. Moreover, this decision was independent of their statistical knowledge. Surprisingly, from a normative point of view, the best strategy is to change to the other closed door; the chances of winning if the contestant changes are  $\frac{2}{3}$  and only  $\frac{1}{3}$  if he or she stays. An informal justification for this conclusion might take the following form: The host will never open the door concealing the prize and, obviously, he will not open the guest's door either. So, taking into account the fact that two thirds of the times the prize will be in one of the nonchosen doors, the nonchosen door that is still closed will hide the prize in two thirds of the trials (see, e.g., Eisenhauer, 2000, for a mathematical analysis of the answer).

### Probability/Frequency Learning in the MHD

In some experiments with the MHD, it has been reported that subjects change their initial decisions after some practice (Friedman, 1998; Granberg, 1999; Granberg & Brown, 1995; Granberg & Dorr, 1998). Granberg and Brown showed an increment from 10% switching (in the first trial) to 55% (in the last 10) after 50 practice trials. Similarly, Friedman found an increment from 10% switching to 48% after 15 practice trials. However, it is not clear if this training modifies the mental representation of the problem and, consequently, the reasoning process. Granberg and Brown concluded: "Even for those with the greater incentives to play a switching strategy, full insight into the correct solution and purely rational behavior did not emerge" (1995, p. 720). Moreover, the probability estimations of subjects who were clearly informed about the relevance of

the host's knowledge did not differ from those of subjects in the control group; their initial probability estimation for the first decision was around .35, and their final estimation for the second decision was .50. As the authors commented, "people do not adequately take into account the knowledgeable host cue" (Granberg & Brown, 1995, p. 720).

On the other hand, Friedman concluded that, although the MHD is not easy to learn, "most of the people made the rational choice most of the time in the more favorable conditions" and that "the rationality rate sometimes exceeded the classic probability matching benchmark of 2/3 for transparent risky choices with the same odds" (1998, p. 941). Friedman based this conclusion on the observation that some of his subjects were switching in more than 70% of the last trials. However, does the high frequency of correct choices necessarily imply that the subjects were reasoning in accordance with the mathematical structure of the task?

There is experimental evidence that some subjects exceed probability matching in different associative learning paradigms, especially when they are aware of the programmed contingencies (see, e.g., Kudadjie-Gyamfi & Rachlin, 2002; Roth & Erev, 1995). Consequently, it is possible that the individual differences observed in the experiments above could reflect different levels of awareness of the response-reinforcement contingency. In this regard, some of the subjects in Friedman's (1998) study may have been aware of the relation between switching and winning. But does awareness of this contingency necessarily imply a change in the reasoning process?

### Reasoning in the MHD

Shimojo and Ichikawa (1989) ran an experiment with a problem analogous to the MHD: the three prisoners problem.<sup>1</sup> In order to analyze the strategies underlying the reasoning, these authors used a version in which the initial probabilities were unequal for the three alternatives. They found that the erroneous intuitions, based on subjective theorems, did not disappear after subjects were aware of the correct answer. Basically, they identified three subjective theorems: (1) The *number-of-cases theorem*: If the number of alternatives is  $N$ , then the probability of each one is  $1/N$ . (2) The *constant ratio theorem*: If one of the alternatives is eliminated, the ratio of probabilities between the remaining alternatives does not change. (3) The *irrelevant, therefore invariant theorem*: If it is true that at least one of several alternatives will be eliminated, the information that specifies the alternative to be eliminated is irrelevant and does not affect the probabilities of the other. So, from this point of view, the probability estimates depend on the subjective theorem applied in each case. Similar conclusions can be read in Falk (1992), who suggested that the more frequent theorems or heuristics are the number-of-cases, or, in Falk's words, *uniformity* heuristic and the irrelevant and, therefore, invariant, or *no-news, no-change* heuristic.

Johnson-Laird et al. (1999) analyzed some counterintuitive probabilistic problems in terms of the MMT (Johnson-Laird & Byrne, 1991). The MMT postulates that

extensional probabilistic reasoning is based on three fundamental principles. First, the *truth principle* states that people construct sets of mental models, each of which represents what is true in a true possibility. Second, according to the *equiprobability principle*, each model represents an equiprobable alternative unless people have knowledge or beliefs to the contrary. Third, according to the *proportionality principle*, the probability of an event depends on the proportion of models in which the event occurs. In the case of the MHD, people build three equiprobable models (one for each possibility). Then, if, for example, Door 1 is chosen and the host opens Door 2, people remove the model representing the possibility that the prize could be behind Door 2. This will automatically yield two equiprobable models (one with the prize behind Door 1 and the other with the prize behind Door 3). But people fail to represent (and, as a consequence, do not consider) the crucial underlying assumption (the elimination condition) that, if the contestant chooses Door 1, the host will *always* open Door 2 when the prize is behind Door 3 and only *half the time* when the prize is behind Door 1. Therefore, if Door 2 is opened, the chances of winning are doubled if the contestant switches to Door 3 instead of sticking with Door 1.

In conclusion, psychological research with the MHD or analogous problems has shown that people base their reasoning on strong theorems or heuristics that, as a consequence, frequently produce erroneous intuitions. In terms of the MMT, the principal cause of these intuitions is the difficulty of representing all the information (e.g., different possibilities) that should be taken into account.

### Overview of the Experiments

The aim of the experiments presented here was to analyze the effects of practicing the game several times, making the underlying structure explicit, and enhancing the ability to imagine different possibilities on changing the erroneous intuition associated with the MHD.

