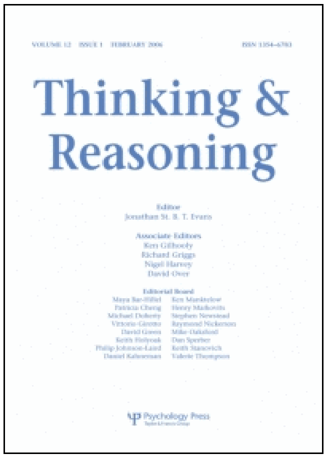

Thinking & Reasoning
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713685607

Let's Make a Deal: Quality and availability of second-stage information as a catalyst for change

Jeffrey N. Howard [a]; Charles G. Lambdin [a]; Darcee L. Datteri [a]
[a] Wichita State University, Wichita, KS, USA

Online Publication Date: 01 August 2007
To cite this Article: Howard, Jeffrey N., Lambdin, Charles G. and Datteri, Darcee L. (2007) 'Let's Make a Deal: Quality and availability of second-stage information as a catalyst for change', Thinking & Reasoning, 13:3, 248 - 272
To link to this article: DOI: 10.1080/13546780600848049
URL: http://dx.doi.org/10.1080/13546780600848049

THINKING & REASONING, 2007, 13 (3), 248 – 272

Psychology Press
Taylor & Francis Group

Let's Make a Deal: Quality and availability of second-stage information as a catalyst for change

Jeffrey N. Howard, Charles G. Lambdin, and Darcee L. Datteri
Wichita State University, Wichita, KS, USA

The Monty Hall Problem (MHP), a process of two-stage decision making, was presented in atypical form via a custom software game. Differing from the normal three-box MHP, the game added one additional box on-screen for each game—culminating on game 23 with 25 on-screen boxes to initially choose from. A total of 108 participants played 23 games (trials) in one of four conditions; (1) "Vanish" condition—all non-winning boxes totally removed from the screen; (2) "Empty" condition—all non-winning boxes remain on-screen, but with an "empty" label on them; (3) "Steroids" condition—all non-winning boxes removed from the screen, with initially chosen box becoming 25% larger; (4) "Steroids2" condition—all non-winning boxes removed from the screen, box not currently chosen becomes 25% larger. Results indicate second-stage on-screen presence of boxes influences switching; with their absence having the opposite effect. Size manipulation appears to elicit demand characteristics resulting in indeterminate influence.

This investigation explores choice patterns within the popular cognitive illusion known as the "Monty Hall Paradox", also known as the "Monty Hall Problem", "Monty Hall Dilemma", or "Three Box Problem". The Monty Hall Paradox (MHP), named after the game-show host from the popular TV show "Let's Make a Deal", has for years been the subject of heated academic controversy. The MHP was first presented by prolific science writer Martin Gardner in 1959 (Piatelli-Palmarini, 1994). Somewhat later in 1976, it appeared in an article published in the journal *American Statistician*, and again in 1990 – 1991 in a series of articles in *Parade* magazine (Piatelli-Palmarini, 1994; see also Granberg & Dorr, 1998). A popular topic in academic circles in the 1990s, the MHP has been extensively

Correspondence should be addressed to Jeffrey N. Howard, Wichita State University, Department of Psychology, 1845 Fairmount, Wichita, Kansas, USA 67260.
E-mail: jnhoward@wichita.edu

http://www.psypress.com/tar DOI: 10.1080/13546780600848049

analysed using a number of methods, including computer simulations and decision trees (Granberg & Dorr, 1998). As testimony to the MHP's counterintuitiveness, even some professional statisticians and logicians deny that the conclusion can be correct (Piatelli-Palmarini, 1994).

The paradigm of the MHP proceeds very parsimoniously with the standard presentation of the problem as follows (the game show used doors; most experimenters use boxes): there are three boxes lined up on a table, and all three of them are exactly alike. Inside one of the boxes is a prize, let's say, a \$100 bill. The other two boxes are empty. To get the prize, all one has to do is guess which box it is in. The participant makes her or his choice, pointing to the desired box. The experimenter then opens one of the other two boxes, revealing it to be empty and stating that there are two boxes left; the prize is in one of the two remaining boxes. The experimenter then asks the participant if they would like to stay with their initial choice, or switch to the other box (this is known as a "two-stage" decision). According to this aforementioned standard presentation methodology, as studies have indicated, participants tend to stick with their initial choice (Bown, Read, & Summers 2003). (For a thorough description of why it is in fact the correct choice to switch, please see the Appendix.)

The reasoning behind the strategy to remain with one's initial choice is not clear—perhaps people tend to stick with their initial choice in demonstration of some type of overriding personal confidence in their initial decision. A priori to the initial decision, the box that actually contains the prize is entirely random; similarly, the box that the participant initially picks is random. Granberg and Dorr (1998) point out that a common error participants make is to assume that the box the experimenter reveals next is revealed at random. If this were true, then the decision as to whether participants should switch or remain with their initial choice would indeed be arbitrary. Ultimately, this confusion could perhaps be clarified simply by having the experimenter state in the presentation: "I will now open a box I know does not contain the prize, and show you that it's empty".

Another suggested hypothesis as to why people tend to stick with their initial choice is a suggested participant "control bias". This cognitive bias, it is suggested, would operate in the following way: when people make their initial choice, the action of making the choice creates in their minds the illusion of having some type of control over the outcome. When people are then asked if they would like to stay or switch, they prefer to stay because switching would remove this perceived "control over the situation" created by the action of having made the choice. Granberg and Dorr (1998) found evidence supporting this hypothesis in an MHP research study which found that participants switched a significantly greater number of times ($p < .001$), when someone else made the initial choice for them. Interestingly, although this led to greater switching initially, it did not lead to greater insight into the task over the total number of games, as the total number of times participants switched was still only 57% (Granberg & Dorr, 1998).

It has also been demonstrated that revealing all but one incorrect box serves to make the MHP less ambiguous and makes the need to switch more salient as the total number of boxes is increased; however, increasing the number of boxes while only revealing one incorrect box (adding boxes yet still only opening one and showing it to be empty) makes the situation more ambiguous and encourages staying (Granberg & Dorr, 1998).

Additional studies have investigated the MHP beyond its "normal" presentation state of three boxes—in line with the present experimental inquiry via investigating the influence that supplemental information beyond that of the standard MHP dilemma might have. Franco-Watkins, Derks, and Dougherty (2003) manipulated the number of prizes and doors to determine how participants interpreted the problem space of the MHP. Their findings indicate that participants will partition the probability judgement to reflect the number of prizes over the number of unopened doors. Similarly, the aforementioned study by Granberg and Dorr (1998) included an experimental manipulation where it was found that varying the number of doors, and the incorrect number of doors revealed, had no effect on the initial propensity to stick, but did show a strong effect across all experimental trials. Granberg (1999) also investigated the potential outcomes if participants in a four-door MHP presentation either (1) decided in their mind to switch *before* any two of the non-winning doors are revealed to them—known as the *unconditional* approach; or (2) implement their decision process to switch or stay only *after* having any two of the non-winning doors revealed to them—known as the *conditional* approach. Each of these approaches results in the same solution probability for the second stage: .75 probability of picking the winning box by implementing the switching strategy. What is most important about this design implemented by Granberg is its association with the present study, and how the current study uses an increasing number of boxes presented on each trial, in an attempt to influence the decision-making process of participants.

