Running head: CAUSALITY IN REASONING

The collider principle in causal reasoning: Why the Monty Hall dilemma is so hard

Bruce D. Burns and Mareike Wieth

Michigan State University

Address for correspondence:
Bruce Burns
Department of Psychology
Michigan State University
East Lansing, MI 48824-1117
USA
e-mail: burnsbr@msu.edu

Abstract

Glymour (2001) speculated that people have difficulty applying the *collider principle* when two independent causal factors influence a single outcome. The principle states that such causal factors are dependent conditional on the outcome. The Monty Hall Dilemma (MHD) has this causal structure, so we tested the thesis that people find the MHD hard because they fail to understand the implications of its causal structure. Four experiments showed that participants performed better in versions of the MHD involving competition, which emphasizes causality. This manipulation resulted in more correct responses to questions about the process in the MHD and a counterfactual that changed its causal structure. Correct responses to these questions were associated with solving the MHD *regardless* of condition. In addition, training on the collider principle transferred to a standard version of the MHD. The MHD taps a deeper question, when does knowing about one thing tell us about another?

The collider principle in causal reasoning: Why the Monty Hall dilemma is so hard

The Monty Hall dilemma (MHD) is interesting because both the general public and experts in probability consistently give an incorrect answer, and fail to understand why they are wrong. Named after the host of the television show “Let’s Make a Deal”, the MHD presents three doors to a participant who is told that behind one door is a prize, but behind the other two are worthless items. The participant selects a door hoping it conceals the prize, and then the host always opens one of the unselected doors. The door the contestant selected is never opened at this stage and the host, who knows where the prize is, never opens the door concealing the prize. The participant is then given a choice: either stay with the door initially selected, or switch to the other unopened door. When faced with the two remaining doors, most people strongly feel that it's a 50% chance either way, and usually stay with their first choice. However, participants have a two-thirds chance of winning if they switch (Nickerson, 1996; Selvin, 1975a,b; for a simulation see Shaughnessy & Dick, 1991), therefore the optimal strategy is to switch.

A number of empirical studies have tested standard versions of the MHD (Brown, Read, & Summers, 2003; Friedman, 1998; Granberg, 1999; Granberg & Brown, 1995; Granberg & Dorr, 1998; Krauss & Wang, 2003; Page, 1998; Tubau & Alonso, 2003). We define a *standard version* as one in which the problem is broadly stated as above, it is the first presentation of the problem, and no clues are given. These previous papers reported thirteen studies using standard versions and switch rates ranged from 9% to 23% with a mean of 14.5% (*SD* = 4.5). This consistency is remarkable given that these studies range across large differences in the wording of the problem, different methods of presentation, and different languages/cultures. Thus it appears that failure on the MHD is a robust phenomenon unlikely to be due to confusion arising from minor aspects of the wording or presentation of the problem.

People's resistance to the correct MHD solution suggests that their failure is not just due to lack of knowledge about logic or probability. After Marilyn vos Savant in 1990 published the MHD and its correct answer in *Parade* magazine she reports receiving thousands of letters commenting on her answer. Of those letters 92% of the general public disagreed with her but so did 65% of letters with university addresses (vos Savant, 1997). Earlier Selvin (1975a) published a solution to the MHD in *The American Statistician* (distributed to all members of the American Statistical Association) but quickly follow-up because "I have received a number of letters …. Several correspondents claim my answer is incorrect." (Selvin, 1975b) Schechter (1998, pp. 108-109) relates that Paul Erdős, one of the greatest mathematicians of the twentieth century, rejected the correct solution to the MHD and only a computer simulation convinced him that switching was correct. Only after several days could a fellow mathematician make him understand why. Thus expert knowledge of probability appears insufficient for solving the MHD, which strongly suggests that its difficulty is because people commonly misrepresent the task.

Explanations for the MHD

Krauss and Wang (2003) provide an excellent review of the existing literature on the MHD. They address a number of possible explanations for people’s failure on the MHD, so we will not repeat their analysis here (see also Nickerson's, 1996, analysis of the danger of ambiguities in experiments on the MHD).