In Experiment 1, the participants played several times with a computer version of the MHD. In light of the results from the probability learning literature reviewed above, we expected to find an improvement of the correct response (i.e., to switch) through the training phase. In order to investigate the relation between this improvement and explicit judgments, the participants had to estimate the relative probability of winning after switching and after sticking, and the frequencies of wins for both events (i.e., switching and nonswitching trials). If the frequency information is important for improving probabilistic reasoning, participants with more switching experience should have to change their erroneous intuition more often. The main goal of Experiment 2 was to compare the standard version of the MHD with an analogous problem designed to present the underlying structure more explicitly. The analogous problem took the form of a card game between two participants (*decision maker* and *informant*). The goal of the game was to draw an ace from a pool of three cards (one ace and two non-aces). The decision maker, analogous to the contestant in the MHD, chooses one of the

cards. The informant, analogous to the host, keeps the remaining cards. So, the relevant sets or partitions (Giroto & Gonzalez, 2001) were presented more explicitly by means of two players with different probabilities of having the ace. As was stated above, from the point of view of the MMT (Johnson-Laird et al., 1999), reasoning depends on the ability to imagine models or possibilities. Consequently, assuming that the adversary version presented the different possibilities of both players more explicitly, we expected to find a significant improvement of reasoning with this new version. Finally, in Experiment 3, the effect of verbalizing or representing explicitly the underlying possibilities before producing the answer to the problem was studied. We expected that this imagination exercise would help participants to pay more attention to the different possibilities and, therefore, to facilitate the emergence of correct reasoning.

### EXPERIMENT 1

The MHD is an interesting setting for the study of the practice effect on changing strong erroneous intuitions. As we said above, Granberg and Brown (1995) also analyzed the effect of training in the MHD, but they considered only the proportion of switching responses—they did not consider the verbal judgments of the participants. In Experiment 1, we tried to answer one main question: Could the learning phase help to uncover the mathematical structure of the problem? Taking into account the strong erroneous intuitions associated with the problem, it is conceivable that the training phase would improve performance (observed as an increment in the tendency to switch) without changing the participants' explicit judgments about the best strategy for solving the problem. Hence, after a training phase, the participants had to make an explicit judgment about which strategy was better, and they also had to estimate the frequency of wins from 10 switching and 10 nonswitching trials. From the verbal judgments, the participants were grouped according to the correctness of their answers.

Taking into account the results of Granberg and Brown (1995) and of Granberg and Dorr (1998), we designed a modified presentation of the MHD. Granberg and Brown found a high tendency to stick in the MHD even after 50 practice trials, reflecting the general phenomenon of belief perseverance or an illusion of control (Langer, 1975). People tend to attribute a higher probability value to their own original choice, or they protect themselves from the unpleasant possibility of missing out on a prize precisely because they changed their minds. It has been observed that the illusion-of-control phenomenon may be attenuated when people are given an alternative so that they have no chance to choose (Langer, 1975). So, in order to facilitate the switching experience in the context of the MHD, our computer program simulated a lottery in which, at the beginning of the game, the participant was randomly assigned one of the three options. By means of this procedure, we expected that the participants would not be so re-

**Table 1**  
**Percentage of Responses From the Verbal Judgments in Each Answer Category in Experiment 1**

Initial Answer	Final Answer			Total
	<i>Equal</i>	<i>Switch</i>	<i>Stick</i>	
<i>Equal</i>	40 (23)	19 (11)	11 (6)	70 (40)
<i>Switch</i>	12 (7)	7 (4)	–	19 (11)
<i>Stick</i>	4 (2)	7 (4)	–	11 (6)
Total	56 (32)	33 (19)	11 (6)	100 (57)

Note—The responses to the initial question are presented in rows, and the responses to the final question are shown in columns. Raw frequencies are shown in parentheses. *Equal*: Both to stick and to switch have the same advantage. *Switch*: better to switch; *Stick*: better to stick.

luctant to switch and, as a consequence, they would have more opportunities to associate (either implicitly or explicitly) the switching strategy with a higher probability of winning.

## Method

**Participants.** Fifty-seven students from the University of Barcelona participated in the first experiment. The participants were grouped according to the accuracy of their verbal judgments (see the Procedure section).

**Materials.** A computer program simulated an adapted version of the game *Let's Make a Deal*, hosted by Monty Hall. The participants were shown three gray (closed) windows and were informed that only one of them would be colored red once opened (the red window representing the prize) and the other two would be white. At the beginning of each trial, the computer randomly assigned one of the windows to the participant. Then, one of the remaining windows was opened in order to show that it was not red (i.e., it did not contain the prize). Finally, the participant had the chance to decide whether to accept the assigned window or to switch to the other one that was still closed. Once the final decision was made, all the windows were opened and the participant could see the location of the red window. The number of win and no-win trials was constantly shown. The program saved the responses made by the participants in each trial, as well as the beginning and final verbal judgments (see below).

**Procedure.** The instructions explained the computer version of the MHD, with a clear presentation of the randomness of the initial lottery, as explained in the Materials section. The participant's goal was to win the prize (red window) as many times as possible. After reading the instructions and before beginning the learning phase, the participant had to answer the question, "Which option is better: switching, sticking, or no preference?" This question was repeated at the end of the learning phase. The answers, or explicit judgments about relative probability, were used to group the participants into different categories (correct and incorrect; see the Results and Discussion section). Finally, the participants had to estimate the frequency of wins for both 10 switching trials and 10 sticking trials. The participants were presented the MHD 45 times, grouped, for the analyses, into three blocks of 15 trials each.

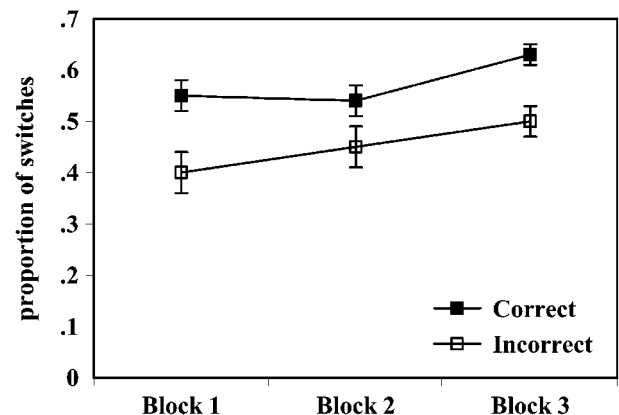
## Results and Discussion

According to their explicit judgments, only 33% of the participants apparently noticed, at the end of the training phase, the advantage of switching over sticking (see Table 1). Most of them (56%) still considered that switching and sticking were equally good strategies after 45 practice trials.