What the current study accomplishes, with respect to this aforementioned Granberg four-door paradigm, is to attempt to force a greater tendency towards thought processes of the earlier-mentioned *conditional* nature from participants. This is accomplished by participants eventually recognising their decreasing odds of picking the winning box on their first choice, due to the great disparity between any one box they *could* choose, and *all* available boxes on the screen—the number of which, in the current investigation, becomes greater on each trial. The aforementioned is critical to the present study: if participants never implement an unconditional approach, then they are reflexively reacting to the first stage of the problem and they do not accurately develop a solution strategy. The current study attempts to reveal this—that developing "insight" into implementation of an unconditional first-stage strategy can be influenced by supplemental information (or lack thereof), and how it is presented to participants across the second stages of their experimental trials.

The current study primarily embeds itself in the lore of MHP research via its focus on the second-stage decision-juncture dilemma facing the player, and the arrival at and manipulation of the universal second-stage point in the MHP that can be deemed the "probability contrast point"—the contrast between the initial choice that has been made, and the quantity of that which remains in the form of alternatives not chosen. This decision point is directly related to that known as the "uniformity belief" (Bar-Hilel & Falk, 1982; Falk, 1992)—the belief that uniform distribution of sample space probability is to be made across all possible alternatives. This probability contrast point is theorised to be the application point of the uniformity belief, and that any occurring belief in "equiprobability of alternatives" could be dependent on the magnitude of the contrast, the physical qualities of the alternatives, and their second-stage visual presence or absence.

In a study relating to second-stage inference, Goldstein, Gigerenzer, and the ABC Research Group (1999) found that the operation of a "recognition heuristic" facilitates the ability to infer within the second stage—an outcome they deem to be influenced by the fact that at times "knowing less is more"—a heuristic endorsing the limiting of available information so as to simplify the situation and minimise confusion. Additionally, this "less is more" heuristic, in differing from a "recognition heuristic", is driven by the influence of readily available cues whereby the most valid available cue might be used (take the best; one-reason decision making) given the failure of any recognition process. Gigerenzer, Todd, and the ABC Research Group (1999) indicate that the relationship between the number of objects and number of cues can influence such a take-the-best one-reason decision performance. Conversely, with respect to the aforementioned second-stage inference and cue availability, the current study attempts to reveal a theorised pervasive influence on the part of non-winning un-chosen alternatives; in particular, influence derived from the *quality* and *availability* of un-chosen alternatives as a driving force in correct inference, despite posterior second-stage knowledge of their devalued nature.

Other paradigms endorse problem structure as a critical factor in participant success for the MHP. Kraus and Wang (2003) have shown that by not revealing an "opened door" or a non-winning choice one can eliminate introduction of any restrictions on participants' mental simulation of the game as they attempt to solve the problem.

Likewise, problem structure devoid of opportunity to infer causality, according to Burns and Wieth (2004), emphasises participant failure to understand implications of causal structure via what they deem is a "collider structure", whereby two autonomous factors of causation influence a sole outcome. Their series of MHP experiments using the random choice of a boxer, a fight between the two un-chosen boxers, and the subsequent choice to switch from one's chosen boxer to the other boxer for a second fight, not only provided conceptual insight into causality, but yielded increased correct participant responses to questions regarding processes involved in the MHP. In addition, they showed that participant training on understanding the implications of causal structure yielded transfer results of a positive nature to a standard version of the MHP.

In a presentation of the MHP by Friedman (1998) students used real money across a 10-trial sequence of the standard MHP that addressed solution accuracy as a function of group differences such as age, gender, college major, previous statistics class experience, and upper vs lower division students. According to Friedman's observations, the resulting responses did not begin to approximate rational decision making. His indication is that the majority of the time participants made irrational choices in the face of an equally available rational option—and did so with clear opportunity to familiarise themselves with the task and uncover the rational decision. Such an outcome, with respect to the present study, is unique to the standard three-door MHP in a highly specific way: the physical presence of revealed non-winning entities within the second stage. Our hypothesis is that there exist second-stage factors unique to the standard MHP that in fact can influence rational or irrational participant choice—the mere *presence* of a revealed non-winning entity or entities—and that a comparison between such second-stage "non-winning boxes present vs non-winning boxes absent" conditions could potentially reveal an important catalyst of the perseveration–confusion factor of the standard MHP.

Aaron and Spivey (1998) addressed the presentation format of the MHP by demonstrating the importance of framing the problem in a specific format: either the "probability format", whereby the concept of probability is used and percentage values are mentioned (e.g., the probability of having breast cancer per 1000 women is 1%); or the "frequency format", which merely consists of the mentioning of comparison quantities as numbers (e.g., "10 out of every 1000 women have breast cancer"). Their findings indicated that information presented in the frequency format clearly bolsters Bayesian mathematical understanding driving the three-doors problem. Likewise, the current study presents just such a scenario—frequency format information in the form of a computer game that presents boxes on-screen, allowing direct contrast comparison of one's alternative choices with one's current choice state—devoid of mention of probabilities, percentage values, or requests for comparison using mathematical terminology.

Ultimately, it is proposed that there are additional supplemental information influences that prove advantageous to solving the MHP in the second stage. It is theorised that there is an influence via the visible presence of these non-winning entities, which can be shown as a trend across trials (the aforementioned "probability contrast") that is gradually magnified by adding and revealing one additional non-winning entity for each trial across a series of MHP presentations. Additionally, and perhaps even more importantly, it is theorised that the *lack* of this visual exposure to non-winning entities during the second-stage process would prove detrimental to any participant benefit (conscious or unconscious) from this probability contrast—thus countering the "knowing less is more" recognition heuristic benefit as indicated by Goldstein et al. (1999).

Simply put, for those lacking insight into the standard MHP, the most common and pervasive error in judgement at the probability contrast point is likely the earlier-mentioned uniformity belief (Bar-Hilel & Falk, 1982; Falk, 1992) that one has a 50-50 chance of winning with one's current choice merely because there are two boxes remaining. For example, the pervasiveness of the uniformity belief is so strong that hundreds if not thousands of letters propagating this incorrect belief (many were from university mathematicians and professors) arrived in response to publication of the correct answer to the "three-prisoners problem"—a parallel two-stage decision paradigm to the MHP—by Marilyn vos Savant in her column (2 December 1992) in *Parade* magazine (Falk, 1992; Frazier, 1992).

Thus the current study proposes two primary questions regarding pervasiveness of the uniformity belief, and the inferential value of information at the probability contrast point within the MHP second stage: First, in an MHP situation where one new box is added on each trial over a series of 23 trials (begin trial 1 with three boxes, trial 2 with four boxes, trial 3 with five boxes, and so on), how many revealed non-winning boxes would be required before the MHP becomes transparent?—that is, what is the required "magnitude" of the probability contrast between the current choice and visible non-winning choices that is required to facilitate greater implementation of a switching strategy? A simple calculated ratio represents such a magnitude—the ratio of 1 (the desired choice) to the total number of revealed second-stage non-winning boxes presented for any one trial ($1/n$). Such a ratio additionally represents the probability that one would pick the winning box with one's first choice prior to entering the second-stage—e.g., 1 out of 3 or .33 for the typical three-door MHP.

The second question asks what influence does manipulation of the *quality* or *visual access* to second-stage information have over effective solution to the MHP?