People's strong tendency to stay rather than switch, even if they think each has an equal chance, appears to be a more general bias that has long been noted in responses to multiple-choice questions (Geiger, 1997; Matthews, 1929). Gilovich, Medvec, and Chen (1995) suggest that people stay because of anticipation of regret, and Bar-Hillel and Neter (1996) presented evidence that the reluctance to exchange lottery tickets is due to this. (Tor & Bazerman [2003] presented a version of the MHD in which the participants made a decision about *someone else* making a trade. They found that 41.4% correctly said that person should trade, but their protocols indicate that only 7.3% of the total sample reasoned correctly.) We will not address this more general bias in this paper. However, this strong bias to stay has an important methodological implication because it results in base-rates for staying that are much lower than 50%, allowing us to compare switch rates to a low baseline. Participants deciding to switch appear to do so only if they see compelling reasons to believe that switching raises the chance of winning.

Switch rates improve when participants experience multiple trials (Franco-Watkins, Derks, & Dougherty, 2003; Friedman, 1998; Granberg & Brown, 1995; Granberg & Dorr, 1998; Tubau & Alonso, 2003). However, people are good at frequency detection (Hasher & Zacks, 1984) so presenting multiple trials may only teach an association between switching and winning, without improving understanding of the problem. Granberg and Brown, Franco-Watkins et al., and Tubau and Alonso all found little evidence that multiple trials improved reasoning about the MHD. Some people learn the right response after multiple presentations with feedback but this does not explain why they find the MHD so difficult to solve the first time it is presented.

Research has focused on the MHD as a failure of representation. Bar-Hillel and Falk (1982) suggested that critical to solving the MHD is understanding that the way in which the information was obtained is important. Falk (1992) saw it as an example of problem solvers' failure to differentiate options. Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni (1999) have argued that people fail to differentiate the options in the MHD because they create the wrong set of mental models (three instead of six). Franco-Watkins et al. (2003) suggest that the problem seems counterintuitive because people reinterpret the problem space once the host has revealed a door. Both Krauss and Wang (2003) and Tubau and Alonso (2003) provided direct evidence that presenting the MHD in ways that help participants form the correct representation improved their performance, in particular it seems by leading them to focus less on the two doors remaining after one was opened. This work is important and provides insight into the MHD, however this paper is addressing a different issue: Why is it so hard to spontaneously form the appropriate representation once an option has been eliminated?

That people fail at problems due to misrepresentation is a point that goes back at least as far as the Gestalt psychologists (e.g., Maier, 1931). If the MHD is just another example of how people misrepresent problems, then it may be of limited general interest. However, we set out to examine the MHD based on the assumption that a problem that produces such consistent failure and resistance to the correct solution may reveal something new about reasoning.

Causality in the Monty Hall dilemma

There has been extensive empirical work on causality (for reviews see Sperber, Premack, & Premack, 1995; Shanks, Medin, & Holyoak, 1996) showing that people often make errors when reasoning about causes. For example, misunderstandings regarding causality can lead people to give significance to spurious correlations (Chapman & Chapman, 1969), weigh the same evidence differently (Pennington & Hasties, 1988), or draw incorrect conclusions about conditional probabilities (Tversky & Kahneman, 1980). Even when people correctly identify causal relationships, they may still fail to grasp the implications of a situation's causal structure.

Some spurious causes arise from a causal structure that can be represented by a graph in which one factor has unidirectional causal links to two (or more) effects, for example (arrows indicate directional causal links):

(murder rate) ← (summertime) → (ice cream consumption)

Such common-cause structures can produce statistical dependency between two independent factors (e.g., murder rate and ice cream consumption) despite the lack of any causal relationship between these factors. That "correlation is not causation" is well known. Less well known are the implications of the opposite causal structure, when two independent variables influence a third variable. Glymour (2001, pp. 69-71) refers to such common-effect structures (to use Rehder's, 2003, terminology) as *collider phenomena* and to their implication as the *collider principle*. This principle states that for a structure containing two independent variables that are both causal with respect to an effect, these causal variables are *dependent* conditional on that effect. Glymour cites Pearl's (2000) example: Assume there are only two reasons for your car not starting, either the fuel tank is empty and/or the battery is dead. Assume also independence, thus knowing the tank state (empty or not) provides no information about the battery state (charged or dead), and vice-versa. However once the outcome of the common effect is known there is a conditional dependency between tank and battery states, because both have a causal influence on the common effect (car starting). For example, if after your car does not start you discover the battery is charged then the tank must be empty; but if instead you first discover that the tank is full then the battery must be dead. Thus the two causes (tank and battery states) are dependent conditional on the value of the effect (car starts) because this situation has a collider structure:

(state of battery) → (car starting) ← (state of tank)

The car example of the collider principle is relatively easy, partly because the answer is deterministic rather than probabilistic. A probabilistic example of the collider principle is provided by Pearl (2000, p.17, though he calls this the *explaining away effect*). If the admission criteria for a certain graduate school are high grades and/or special musical talent, then these two attributes will be found to be negatively correlated amongst students at that school. This will be true even if these attributes are uncorrelated within the general population. Once again two independent causal factors influence a single outcome, and therefore the collider principle predicts that the uncorrelated causes become correlated conditional on that outcome. Knowing about one cause (i.e., grades) provides information about the other cause (i.e., musical ability) if we know the outcome that occurred (i.e., the student has been admitted to the graduate school).