In order to analyze the relation between explicit judgments and switching responses, we grouped the partici-

pants into two categories: *correct*, or those who, at the end of the training phase, responded correctly that it was a better strategy to switch than to stick ( $n = 19$ ), and *incorrect*, or those who considered that both strategies had the same chance ( $n = 32$ ). The analysis of variance (ANOVA) of proportion of switches yielded a significant effect of block [ $F(2,98) = 5.68, MS_e = 0.19, p < .01$ ] and of group [correct vs. incorrect:  $F(1,49) = 13.64, MS_e = 0.56, p < .01$ ]. The interaction between block and group was not significant, denoting a clear effect of block for both groups. As can be seen in Figure 1, there was an increase in switches through the learning phase, even in the case of the incorrect participants. Thus, a dissociation was found between these two measures (explicit judgments and switching responses), suggesting that the incorrect participants noticed the advantage of the switching response, but only implicitly.

In relation to the frequency estimations given at the end of the training phase, the ANOVA on these data, with type of question (number of wins for 10 switching trials and for 10 sticking trials) as the within-participants factor, and group as the between-participants factor, yielded a main effect of type of question [ $F(1,49) = 24.54, MS_e = 27.46, p < .001$ ] and a significant interaction between type of question and group [ $F(1,49) = 20.44, MS_e = 22.87, p < .001$ ]. Taking into account that the programmed probabilities constituted a two-thirds chance of winning after switching and a one-third chance after sticking, the theoretical frequencies were 6.7 and 3.3 for 10 switching trials and 10 sticking trials, respectively. The frequency estimations made by the correct participants were much closer to the theoretical ones (means were 6.6 and 4.6 for 10 switching and 10 sticking trials, respectively) than the estimations of the incorrect participants (means were 4.8 and 4.9, respectively). The difference between the two groups was also significant [ $F(1,49) = 6.19, MS_e = 12.17, p < .05$ ]. As was explained above, the distinction between the correct and the incorrect participants was made on the basis



**Figure 1.** Proportion of switches for each trial block in Experiment 1 for correct and incorrect participants, taking into account the final relative probability judgment. Error bars represent standard errors of the mean.

of the answers given to the final question about the relative advantage of switching or sticking strategies. Surprisingly, only 11 of the 19 correct participants (58%) attributed a higher frequency of wins by switching than by sticking, but most of the incorrect participants (30 of 32, or 94%) were consistent in their frequency estimations (i.e., their estimations were the same for both situations). In light of this finding, we regrouped all of the participants into two new frequency-accuracy groups: those who had estimated correctly a higher frequency of wins by switching, or *frequency learners* ( $n = 14$ ; mean frequency estimations were 6.5 and 3.1 for switching and for sticking trials, respectively) and the *frequency nonlearners* ( $n = 43$ ; mean frequency estimations were 5.2 and 5.1 for switching and sticking trials, respectively). The ANOVA of the proportion of switches with block as the within-participants factor and frequency-accuracy group as the between-participants factor yielded a general effect of block [ $F(2,110) = 3.7, MS_e = 0.14, p < .05$ ] and of frequency-accuracy group [ $F(1,55) = 6.46, MS_e = 0.29, p < .01$ ]. The interaction was not significant.

Hence, a dissociation also appeared between the frequency estimations and the switching response. Frequency nonlearners attributed the same frequency of wins for switching as for sticking; however, their proportion of switches increased significantly through the three blocks. These results suggest that the participants adapted their behavior to the experienced contingency between switching responses and probability of winning. Nevertheless, this sensitivity, reflected in the tendency to switch rather than stick, was not sufficient to cause the participants to make explicit conscious judgments and frequency estimations. One unexpected result found in Experiment 1 was the low consistency between relative probability judgments and frequency estimations shown by the correct participants (with respect to the relative probabilities of winning by switching or by sticking). Almost 50% of them answered that it was better to switch than to stick, but assigned the same frequencies of wins (approximately five) for 10 switching trials as for 10 sticking trials. Therefore, only the stated preference for the switching strategy seemed to be insufficient to prove a complete understanding of the problem.

Moreover, informal questionnaires given to the correct participants showed that none of them was able to explain the reason for the differences in frequencies or probabilities. These participants realized that they were more likely to win after changing the initial selection than after sticking, but could not explain why. Most of them argued that the computer had been programmed to behave in a certain way, but that they could not understand it. From their perspective, "logically," there should be the same probability of winning after switching as after sticking (concretely, .5). With the explanation that the computer behaved in a nonlogical way, our participants were able to keep their initial erroneous intuition (i.e., "the probability of winning after switching should be the same as that for winning after sticking") intact. Consequently, the experience

observed was not sufficient to change the underlying representation of the problem. As was stated above, in Experiment 2 we explored a new version of the game together with more systematic questionnaires for both the standard version and the new one.

## EXPERIMENT 2

One of the problems with the standard version of the MHD is the difficulty of considering the elimination condition. As we said before, from the point of view of the MMT, in this situation people build three equiprobable mental models. Once one of them is eliminated, they assume that the probabilities are distributed equally in the two remaining alternatives, although in fact the probabilities of the remaining alternatives (the one given to the participant and the one that remained closed in our experiment) are unequal.