It is hypothesised that the probability contrast as an inferential aid could potentially be influenced by manipulation of the actual size (physical quality) of remaining second-stage process elements, as well as the number of these elements remaining visually accessible during the second stage—specifically: (1) that the availability of non-winning items on-screen for the second stage would facilitate solution frequency of the paradigm; (2) that the absence of non-winning items on-screen would hamper solution frequency of the paradigm; and (3) that increasing the size of a second-stage item would facilitate selection or retention of the enlarged item. Table 1 presents a synopsis of the experimental conditions and the nature of the manipulation within the second-stage of each paradigm.

METHOD

Participants

A total of 108 graduate and undergraduate students (91 females and 17 males) from Wichita State University participated and received partial course credit for their participation. Of these, 79 participants were arts and sciences majors, with nursing (25) and psychology (27) being the only majors with more than three representatives. A total of 17 participants did not

TABLE 1
Summary of conditions, second-stage manipulations, and final choice options

Condition	*Initial on-screen presentation*	*Second-stage manipulation*	*Choice option*
"Empty"	Begin first trial with 3 boxes, add one box each trial across 23 trials; e.g., 23rd trial starts with 25 boxes on-screen	Non-winning, un-chosen boxes remain on-screen; label on each box indicates "EMPTY"	Stay with first choice or switch to other box
"Vanish"	IDENTICAL	Non-winning, un-chosen boxes removed from screen	IDENTICAL
"Steroids"	IDENTICAL	Non-winning, un-chosen boxes removed from screen; current participant box choice becomes 25% larger	IDENTICAL
"Steroids2"	IDENTICAL	Non-winning, un-chosen boxes removed from screen; box NOT currently chosen by participant becomes 25% larger	IDENTICAL

indicate a major, and the remaining 12 were education (5) and business (7) majors.

Visual stimulus and experiment design

The study was self-contained within a software program that varied on-screen presentation of information during the second-stage decision-making opportunity. For all four conditions, the presentation of on-screen stimuli at the beginning of each game was identical and consisted of several different on-screen configurations of small grey numbered boxes, the on-screen size of which was .75 inches by .75 inches. All four of the experimental conditions started with the presentation of three on-screen boxes for the first game, with each subsequent game having an initial on-screen presentation of one more box than the previous game. Thus, the first-stage on-screen presentation for any one game within any condition always looked exactly the same for each participant. Hence, if it were game number 2 in any condition, the game would begin with the on-screen presentation of four boxes of equal size—if it were game number 3, then five boxes of equal size and in identical locations would be presented on-screen at the beginning of the game, and so on.

One of the small grey on-screen boxes either contained or did not contain a fictitious \$100 "correct-choice" reward. There were a total of 23 opportunities to win the fictitious \$100 reward within each condition, thus a perfect score for all games attempted would be \$2300. Failure to win on any one game presented a pop-up box with the word "LOSER" on it. Since the first game always started by presenting three boxes, and one box was added to the screen at the beginning of each game, there were 23 games needed to attain the desired on-screen box presentation maximum of 25 boxes, which occurred on the final game of the sequence. Additionally, the available presentation space for the boxes occupied an on-screen area of 6 inches by 6 inches. Figure 1 shows the interface layout of the condition known as "Vanish" as it would appear to a participant during game number 20 of the 23-game sequence.

Manipulation of the presented grey on-screen box stimuli (the independent variable) for the four conditions occurred only after a participant had first clicked on one of the small on-screen boxes, and then clicked on a button called "help me out", which disappeared after being clicked. This "help me out" procedure was designed to assist participants by manipulating all of the "non-winning" boxes in one of four unique ways:

First, in the condition called Empty (Figure 2), after the participant had clicked the "help me out" button during the game, all non-winning on-screen boxes remained on-screen, and were automatically affixed with labels indicating that they were "EMPTY", a process that again left only two

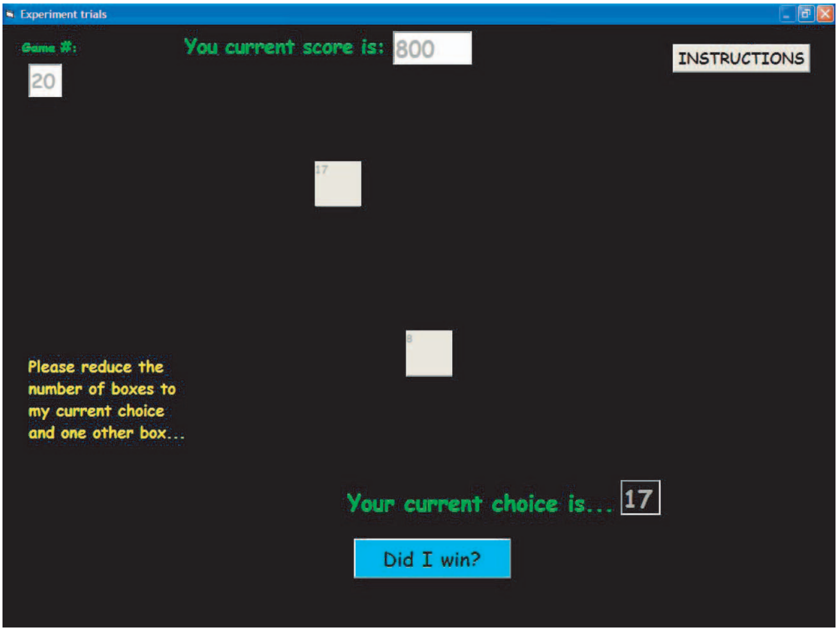

Figure 1. Interface layout for "Vanish" experimental condition on game 20 of the 23-game sequence. To view this figure in colour, please see the online issue of the Journal.

boxes—the one the participant had chosen and one other unlabelled box, one of which contained the fictitious $100 reward.

The box labels were white with red lettering on them and were .75 inches wide and .25 inches high. Thus within the "EMPTY" condition, the independent variable manipulation was the continued presence of all other boxes as on-screen competing information during a choice opportunity, despite the fact that these boxes were clearly labelled as "EMPTY". It was hypothesised that this condition, as the number of boxes increased, would be the most conducive to switching. This Empty condition was the closest analogue of the actual MHP problem with respect to the three conditions tested, in that the non-winning entities remained on-screen and visually accessible to participants during their subsequent second-stage decision making process.

Second, within the earlier-mentioned condition called Vanish (Figure 1), after making their box choice participants would click the "help me out" button, which would subsequently remove from the screen all non-winning boxes, leaving only the participant's initial box choice and one other box. Participants would then be able decide (as was indicated in the instructions for the game) if they wanted to switch from their initial box choice to the

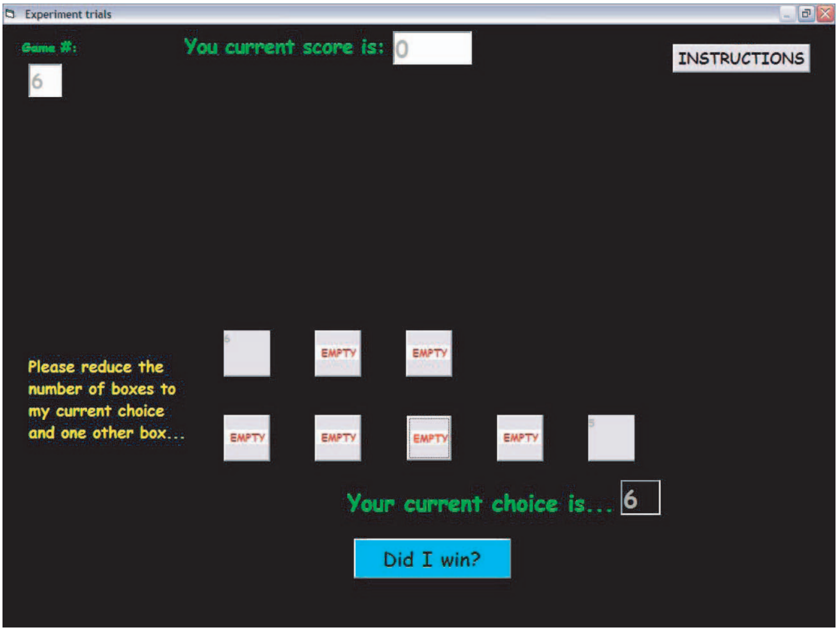

Figure 2. Interface layout for "Empty" experimental condition on game 6 of the 23-game sequence. To view this figure in colour, please see the online issue of the Journal.

other box, or "stay" with their first choice. After making their decision to "stay" or "switch", participants would click on the button called "Did I Win?" so as to find out the status of their decision. Thus for the Vanish condition, the independent variable manipulation was the presence of on-screen competing information during a choice opportunity, which consisted of nothing more than the presence of another box of equivalent size and the opportunity to choose the other box prior to checking the status of their decision.

Third, in the condition called "Steroids", after a participant had clicked the "help me out" button, once again all "non-winning" on-screen boxes were removed from the screen. However, within this "Steroid" condition, the box the participant had chosen would suddenly increase in size to a width of 1 inch and a height of 1 inch (see Figure 3; a 25% size increase).

Fourth, for the condition known as "Steroids2" (the exact opposite of "Steroids"), after clicking the "help me out" button, all "non-winning" on-screen boxes were once again removed from the screen. However, for this "Steroids2" condition, the box not currently chosen (the "other" box) would suddenly increase in size to a width of 1 inch and a height of 1 inch (a 25% size increase).

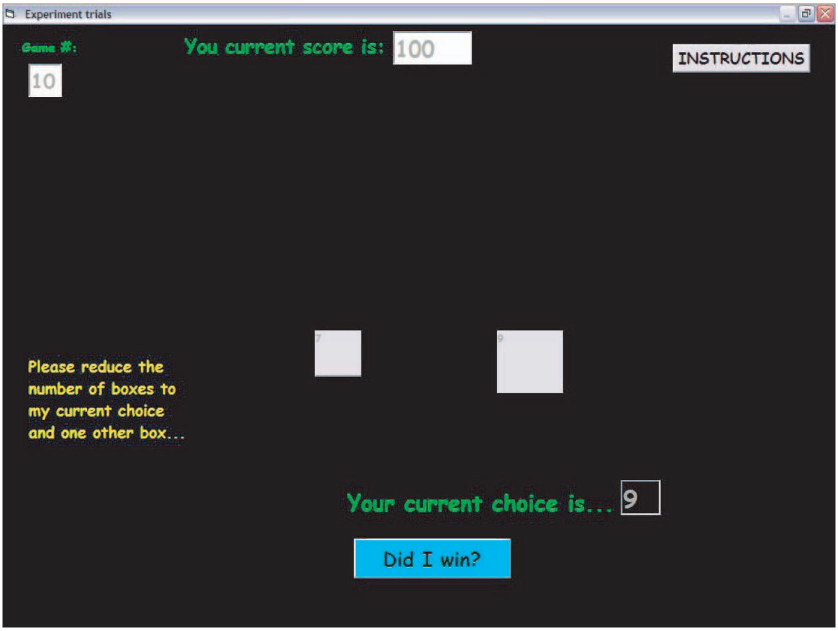

Figure 3. Interface layout for "Steroids" experimental condition on game 10 of the 23-game sequence. To view this figure in colour, please see the online issue of the Journal.

Thus, for these "Steroids" and "Steroids2" conditions, the independent variable manipulation was to create the presence of competing information comprising the sudden increase in size of either one's initial box choice or the "other" box not currently chosen. It was thought that this "enlargement" manipulation would encourage participant desire for the larger box, as it was hypothesised that participants would think it was more likely their probability of winning would be greatest with the larger of the two boxes.

Instructions were continuously available to each participant via a button at the top right of the screen. Once the sequence of games was complete, each participant was informed that they were done and were presented a prompt to proceed through the program which then presented the debriefing for the experiment.

Procedure

Upon entering the lab, participants were seated at a computer station. The experimenter provided the participant with an informed consent document and instructed them to fully review the document before signing. Participants were assigned to conditions via a roll of a die. Ultimately, it was necessary to randomly deselect several cases from three of the groups for the analysis, so as to arrive at equal participant representation for each condition.

After participants had completed their informed consent document, the appropriate condition was started on their machine, and they were told that they were going to be playing a game and were to try to accumulate as many points as possible. They were then instructed to simply fill out the demographics form, click the "submit" button, and then read and follow the directions within the program. They were also told that the experiment would last approximately 15 minutes. Upon completion of the experimental gaming trials, all participants were presented with an on-screen debriefing explaining the nature of the experiment and were thanked for their participation.

RESULTS

All demographics information and game response data were reported to an Excel file. Responses for each game outcome were then coded as to whether or not a "switching" strategy was employed at the end of the game—a "1" was assigned if a switching strategy was employed, and a "0" was assigned if a "stay" strategy was employed.

It is important to mention that the probability that a participant would implement a "switching" strategy was .50, which is the simple probability that they would either choose to "switch" (coded as a "1") or "stay" with their original choice (coded as a "0") at the end of any one game. Thus, if the overall mean for any single game across the sequence of 23 games (as compiled from all participant choices) had a value far above that of .50, then it was clear that on this particular game in the sequence, for some reason, participants were applying the switching strategy more than would be indicated by simple binomial probability—in essence, perhaps participants had "figured out" that switching was a winning strategy. These means, along with frequencies for application of the switching strategy, can be seen in Table 2.

Albeit trials prior to attaining a .50 criterion do encompass any learning that may have occurred up to this point, a .50 criterion is lacking in its ability to reveal the intricacies of any such learning. However, a simple trend across trials for frequency of application of the switching strategy can and does provide such insight. Figure 4 provides a visual representation for frequency of application of the switching strategy (Table 2) for the four groups as a function of trials 1–23.

Of particular interest was the ability for chi-square analyses to reveal significant differences within the individual groups when comparing each group's collapsed aggregate of the first five and last five trials (1–5

TABLE 2
Switching frequencies, switching percentages, and aggregate totals for each game

Game #	*Empty*	*%*	*Vanish*	*%*	*Steroids*	*%*	*Steroids2*	*%*
1	10	0.37	3	0.11	5	0.19	8	0.30
2	12	0.44	3	0.11	9	0.33	8	0.30
3	7	0.26	7	0.26	11	0.41	6	0.22
4	13	0.48	10	0.37	14	0.52	14	0.52
5	14	0.52	8	0.30	14	0.52	12	0.44
6	12	0.44	9	0.33	18	0.67	12	0.44
7	14	0.52	10	0.37	20	0.74	14	0.52
8	14	0.52	9	0.33	20	0.74	16	0.59
9	16	0.59	12	0.44	17	0.63	14	0.52
10	15	0.56	13	0.48	18	0.67	16	0.59
11	20	0.74	14	0.52	20	0.74	13	0.48
12	21	0.78	13	0.48	22	0.81	13	0.48
13	15	0.56	13	0.48	20	0.74	14	0.52
14	23	0.85	13	0.48	22	0.81	15	0.56
15	19	0.70	16	0.59	22	0.81	16	0.59
16	21	0.78	12	0.44	21	0.78	16	0.59
17	24	0.89	14	0.52	23	0.85	15	0.56
18	23	0.85	13	0.48	22	0.81	11	0.41
19	24	0.89	14	0.52	21	0.78	15	0.56
20	22	0.81	17	0.63	22	0.81	19	0.70
21	20	0.74	15	0.56	23	0.85	18	0.67
22	23	0.85	15	0.56	22	0.81	15	0.56
23	23	0.85	12	0.44	21	0.78	19	0.70
Totals	405	0.65	265	0.43	427	0.69	319	0.51

Shading indicates the point where mean value first exceeds binomial probability.