The modified version that we presented in Experiment 2 was a game between two participants. One of them was the *decision maker* and the other the *informant*. The goal of the game was to draw an ace from a pool of three cards (one ace and two non-aces). The decision maker had to choose one of the three cards without seeing it. The remaining cards were for the informant. So, at the beginning, the game involved the building of two nonequiprobable sets: one with one possibility and the other with the remaining two possibilities. In terms of the MMT, the decision maker's chances could be represented by means of three equiprobable models, only one representing the ace (or a model representing the ace with a probability of one third), whereas the representation concerning the adversary's chances would consist of three equiprobable models: two with the ace and one without the ace (or a model representing the ace with a probability of two thirds; see Table 2 and, for more details of these representations, see Johnson-Laird et al., 1999).

Since the informant had the advantage, one of the game's rules was that he or she had to show one non-ace to the decision maker. Then, the decision maker had the opportunity to choose between sticking and switching (i.e., changing the selected card for the informant's other card, still hidden). If the participants were able to mentally represent the models explained above, they would easily realize that the informant would hide the ace two thirds of the time and, therefore, that the best strategy was to change card. On the basis of this mental representation, we expected an improvement in the probabilistic reasoning with this new version of the problem.

**Table 2**  
**Possibilities of Having the Ace for the Decision Maker and for the Informant**

Player 1 (Decision Maker)	Player 2 (Informant)
{ace}	{no ace, no ace}
{no ace}	{ace, no ace}
{no ace}	{no ace, ace}
$p(\text{ace}) = \frac{1}{3}$	$p(\text{ace}) = \frac{2}{3}$

## Method

**Participants.** Sixty-two students from the University of Barcelona participated in this experiment. Twenty-eight participants played the standard version of the Monty Hall game (control condition). The remaining 34 participants interacted in pairs in the new version of the game (17 as decision makers and 17 as informants).

**Materials.** Three poker cards (one ace and two non-aces) were used in this experiment both for the standard version of the Monty Hall game and for the new version. The participants wrote their answers (their preferred strategy and their justification) in a questionnaire at the beginning, middle, and end of the experiment, together with the responses given in each trial.

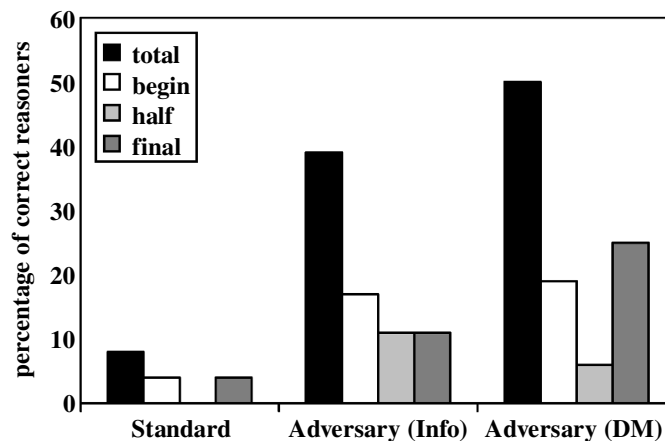
**Procedure.** Two different conditions (control or standard version and experimental or adversary version) were designed. For the control condition, an adapted version of the MHD was presented. In this condition, the participants interacted individually with the experimenter. The experimenter showed the three cards (one ace and two non-aces) and explained the goal of the game: to draw the ace as many times as possible. To approximate the situation to the standard game, the cards were put on a support with three holes so that the participant could see one side and the experimenter the other. The holes were numbered from 1 to 3. For each trial, the experimenter shuffled the cards and put them into the holes with their backs to the participant, and the participant chose one of them, indicating the number. Then, from between the nonselected cards, the experimenter showed a card other than the ace. At that moment, the participant was prompted to choose between sticking with his or her chosen card and switching to the still hidden card. This procedure was run 18 times (2 blocks of 9 trials). In the experimental condition, the participants participated in pairs and played against each other. In each trial, after the experimenter had shuffled the cards, the decision maker chose one of them. The other two cards were given to the informant. The informant, who could see his or her cards, showed one non-ace card to the decision maker, and the latter then decided to stay or to switch. These participants played 18 trials as well. At three points of the experiment (beginning, middle, and end), the participants in both conditions (standard game and adversary game) and in both roles (decision maker and informant) were asked to judge the relative advantage of the two strategies (to switch or to stick,

from the decision maker's perspective) and to write down the reasons for their judgments.

## Results and Discussion

One of the decision makers was eliminated due to her familiarity with a similar problem, which resulted in 28 participants for the standard version and 33 for the adversary version (16 decision makers and 17 informants). Written explanations were considered correct when the reason for changing referred explicitly to the greater probability (whether relative or absolute) of the informant's having the ace and, consequently, that the last hidden card would more often be the ace. One prototypical example of correct reasoning is, "The decision maker should switch cards because the informant has more chances of having the ace." In the case of the standard condition, the correct reasoning had to refer to the greater probability (also relative or absolute) of the nonchosen card's being the ace. With this criterion, both authors evaluated the written answers, and there was a complete agreement in the final categories of correct reasoners and incorrect reasoners (interrater reliability was 100%).

Figure 2 shows the percentage of participants in the two conditions who reasoned correctly (both the total percentages and the partial percentages corresponding to the beginning, middle, or end point in the experiment). For the adversary condition, the results are presented separately for decision makers and informants. Statistical analyses showed a significant difference in the total percentage of correct reasoners between standard (8%) and adversary conditions [decision makers: 50%,  $\chi^2(1) = 10.65$ ,  $p < .005$ ; informants, 38%,  $\chi^2(1) = 7.66$ ,  $p < .01$ ]. The difference between the adversary subgroups (decision makers and informants) was not significant.



**Figure 2.** Percentage of participants who reasoned correctly in Experiment 2 at the beginning, middle (halfway point), and end of the experiment (total columns, in black, represent the sum of the other columns—that is, the percentage of participants who reasoned correctly at the beginning, in the middle, or at the end of the experiment). Standard: Monty Hall dilemma, standard version. Adversary: Monty Hall dilemma, adversary version. Info: informants; DM: decision makers.