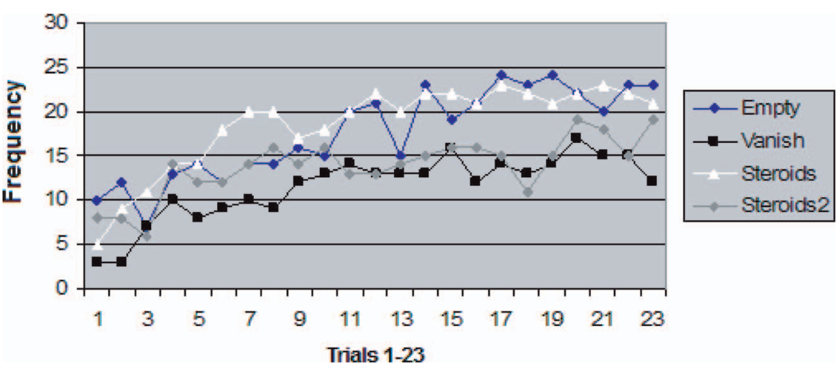

Figure 4. Switching frequency as a function of trial for all trials 1 – 23 (see Table 2).

and 19 – 23 from Table 2) with respect to frequency of application of the switching strategy. The resultant chi-square statistics and probability values for this within-groups aggregate trials comparison can be seen in Table 3.

A chi-square analysis between groups using the collapsed aggregate switching frequencies of trials 1 – 5 and 19 – 23 (Table 2) was also conducted. The results of this comparison, $\chi^2(3, N = 568) = 19.01$, $p < .001$, reasonably concludes that these groups do differ with respect to these collapsed frequency counts for early vs later trials regarding application of switching from one's current choice. Thus the exposure over the course of these trials reveals differences between the groups when frequencies are collapsed and compared. However, this analysis compiles the incremental resolution of the experimental sequences, and a comparison of individual trials through this sequence may yield better insight.

A between-groups comparison subsequently looked at trials 1 – 5 and 19 – 23 individually (non-aggregate) so as glean better insight into the aforementioned resolution of the individual trials making up these early- and later-occurring trial clusters. Interestingly, the between-groups comparison of individual trials 1 – 5 and 19 – 23, do not reveal significant differences between each individual trial for application of the switching choice. Table 2 once again provides the observed frequencies of switching for each individual trial of trials 1 – 5 and 19 – 23 for the four groups, with Table 4 providing the individual chi-square statistics and probability values.

The indication here is that the groups do not differ significantly from one another when one discounts the cumulative effects of exposure to the paradigm across a longer sequence of the trials. This comparison seemingly addresses the base degree of difficulty imposed per trial via the uniqueness of the conditions, and gives insight into long-term differences as revealed by analyses that collapse frequency counts over multiple trials in the sequence. Thus it appears that a comparison of aggregate accumulation of switching choices across several front-end trials vs back-end trials is necessary to reveal any pattern of systematic solution differences contained within different groups.

TABLE 3
Within-groups chi-square statistics and probability values for aggregate trials 1 – 5 vs 19 – 23

Group	*p*	χ^2	*df*	*N*
Empty	$< .001$	18.667	1	168
Vanish	$< .001$	16.962	1	104
Steroids	$< .001$	19.358	1	162
Steroids2	$= .001$	10.776	1	134

Trials 1–5 only, and the manner in which these early trials present themselves performance-wise across the four groups (Table 2, Figure 4), were also analysed with respect to aggregate collapsed frequency counts. The chi-square analysis for this comparison between the four groups, $\chi^2(3, N = 188) = 7.95$, $p = .046$, did arrive at significance. A chi-square analysis of this same structure for trials 1–3 only, $\chi^2(3, N = 89) = 6.235$, $p = .100$, did not yield a significant result, thus bolstering earlier indication that the disclosure of patterned solution differences between groups is likely driven by the trials across the sequence and the cumulative insight gained from such experience.

The global overall frequency totals of switching vs staying for the entire experiment was also of interest. Ultimately, this analysis addresses any significant difference in aggregate application of the switching vs staying choice across all trials between the four groups. The chi-square analysis for all cumulative switching vs staying observed frequencies on all trials (Table 5), $\chi^2(3, N = 2484) = 112.192$, $p < .001$. does indicate that the global accretion of switching vs staying decision frequency varied amongst the four groups over the 23-trial sequence.

TABLE 4
Between-groups chi-square statistics and probability values for individual trials 1–5 vs 19–23

	p	χ^2	df	N
Trial 1	.216	4.462	3	26
Trial 2	.154	5.250	3	32
Trial 3	.593	1.903	3	31
Trial 4	.839	0.843	3	51
Trial 5	.572	2.000	3	48
Trial 19	.292	3.730	3	74
Trial 20	.825	0.900	3	80
Trial 21	.617	1.789	3	76
Trial 22	.388	3.027	3	75
Trial 23	.300	3.667	3	75

TABLE 5
Observed aggregate switching vs staying frequencies from all trials for all four groups

Observed	*Empty*	*Vanish*	*Steroids*	*Steroids2*	*TOTAL*
Switch	405	265	427	319	1416
Stay	216	356	194	302	1068
Total	621	621	621	621	2484

Another of the interesting results in the frequency data was gleaned from individual performances—namely the frequency at which some individuals never won a game over the entire 23-trial sequence. The frequencies of such "unlucky" participants across conditions; Empty = 1; Vanish = 10; Steroids = 2; Steroids2 = 6; $\chi^2(3,\ N = 19) = 10.68$, $p = .013$, form a most interesting result supporting prevalence of difficulty within the Vanish group. It is important to point out here that the total observed frequency count of 19 as distributed across four cells would yield a per-cell expected value of just below 5. However, the performance data alone as seen in Figure 4 and Table 2 for this Vanish condition make clear the fact that this group would likely contain such a population of individuals who had difficulty with the paradigm solution process.

Individual performance data also provided an interesting look at what could be called "maximal performance"—participants who at some specific point began to win on a consistent and continuous basis for the remainder of the trials without losing. Figure 5 provides graphical insight into this trend as it occurred from trials 5 through 20 across the four groups. For example, in the Steroids condition there were only four participants on trial 5 who maintained perfection-of-solution for the remainder of trials (19 wins in a row)—by trial 20, the number of persons who were perfect on the remaining four trials was 18. However, chi-square analyses did not indicate significant differences with respect to these frequencies; it is clear that the groups display patterned disparity regarding potential for sudden and sustained solution insight across the sequence. The trend divergence over trials is a particularly interesting feature—the four groups start out

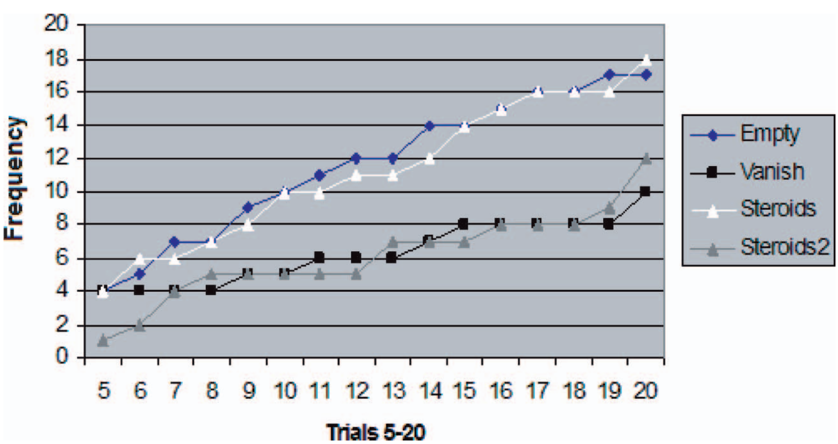

Figure 5. Cumulative frequencies for number of participants reaching "maximal performance" across trials 5–20 for all four groups.

relatively clustered, and then gradually diverge in unique pairings over the 16 trials, an indication that Empty and Steroids groups generate ‘‘maximal performers’’ at a greater rate than do the Vanish and Steroids2 conditions.

Effect sizes comparing differences in performance for individual trials across all four groups were also calculated, as were effect sizes comparing composite overall ‘‘mean switching value’’ as garnered by each group across the 23 trials. The six combinations that compare the effect sizes for the individual trials against one another can be seen in Figure 6.

Additionally, effect size comparisons for the six combinations of composite overall ‘‘mean switching values’’ for each group can be seen in Table 6. It should be noted that, directionally, the experimental minus the control group is utilised here for effect size calculation, thus positive differences favour the experimental group and negative differences favour the control group (‘‘control’’ groups are specified within figure titles).

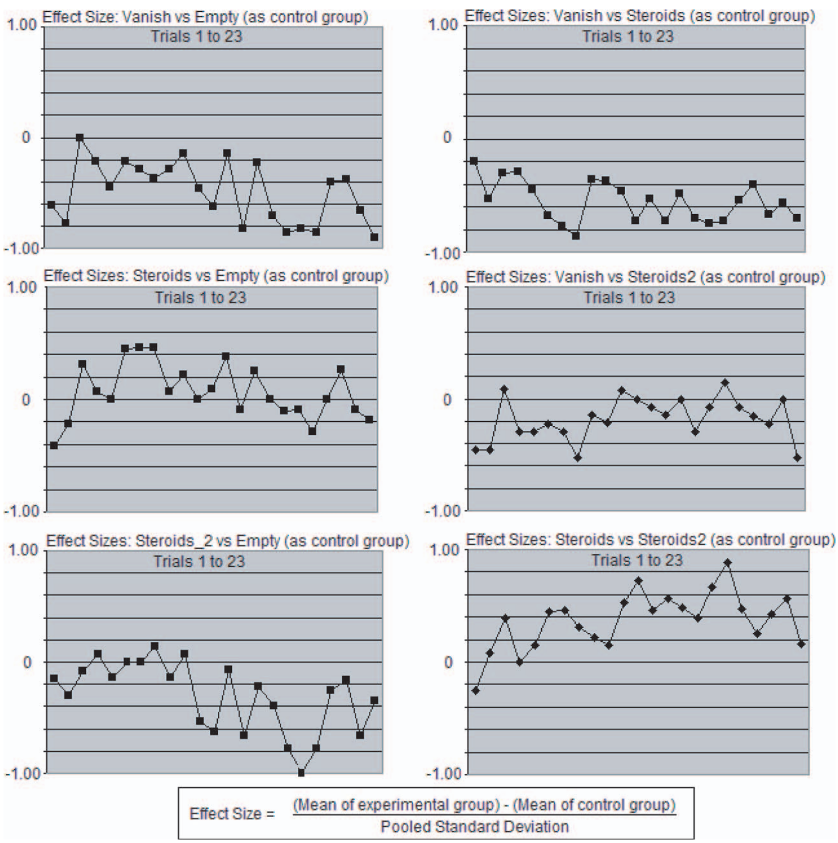

Figure 6. Effect size comparisons for all groups across the 23-trial sequence.

According to Cohen (1988), effect sizes of .2 are considered a "small" effect, .5 as "medium", and .8 as "large".

Next, the analyses looked at the specific differences between the 23 game means generated by each of the four conditions via ANOVA with separate a priori contrast tests—a test sequence that compares the various contrasts between the response scores as presented in Table 2. The between-groups ANOVA result for these scores, $F(3, 88) = 13.532$, $p < .001$, clearly indicates that the groups were different regarding average "switching frequency" scores for all games. An a priori ANOVA contrast test sequence comparing all six one-to-one direct comparisons between groups was also carried out, the results of which can be seen in Table 7.

Of these individual contrasts, only the Empty vs Steroids conditions contrast arrived at an insignificant result—all other a priori contrasts were significant.

Analysis of variance was also conducted so as to look at the differences for each of the four conditions regarding the means garnered on games 4, 5, and 11, which appear as "solution points" where switching frequencies first exceeded a value of .50 (Table 2). The visual indication from Table 2 is that these trials exist as some sort of point whereby participants "caught on" to

TABLE 6
Effect size comparisons for the six combinations of composite overall "mean switching values" for each group

Experimental group	*Control group*	*Effect size*
Vanish	Empty	−0.68
Steroids	Empty	0.13
Steroids2	Empty	−0.44
Steroids	Vanish	0.78
Steroids2	Vanish	0.24
Steroids2	Steroids	−0.54

TABLE 7
Individual a priori ANOVA contrast test comparisons comparing mean switching values garnered across the 23 trials (see Table 2)

Contrast	*t*	*df*	*p*
Empty vs. Steroids	−.655	43.871	.516
Empty vs. Vanish	4.662	40.33	<.001
Empty vs. Steroids2	2.958	37.823	.005
Steroids vs. Steroids2	3.856	39.010	<.001
Steroids vs. Vanish	5.580	41.386	<.001
Vanish vs. Steroids2	2.295	43.408	.027

the switching strategy. An ANOVA comparing these trials across all four conditions was conducted, with the results indicating no significant differences between these groups across the conditions. In addition, a priori contrast tests between all combinations of conditions for these specific games also failed to yield any significant results below the .05 level.

Also of interest was the unusual difference between solution trial mean values (Table 2) as garnered on game 1 for the Vanish vs Empty conditions, and the fact that the only difference between these two conditions within the game 1 second stage was that the Empty condition left the non-winning box on the screen after revealing it. It must be added that the comparison of these two conditions on this particular trial was the only comparison using a pure three-box MHP occurrence that this study had to offer. The result of this comparison, $t(44.674) = 2.294$, $p = .027$, did in fact indicate a significant difference between these two conditions on this first game of the 23-game sequence.