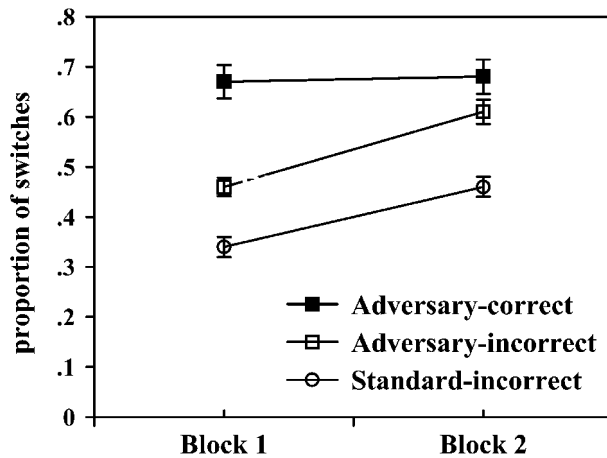


Figure 3. Proportion of switches through the two blocks (nine trials each) in Experiment 2. Standard: Monty Hall dilemma, standard version (incorrect reasoners). Adversary: Monty Hall dilemma, adversary version (correct and incorrect reasoners). Error bars represent standard errors of the means.

The proportion of switches was also calculated for the standard-condition participants and for the adversary-condition participants as decision makers for the two blocks of the experiment. Figure 3 shows these proportions for correct and incorrect reasoners (due to the low percentage of correct reasoners in the standard version, we did not plot the data from those participants). A two-way ANOVA was performed with block (first and second) as the within-participants variable and reasoner (correct and incorrect) as the between-participants variable. Both effects were significant [ $F(1,42) = 5.94, MS_e = 0.10, p < .05$ , and  $F(1,42) = 14.91, MS_e = 1.33, p < .001$ , respectively]. Separate comparisons in the adversary condition showed a significant difference in proportion of switches between correct and incorrect reasoners, but only in the first block of the experiment [ $t(14) = 2.61, p < .05$ ]. There were no significant differences between incorrect reasoners from the two conditions (standard and adversary). As can be observed in Figure 3, and similarly to previous results, the participants from the standard condition who did not reason correctly were very reluctant to change, especially in the first block of the experiment, in which the proportion of switches was .34.

The results of Experiment 2 showed that our prediction was confirmed: Presentation of the MHD in an adversary context that make the relevant sets more explicit facilitated the emergence of correct reasoning. It is worth noting that, in both the roles of decision maker and informant, some of the correct reasoners (39%) were already able to produce the correct argument at the beginning of the game. This point is analyzed further in Experiment 3.

### EXPERIMENT 3

One possible explanation that can account for the differences between correct and incorrect reasoners in Experi-

ment 2 could be related to differences in the explicitness of the different possibilities involved in the game. The verbal arguments suggested that the correct reasoners had built a more complete representation of the different possibilities of both adversary players (decision maker and informant) than had their incorrect counterparts. Moreover, the formation of this complete representation seemed to be independent of amount of practice. According to the MMT, considering the different possibilities of both players involved in the game should be enough to improve probabilistic reasoning. Therefore, a procedure that pays more attention to the different possibilities should produce correct reasoning more often. The goal of Experiment 3 was to test this hypothesis by means of a comparison of participants who represented graphically the different possibilities involved in the game (e.g., locations of the ace or the prize) with those who did not complete this initial task.

### Method

**Participants.** Sixty students from the University of Barcelona participated in Experiment 3. Twenty participants were randomly assigned to each of three conditions: (1) Monty Hall, standard version with explicit representation (standard-ER); (2) Monty Hall, adversary version with explicit representation (adversary-ER); and (3) Monty Hall, adversary version without explicit representation (adversary).

**Materials.** Three different questionnaires were constructed. For the standard-ER condition, the questionnaire explained first the standard MHD involving a host, a guest, and three envelopes (instead of windows or doors), and the participants were asked to represent graphically the different possible locations of the prize. Afterwards, the participants had to (1a) estimate the probability of choosing the envelope with the prize, (1b) estimate the probability that the remaining alternatives would contain the prize, (2a) estimate the probability that the chosen envelope would contain the prize after elimination of one of the remaining alternatives, (2b) estimate the probability of the nonchosen and noneliminated alternative would contain the prize, (3) estimate the relative probability of winning after switching or after sticking (i.e., which strategy was considered better); and (4) explain verbally the reason for their previous answers (see the Appendix). For the adversary-ER condition, the questionnaire presented the rules of the adversary-game version used in Experiment 2, and the participants were also asked to represent the different possible locations of the ace. The participants then had to answer the same questions (adapted to the adversary game) as in the previous condition (see the Appendix). Finally, the questionnaire for the adversary condition was exactly the same as that for the adversary-ER condition, except that the participants were not required to represent the different possible locations of the ace explicitly.

**Procedure.** The participants, in small groups of 4 or 5, were given the questionnaires and told to try to imagine the different situations explained in the text as accurately as possible and to think carefully before answering the questions. They were allowed as much time as they needed. All of the participants completed the questionnaire in 15–20 min.

### Results and Discussion

Only 1 of the participants in the adversary-ER condition was unable to correctly represent the three possibilities of the distribution of the ace. All of the participants in the standard-ER represented the three possibilities correctly. In relation to the question of the alternative that might be eliminated in each possibility (see the Appendix

**Table 3**  
**Percentage of Correct Responses in Experiment 3**  
**(by Group and by Question)**

Question	Standard-ER	Adversary-ER	Adversary
1A. <i>p</i> (ace DM)	95	95	95
1B. <i>p</i> (ace Inf.)	55	85	65
2A. <i>p</i> (ace DM Elim.)	20	75	70
2B. <i>p</i> (ace Inf. Elim.)	10	65	10
3. Best strategy?	25	60	50
4. Why?	0	60	25

Note—ER, explicit representation of the possibilities; DM, decision maker (or contestant); Inf., informant (or remaining envelopes); Elim., after the elimination of one alternative (card or envelope). The questionnaires are included in the Appendix.

for details), 75% of the participants in the adversary-ER condition and 65% of those in the standard-ER condition responded correctly. These differences were not significant. The first author analyzed the verbal reasoning question by applying the same criterion of correctness as in Experiment 2. For the other items, the percentages of mathematically correct responses were compared (see Table 3).