It must also be added that due to occasional violation of homogeneity of variance for the ANOVA tests conducted herein, these results subsequently reflect test statistic values and degrees-of-freedom as per "non-assumption of equal variances" where relevant. Thus a conservative approach to the presentation of results was implemented, albeit the F statistic is known for its robustness despite violations of such assumptions (Lindman, 1974).

DISCUSSION

Generally speaking, the results of this experiment indicate that the MHP, when removed from its pure form as best represented within this experiment by the Vanish condition, is particularly difficult for individuals to solve. The Empty condition was perhaps the closest analogue of the MHP problem, and most purely representative of the actual game show in that although the non-winning boxes were revealed, they remained available as second-stage information that participants could potentially utilise—all for the intent and purpose of influencing participant application of a first-stage unconditional solution approach. The results seemingly indicate that the presence of the non-winning boxes for the second-stage decision process (the Empty condition) did in fact facilitate participant adoption of an "unconditional" probability solution approach for attempts beyond game 7.

Participant accuracy for applying the proper switching strategy in the Vanish condition can be seen in Table 2 to revert to below .50 chance probability at game number 16, then to surpass .50—only to once again fall below .50 on game number 23, the final game of the entire sequence. The observed frequency of application of the switching strategy as seen in Figure 4 corroborates the erratic and poor application of the switching strategy for this Vanish condition, and endorses that there is participant advantage in leaving first-stage information accessible to participants for the second stage of the decision-making process. In addition, the $-.68$ effect size difference (Table 6) seen between the Empty and Vanish groups bolsters support for this result even further.

Thus, for the present experiment, the results indicate that any consideration of alternative possibilities was dependent on the amount and type of information construed by those possibilities during both first and second stages of each trial. These experimental results are a little in line with findings such as those of Tubau and Alonso (2003) who found that correct reasoning seems to have greater dependence on ability to consider different possibilities than on extensive practice with the game. The current experiment shows via results prior to trials 4, 5, and 11, particularly trials $1-3$ for all groups, that a minimum amount of learning must seemingly be attained, and is a necessary condition prior to application of different (but potentially more accurate) possibilities. Were this not the case, the clear pattern of values below .50 and the sudden change to values above .50 at trials 4, 5, and 11, would not likely have occurred as observed.

Additionally, the Vanish condition solution accuracy outcomes and the frequency of application of the switching strategy, particularly beyond trial 11 (Figure 4), appear to reflect nothing more than random chance "practice"-type results across all 23 games. However, what is not precisely known from this experiment is whether early trials led to acquisition of any "ability" to consider different possibilities, and subsequent correct reasoning—or if simple reinforcement of winning after switching on some occasion merely facilitated repetition of the switching strategy independent of any understanding of its efficacy.

With respect to second-stage information presentation (feedback) as catalyst for switching strategy application, Krauss and Wang (2003) emphasise the connection between change in perspective and Bayesian rule structure. Their indication is that calculation of conditional probability of an arbitrary event "A" given a condition "B"—the probability of "A" given "B"—the rule structure requires that one needs to consider inverse conditional probabilities. Thus the cognitive process in assessment of three probabilities for a three-box paradigm is independent of contestant behaviour and is reliant on imagining the behaviour of Monty Hall (the host) for all three possibilities of the paradigm. Furthermore, their indication is that Bayesian problem solution, be it formally based on Bayesian rule structure or intuitively driven, centres on host behaviour rather than on participant behaviour. Ultimately, the indication here is that the change from the perspective of the participant to that of the host (Monty Hall) corresponds to a change from non-Bayesian- to Bayesian-driven thinking. It is important to emphasise that the "host behaviour" (feedback) as presented in the Empty condition was quite different from "host behaviour'' of the Vanish condition and that, ultimately, the calculation of conditional probabilities, via consideration of inverse probabilities, would likely be affected by the presence or absence of adequate feedback information to calculate such inverse probabilities. The salient disparity between the Empty and Vanish condition data supports such a result.

Likewise, the findings of Aaron and Spivey (1998) can be related back to the importance of these feedback (host behaviour) manipulations within conditions of the present study, in that the Empty and Vanish conditions clearly differ with respect to a type of ''frequency'' presentation of information. The Empty condition in essence presents second-stage information that not only contains the Vanish condition as a subset (only two viable boxes remain), but also contains information that provides a procedural solution strategy structure—a structure whereby the presence of the word ''empty'' on the remaining boxes seemingly functions as a passive mention of ''comparison quantities'', which may in turn facilitate ''frequency format'' solution strategy application.

Another interesting result is that of the Steroids vs Steroids2 comparison using effect sizes from Table 6. It was theorised that there could be a demand characteristic that would arise within participants regarding these conditions, causing participants to feel that the experimenter was explicitly tempting them to either retain or choose the box that became larger—thus participants perhaps felt that maintaining control over their destiny meant doing the opposite of this perceived temptation. Essentially, demand characteristics derive from the ''constructivist'' human nature of participants, who desire to find patterns in the data, guess hypotheses, or to satisfy what they perceive to be experimenter-desired responses—and the Steroids and Steroids2 enlargement manipulation appeared to provide ideal conditions for manifestation of such a result. Table 6 provides some indication of this occurrence via the negative effect size difference favouring the Steroids condition in the Steroids vs Steroids2 comparison—a difference indicating that solution accuracy was greater in the Steroids condition. Had participants in both conditions been consistently and rather equivalently implementing the strategy of switching, independent of the perceptual change in the stimulus, the effects size difference would have been very small. The case in point here is that if the theorised demand characteristic ''wins out'' in the Steroids condition, participants would end up applying the effective switching strategy in the end (''A-HA! … my currently chosen box just became larger, the experimenter is tempting me to stay, so I'll switch''), resulting in a high successful solution frequency count for the Steroids condition—which did in fact occur (Figure 4). Similarly, if the theorised demand characteristic ''wins out'' in the Steroids2 condition, participants *would not* end up applying the effective switching strategy (''A-HA! … the other currently UN-chosen box just became larger, the experimenter is tempting me to switch, so I'll stay''), resulting in a low successful solution frequency count for the Steroids2 condition—which also occurred. Ultimately, the unfolding of such a demand characteristic scenario would create an effect size difference favouring the condition whereby the demand characteristic catalysed implementation of the switching strategy—in this case, the Steroids condition.

With respect to any ''standard'' three-box MHP paradigm, the first game within the Vanish and Empty conditions produced statistically significant results whereby the presence of the revealed non-winning box on the screen in the Empty condition seems to have provided some advantage regarding solution accuracy. Given that this is the closest approximation of the pure MHP, this result is very interesting—particularly so in comparison to the other game-1 trial results of the other conditions. In addition, the longitudinal results of the entire Vanish condition endorse the difficulty that participants had with this condition. The absence of visible second-stage information on-screen within the Vanish condition—as opposed to the Empty condition—seems to be detrimental to long-term performance within this paradigm. Ultimately, this outcome was serendipitous and not an intended directive of the study. Further, the question remains as to whether or not the result of such a basic three-box manipulation would be pervasive over future experiments.

The individual performance results, showing the increase in the number of people who ''caught on'' as trials progressed, were also a most interesting finding. This result supplements the Empty vs Vanish switching data in that the number of ''maximal performers'' was disparate among these two conditions—just as their ''switching'' frequency data would suggest. The rate at which the gap gradually grows between the Empty and Vanish groups across the trial sequence (Figure 5) appears very consistent and continuous, and supplements the trend increase in switching strategy by showing that the accrual rate of ''maximal performers'' who are applying the strategy also increases.

It is possible that the nature of learning within the restrictions of this paradigm does not allow for a full revealing of the learning process regarding the conditions—perhaps a set of 100 trials would in fact provide greater insight into such individual performances. It is possible that this disparate divergence, which one sees occurring between the Empty and Vanish conditions across trials, will converge or blend at some point. Clearly the ability to catch-on to the strategy of switching differs between these groups, and perhaps something as simple as a greater number of trials would allow their solution frequency data to eventually equate.

It was also surprising to find that a number of individuals for some reason failed to solve the paradigm at all across the 23-trial sequence. These ''minimal performers'' were predominant in the Vanish group and clearly contributed greatly to the group's comparatively discouraging results. The story that remains untold regarding this performance subgroup is whether or not difficulty of this magnitude would continue to grow disparately between groups were there a much greater number of trials. Additionally, the question looms as to whether or not these individuals simply gave up, didn't care about their performance, were perseverate via personal stubbornness, or perhaps were exuding a demand characteristic to resist winning—that is, intentionally trying to achieve a "maximal failure" rate.

Comparison of the individual trial sequences 1–5 with 19–23 as made between groups was also very revealing, in that it allows for insight into the learning process across the trial sequence—particularly the manner in which these single trials did not yield significant differences, but aggregate collapsed trial frequencies did. The indication here is that learning within this paradigm is not encompassed within any one trial, but rather consists of a process that unfolds across time at a resolution that cannot be captured on a per-trial basis.

One must also consider the fact that avoidance of the "regret" at having made an incorrect decision may be a catalyst or driving factor behind these experimental outcomes as well. According to Loomes and Sundgren (1982) the regret theory rests on both the premise that one will experience the sensations of rejoicing or regret, and that under uncertainty conditions decisions are made in anticipation and consideration of these feelings. Additionally, they indicate that an individual experiencing such feelings, who takes such feelings into account when making decisions, is unlikely to be irrational regarding their decision making. As it relates to the current study, the avoidance of regret would not seem to have a large impact on decision making for participants, due to the artificial nature of the performance reward this study offered. In comparison, it would be relatively easy to imagine an avoidance of regret within a paradigm offering legitimate monetary payout upon winning (reward) and losing (punishment), as opposed to accumulation of points across the gaming sequence. Most importantly, although the regret of decision making can be easily attached to those extrinsically motivating elements society deems to be of higher value (e.g., money, property), it is far more difficult to relate this theory to motivation-to-compete desire-to-achieve intrinsic performance values and desires.

To its advantage was the fact that this MHP "game" itself was internally consistent and provided identical environment of presentation across all participants—a result that would be difficult to achieve, or at the least fraught with more error, if presented using paper and pencil, or within some lesser-controlled open type of environment. The presentation of the game and the subsequent results are without a doubt very straightforward and, given the degree of experimental control offered up by the differing conditions, provide some interesting insight into the presence and absence of supplemental information in a decision-making opportunity scenario of this type.

It must be emphasised that the responses as garnered by this experimental sequence of conditions are in fact expressly unique to this experiment and its conditions. After all, the experiment leaves to the imagination questions such as: What is the influence of colour combinations of supplemental competing information with respect to decision-making processes during such an experimental scenario as this? What influence does the presence of a ''real'' monetary reward have on the concentration and stratagem development by the participant during the decision-making process? What influence would a realistic monetary reward have with respect to stratagem development across genders? There are many more questions such as these that can and do remain regarding the competitiveness of supplemental information on our decision-making processes—all of which should be explored in continuance of this line of research.

Ultimately, and perhaps most importantly, one should never lose sight of the fact that the decision-making process that leads to greatest failure is not the ''incorrect decision'', but rather the process of ''indecision''—the failure to make any decision at all. For here, one has in fact ''decided not to decide''.

Manuscript received 24 May 2005
Revised manuscript received 2 June 2006
First published online 28 September 2006

REFERENCES

Aaron, E., & Spivey, M. (1998). Frequency vs. probability formats: Framing the three doors problem. *Proceedings of the 20th Annual Conference of the Cognitive Science Society* (pp. 13–18). Mahwah, NJ: Lawrence Erlbaum Associates Inc.

Bar-Hilel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition*, *11*, 109–122.

Bown, N. J., Read, D., & Summers, B. (2003). The lure of choice. *Journal of Behavioral Decision Making*, *16*, 297–308.

Burns, B. D., & Wieth, M. (2004). The collider principle in causal reasoning: Why the Monty Hall Dilemma is so hard. *Journal of Experimental Psychology: General*, *133*(3), 434–449.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

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

Franco-Watkins, A. M., Derks, P. L., & Dougherty, M. R. P. (2003). Reasoning in the Monty Hall problem: Examining choice behaviour and probability judgements. *Thinking and Reasoning*, *9*(1), 67–90.

Frazier, K. (1992). "Three Door" problem provokes letters, controversy. *Skeptical Inquirer*, *16*, 192–199.

Friedman, D. (1998). Monty Hall's three doors: Construction and deconstruction of a choice anomaly. *The American Economic Review*, *88*(4), 933–946.

Gigerenzer, G., Todd, P., & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Goldstein, D. G., Gigerenzer, G., & the ABC Research Group (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 37–58). New York: Oxford University Press.

Granberg, D. (1999). A new version of the Monty Hall Dilemma with unequal probabilities. *Behavioural Processes*, *48*(1–2), 25–34.

Granberg, D., & Dorr, N. (1998). Further exploration of two-stage decision making in the Monty Hall dilemma. *American Journal of Psychology*, *111*(4), 561–579.

Kraus, S., & Wang, X. T. (2003). The psychology of the Monty Hall problem: Discovering psychological mechanisms for solving a tenacious brain teaser. *Journal of Experimental Psychology: General*, *132*(1), 3–22.

Lindman, H. R. (1974). *Analysis of variance in complex experimental designs*. San Francisco: W. H. Freeman & Co.

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *Economic Journal*, *92*, 805–824.

Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds*. New York: John Wiley & Sons, Inc.

Tubau, E., & Alonso, D. (2003). Overcoming illusory interferences in a probabilistic counterintuitive problem: The role of explicit representations. *Memory and Cognition*, *31*(4), 596–607.

APPENDIX

Plausible explanation of errant judgement within the Monty Hall Paradox

Within the original “Let’s Make A Deal” dilemma paradigm, a participant’s first choice has a .33 probability of being correct, as do each of the other two choice options. However, after removal of one of the “non-winning” choice options, which now leaves only the participant’s choice and the remaining un-chosen option, many individuals will perceive that their probability of winning has shifted to .50, or 50%. However, this is in fact a probability miscalculation illusion that occurs due to the tendency for the perceived odds to be re-calculated and re-distributed based on an immediate “change of condition” occurrence regarding the number of available options. Thus the participant has errantly re-assessed their chances of success by dividing a probability of 1.0 by the two remaining choices—having failed to recognise that their chances of winning are still based on their initial choice, which was made based on odds calculated using three choice options and not two.

To clarify a little further, the two options *not chosen* always garner a .66 probability between them (.33 each) of containing the correct choice—a probability that remains constant even after revealing one of the two options as a non-winning choice. The .33 probability from the “revealed non-winning option” merely adds to the .33 probability of its non-chosen partner, thus making the participant strategy of “switching” twice as likely (.66) to yield the winning result.