Almost all of the participants in the three conditions (95%) correctly estimated that the probability of choosing the prize (or the ace) was one third (Item 1a). However, more participants in the adversary-ER condition expressed correctly the probabilities of the remaining alternatives of containing the ace (85%, 65%, and 55% for adversary-ER, adversary, and standard-ER, respectively; see Table 3, Item 1b). The percentage comparisons showed that only the adversary-ER condition differed significantly from the standard-ER condition [ $\chi^2(1) = 4.28$ ,  $p < .05$ ]. The other comparisons were not significant.

Most of the participants in the two adversary conditions also realized that the initial probability of choosing the ace did not change after the elimination of one of the remaining alternatives [Item 2a; percentage comparisons between standard-ER (20%) and adversary-ER (75%),  $\chi^2(1) = 12.13$ ,  $p < .005$ , and between standard-ER and adversary (70%),  $\chi^2(1) = 10.10$ ,  $p < .005$ ], but many more participants in the adversary-ER condition estimated correctly the probability of the noneliminated alternative [Item 2b; comparison between adversary-ER (65%) and both standard-ER (10%) and adversary (10%),  $\chi^2(2) = 19.86$ ,  $p < .001$ ].

Approximately half of the participants in the two adversary conditions considered that it was better to switch than to stick. Both conditions differed from the standard-ER condition [ $\chi^2(2) = 11.22$  and  $\chi^2(2) = 8.37$ ,  $p < .05$ , in comparison with adversary-ER and adversary, respectively], but many more participants in the adversary-ER condition were able to reason correctly. In this case, all comparisons were significant [standard-ER (0%) vs. adversary-ER (60%),  $\chi^2(1) = 17.14$ ,  $p < .001$ ; standard-ER vs. adversary (25%),  $\chi^2(1) = 5.71$ ,  $p < .05$ ; adversary-ER vs. adversary,  $\chi^2(1) = 5.01$ ,  $p < .05$ ].

In summary, the results showed that, as in Experiment 2, the adversary condition facilitated the emergence of cor-

rect probabilistic reasoning in comparison with the standard condition. Moreover, Experiment 3 showed that the adversary condition yielded more correct estimations of probabilities, especially after the different possibilities were made explicit (adversary-ER condition). As was argued above, these results seem to support the mental model approach (Johnson-Laird et al., 1999) in the sense that the direct attention to the possibilities involved in the game significantly decreases the emergence of the illusory inference. Alternatively, it may be that the adversary context produced more correct reasoning due to the ambiguity of the datum, "once the informant has shown a non-ace card" (see the Appendix) instead of, for example, "once the informant has shown the 7 of diamonds." It is possible that without the ambiguity, more participants would have produced the illusory inference. However, the same ambiguity was present in the standard version of Experiment 3, and in that condition it did not seem to be helpful; neither was the activity of explicitly representing the different possibilities. That is, the results of the latter condition did not differ from those of the standard condition of Experiment 2, in which no ambiguity was present and in which the explicit representation of the different possibilities was not required.

## GENERAL DISCUSSION

People learn to discriminate between different frequencies or probabilities relatively easily. As far as the probabilities involved in the MHD (two thirds vs. one third) are concerned, a pilot experiment with a neutral situation (two boxes with different probabilities of containing a red circle)<sup>2</sup> showed that, on average, the participants preferred the high-probability alternative over the low-probability one after 30 trials. As in animal learning literature (Herrnstein, 1961) and learning experiments in economics (Roth & Erev, 1995), the likelihood of subjects' choosing the high-probability box increased to its programmed probability (two thirds of high-probability box choices). Moreover, we found that preference responses did not always correlate with awareness of the different frequencies or probabilities involved; half of the 60 participants preferred the high-probability box most of the time, but their frequency estimations were the same for both boxes.

Experiment 1 showed that in a more complex and counterintuitive situation (as is the case of the MHD) there was also an increase in selecting the high-probability alternative (i.e., to switch). Given that the participants preferred to stick with the chosen alternative (although their intuition supported the idea that the two alternatives were equally likely), the observed increase seemed to reflect that they had indeed learned that there was a difference in probability. Nevertheless, only 33% of the participants were aware of the advantages of switching over sticking, and still fewer (24%) gave the correct win frequencies for 10 switching and 10 sticking trials.

However, the informal verbal reports obtained in Experiment 1 suggested that learning occurred in a way sim-

ilar to that of a standard probability discrimination paradigm. That is, our correct participants noticed the difference between switching and sticking responses (some of them estimated the win frequencies in both situations quite accurately) but were not able to explain it. In other words, the correct participants in Experiment 1 did not learn the mathematical structure that produces the correct answer (i.e., switching). They only noticed that one option (switching) led to more wins than did the other (sticking), just as the “aware” participants in the pilot study explained above noticed that one box had twice as many red circles as the other.

Verbal reports from Experiment 2 were clearer on this point. Most of the participants in the standard version realized that in the experimental situation there was a greater chance of winning after switching than after sticking. But their reasoning (except for that of the two participants who reasoned correctly) showed that their initial intuitions had not changed. The most frequent answer was that the probability observed was different due to chance factors, but they thought that the winning probability should be the same (.5 for switching and .5 for sticking). Incorrect reasoners in the adversary version gave similar arguments. Hence, the frequency information obtained from the 18 repetitions of the game (and the 45 in Experiment 1) was not enough to change the incomplete representation responsible for this strong erroneous intuition.

Experiments 2 and 3 also showed that presenting the dilemma as a game between two adversaries (decision maker and informant) significantly improved the participants’ reasoning. One of the principal differences between the standard and the adversary conditions is in terms of the transparency of the underlying assumptions (Falk, 1992). In the case of the MHD (the standard version adapted in Experiment 2), one has to take for granted the honesty of the experimenter (i.e., that there are no tricks) and that the card to be shown is not selected at random. On the first point, some participants were not convinced that the experimenter was entirely to be trusted: Some of them suspected that he or she might cheat. The second point was basically not taken into account. As the three cards were presented separately in three independent holes, from the point of view of the participants there were three different sets. Once the experimenter showed a non-ace card, the decision had to be made between two “equally probable” alternatives (i.e., according to the equiprobability principle of the MMT; Johnson-Laird et al., 1999).

Unlike the standard version, the adversary version presented the first assumption more clearly (both the informant and the decision maker were naive to the aim of the experiment), and the two sets to be considered were consistent with the game structure: one card for the decision maker [ $p(\text{ace}) = 1/3$ ] and two cards for the informant [ $p(\text{ace}) = 2/3$ ]. As this information was more transparent, many more participants reasoned according to the underlying mathematical structure, even from the beginning of the game; 18% of all the participants in Experiment 2 gave the correct argument from the beginning of the experiment. So, they were able to represent the problem cor-

rectly without any training (e.g., in the form of frequency data).

Experiment 3 replicated this result: 25% of the participants in the adversary game context (without any training and without explicit representation of the different possibilities) reasoned correctly. Experiment 3 also showed that the activity of representing explicitly the different possibilities increased the emergence of correct reasoning (correct responses increased from 25% to 60%), but only in the adversary game context. In the standard context, the explicit representation of the different possibilities had no effect (i.e., none of the participants in this condition was able to give the correct argument). Explicit representation seemed to help to represent more clearly the advantage of the informant in the adversary context: After the elimination of one card, the informant still had two possibilities out of three of having the ace. This effect seemed to work only when the two subsets or partitions were presented explicitly by means of the two players. But explicit representation did not help when the two subsets were presented more implicitly (standard version).

Taken together, our results seem to support the proposal of Girotto and Gonzalez (2001, 2002) that one of the crucial determinants of correct reasoning is the ability to represent relevant sets or partitions. As was stated above, the adversary version presented this information more explicitly than did the standard one: The two relevant sets were represented by means of two adversary players (decision maker and informant). Another important difference between the versions may be related to motivation or personal involvement in the experiment. In the adversary version, the participants were more aware of the adversary’s probability of winning. But, as we discussed later, actual experience with the game was not necessary in order to achieve correct reasoning. The efficiency of the adversary context in helping to build a more complete mental representation could be understood better by means of the perspective effect and its corresponding representation of the sets of possibilities. Whereas, in the standard version, the participants build mental models from only one perspective (either that of the guest or that of the contestant), in the adversary version the participants build mental models more easily from both perspectives: that of the decision maker and that of the informant.

With respect to the differences between correct and incorrect reasoners in Experiment 2, it is worth noting that the correct reasoners (in both MHD versions) switched more often than did their incorrect counterparts, especially in the first block of the experiment. Although this tendency was stronger in the case of the participants who reasoned correctly from the beginning, a similar pattern was observed for the remaining correct reasoners (the switching proportion was always higher than .5, except in the case of 1 correct reasoner who switched three times during nine trials). In some respects, the greater tendency to switch shown by the correct reasoners may be compatible with the frequency format accounts (Cosmides & Tooby, 1996; Gigerenzer & Hoffrage, 1995). Participants

with a greater number of switching experiences (frequency data) would have more opportunities to change their initial (and incomplete) representation. But, as the results from our three experiments have shown, this frequency information was not enough to produce the correct reasoning.

In conclusion, the fact that reasoning performance was very similar for participants in both roles—decision maker and informant (Experiment 2)—and the fact that no experience with the game was needed to change the initial intuition (Experiment 3) suggest that one important determinant for the emergence of correct reasoning is the completeness of the underlying representation rather than direct experience with the game. Moreover, Experiment 3 showed that the activity of imagining and representing explicitly the different possibilities had a strong effect on the accuracy of probabilistic reasoning. This result is also consistent with the mental model approach (Johnson-Laird et al., 1999) in the sense that the activity of representing explicitly the possibilities involved seems to permit the representation not only of what is true (truth principle) but also of what is possible and, accordingly, to be able to reason extensionally. Similar results have been observed in deductive reasoning experiments (i.e., Santamaría & Johnson-Laird, 2000) in which problems that facilitate the building of the different models by means of specific cues also allow subjects to overcome erroneous inferences.

## REFERENCES

- BAR-HILLEL, M., & FALK, R. (1982). Some teasers concerning conditional probabilities. *Cognition*, **11**, 109-122.
- 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.
- EISENHAUER, J. G. (2000). The Monty Hall matrix. *Teaching Statistics*, **22**, 17-20.
- FALK, R. (1992). A closer look at the probabilities of the notorious three prisoners. *Cognition*, **3**, 197-223.
- FRIEDMAN, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *American Economic Review*, **88**, 933-946.
- GIGERENZER, G., & HOFFRAGE, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, **102**, 684-704.
- GIGERENZER, G., & MURRAY, D. J. (1987). *Cognition as intuitive statistics*. Hillsdale, NJ: Erlbaum.
- GIROTTI, V., & GONZALEZ, M. (2001). Solving probabilistic problems: A matter of information structure and question form. *Cognition*, **78**, 247-276.
- GIROTTI, V., & GONZALEZ, M. (2002). Chances and frequencies in probabilistic reasoning: Rejoinder to Hoffrage, Gigerenzer, Krauss, and Martignon. *Cognition*, **84**, 353-359.
- GRANBERG, D. (1999). A new version of the Monty Hall dilemma with unequal probabilities. *Behavioural Processes*, **48**, 25-34.
- GRANBERG, D., & BROWN, T. A. (1995). The Monty Hall dilemma. *Personality & Social Psychology Bulletin*, **21**, 711-723.
- GRANBERG, D., & DORR, N. (1998). Further exploration of two-stage decision making in the Monty Hall dilemma. *American Journal of Psychology*, **111**, 561-579.
- HERRNSTEIN, R. J. (1961). Relative and absolute strength of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, **4**, 267-272.
- JOHNSON-LAIRD, P. N., & BYRNE, R. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.
- JOHNSON-LAIRD, P. N., LEGRENZI, P., GIROTTI, V., SONINO-LEGRENZI, M., & CAVERNI, J. P. (1999). Naive probability: A model theory of extensional reasoning. *Psychological Review*, **106**, 62-88.
- KAHNEMAN, D., & TVERSKY, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, **3**, 430-454.
- KUDADJIE-GYAMFI, E., & RACHLIN, H. (2002). Rule-governed versus contingency-governed behavior in a self-control task: Effects of changes in contingencies. *Behavioural Processes*, **57**, 29-35.
- LANGER, E. (1975). The illusion of control. *Journal of Personality & Social Psychology*, **32**, 311-328.
- PIATTELLI-PALMARINI, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds* (M. Piattelli-Palmarini & K. Botsford, Trans.). New York: Wiley.
- ROTH, A. E., & EREV, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games & Economic Behavior*, **8**, 164-212.
- SANTAMARÍA, C., & JOHNSON-LAIRD, P. N. (2000). An antidote to illusory inferences. *Thinking & Reasoning*, **6**, 313-333.
- SHIMOJO, S., & ICHIKAWA, S. (1989). Intuitive reasoning about probability: Theoretical and experimental analyses of the "problem of three prisoners." *Cognition*, **32**, 1-24.
- TVERSKY, A., & KAHNEMAN, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, **90**, 293-315.

## NOTES

1. *The three prisoners problem* is as follows:

Three men, A, B, and C, were in jail. A knew that one of them was to be set free and the other two were to be executed, but he didn't know who was the one to be spared. To the jailer, who did know, A said, "Since two of the three of us will be executed, it is certain that either B or C will be, at least. You will give me no information about my own chances if you give me the name of one man, B or C, who is going to be executed." Accepting this argument after some thinking, the jailer said "B will be executed." Thereupon A felt happier because now either he or C would go free, so his chance had increased from  $\frac{1}{3}$  to  $\frac{1}{2}$ . This prisoner's happiness may or may not be reasonable. What do you think?

2. A computer program simulated two closed boxes which, when the participants clicked on them, showed either a red or a white circle. The participants had to learn to discriminate between two probabilities: one of the boxes showed red circles two thirds of the time; the other, one third of the time.

**APPENDIX**  
**English Translation of the Catalan Texts**  
**Presented to the Participants in Experiment 3**

---

**Monty Hall (Adversary Version) Questionnaire**

Imagine a card game between two players. The game involves three cards; only one of them is an ace. One of the players is the decision maker and the other the informant. The goal of the game is to draw the ace as many times as possible. At the beginning of each game, the decision maker chooses a card without seeing it. The informant keeps the other two.

\*Imagine that the three cards are the ace, the 7, and the 8 of diamonds. Try to represent the different possibilities of distribution of the three cards between the two players:

	Decision Maker	Informant
Possibility 1		
Possibility 2		
Possibility 3		
Other?†		

Item 1a. What is the probability of the decision maker having the ace?

Item 1b. What is the probability of the informant having the ace?

Besides having more cards, the informant has to show one card different from the ace to the decision maker. So, the informant has to inspect his or her cards and show one non-ace card to the decision maker.

\*Which card will the informant in each of the possibilities previously stated show?

Possibility 1  
 Possibility 2  
 Possibility 3  
 Other?

In how many of these possibilities would the decision maker finally have the ace? †

Once the informant has shown a non-ace card:

Item 2a. What is the probability of the decision maker having the ace?

Item 2b. What is the probability of the informant having the ace?

Imagine now that after the informant has eliminated a non-ace card, the decision maker can choose between sticking with his or her initial card and changing it for the one that the informant is still hiding.

Item 3. What should the decision maker do?

- a) Switch the card
- b) Stick with his or her initial card
- c) Either of the above options. The chances are even.

Item 4. Justify your answer.

**Monty Hall (Standard Version) Questionnaire**

Imagine a TV game between the host and a contestant. The game involves three envelopes numbered from 1 to 3; only one of them contains a valuable prize (e.g., a 5,000 check). The other envelopes contain nothing. The contestant has to choose one of the envelopes by its number. Try to represent the different possibilities of the distribution of the prize in the different envelopes.

	Envelope 1	Envelope 2	Envelope 3
Possibility 1			
Possibility 2			
Possibility 3			
Other?			

Item 1a. What probability does the contestant have of choosing the check?

Item 1b. What probability do the remaining envelopes have of containing the check?

The host, who knows the location of the prize, opens one of the nonselected envelopes to show that it contains nothing.

---

**APPENDIX (Continued)**

---

Imagine that the contestant chooses Envelope 1. In each of the previously stated possibilities, which envelope will the host open?

Possibility 1

Possibility 2

Possibility 3

Other?

In how many of these possibilities will the contestant's envelope contain the check?

After the host has opened an empty envelope:

Item 2a. What probability does the one selected by the contestant have of containing the check?

Item 2b. What probability does the remaining, still closed, one have of containing the check?

Imagine now that the contestant can choose between sticking with his or her initial envelope and changing it for the one still closed.

Item 3. What should the contestant do?

a) Switch the card

b) Stick with his or her initial card

c) Either of the above options. The chances are even.

Item 4. Justify your answer.

---

Note—The texts between the asterisk (\*) and the dagger (†) were included in the explicit representation condition only. The asterisk and the dagger did not appear in the text presented to the subjects.

(Manuscript received June 11, 2002;  
revision accepted for publication February 21, 2003).