Glymour (2001, pp. 70-71) sees the MHD as evidence that people find the collider principle hard to apply. His claim is not that the MHD is isomorphic to the car (or graduate school) examples, but that the causal structures for all these scenarios are collider structures. Critical to the MHD is the following collider structure:

(Placement of prize) → (Door opened by host) ← (Initial choice)

Application of the collider principle means that conditional on the information regarding which door the host opened, the contestant's initial choice provides information about where the prize was placed. This information can be used to increase the probability of winning. One third of the time the placement of the prize and the initial choice have the same value and thus the host's choice is less constrained, but the other two-thirds of the time the two causes determine exactly which door the host will open. Thus the contestant wins two-thirds of the time by switching because the host's actions reveal exactly where the prize is two-thirds of the time. If the host could open a door with the prize behind it or open the door the contestant chose then the collider principle would not apply and the outcome of the host's actions would be uninformative.

We can calculate the conditional correlation arising from knowing which door the host opened. There are nine combinations of the door initially chosen and the location of the prize but, because the host sometimes has two choices, we need to represent eighteen cases in order for all combinations to have an equal probability of occurring. Johnson-Laird et al. (1999) pointed out the need for such identical cases and they show six of the eighteen possible. Of course, in twelve of the eighteen equally probable cases contestants win by switching, but the correlation between the prize location and initial choice is zero. However once a door is opened only six cases remain, and regardless of which door was opened the correlation between prize location and initial choice is now $r(6) = -0.33$. These conditional correlations mean that once it is known which door the host opens (and thus which six cases are still viable) the knowledge of which door was initially chosen provides information about the location of the prize.

Hypotheses and Outline

Glymour's (2001) insight that at the heart of the MHD is the collider principle provides a possible explanation for why people find it hard to form or maintain the appropriate representation. If failure to understand the implications of the collider principle underlies why people find the MHD so hard, then this suggests a set of empirically testable hypotheses.

Hypothesis 1. The collider principle appears to be easier to understand in some contexts than others, therefore if we put the MHD into such a context then participants should perform better in terms of switch choices and correct calculations of the probability of winning if they switch. As we will explain, competition should be such a context.

Hypothesis 2. The collider principle applies to the *process* by which an option is eliminated, therefore if a competition context improves the performance of participants it should also increase the likelihood of participants recognizing that the nature of the elimination process in the MHD should influence their answer.

Hypothesis 3. If we present participants with a counterfactual in which we change the causal structure, such that the collider principle no longer applies, participants given a competition context should be more likely to indicate that this may affect their answer.

Hypothesis 4. Participants who recognize the implications of the collider principle should be better at solving a standard version of the MHD. Thus correct answers to process questions (see Hypothesis 2) and counterfactual questions (see Hypothesis 3) should be associated with solving the MHD *regardless* of which MHD version a participant is given.

Hypothesis 5. Participants given training that helps them understand the implications of collider principle should be better able to solve a standard version of the MHD.

We tested these hypotheses in four experiments. All experiments supported Hypothesis 1 by showing that a competition manipulation improved performance. Experiments 2 and 3 asked participants about the process eliminating an option, and supported Hypotheses 2 and 4. Experiments 3 and 4 supported Hypotheses 3 and 4 by presenting a counterfactual that changed the MHD's causal structure. Experiment 4 supported Hypothesis 5 by giving participants training on the collider principle. Thus we provided empirical evidence for the thesis that failure to understand the implications of the collider principle can lead to errors in the MHD.

Before reporting these experiments we should clarify some general points. First, if people cannot identify a problem's causal structure then they will fail to apply the implications of that structure. However, for reasons we will elaborate on later, we think that people's failure to apply the implications of the collider principle is not just due to the failure to understand what the causes are. Second, we are not claiming that participants who solve the MHD will be able to explicitly state the collider principle; rather we expected their answers to reveal an understanding of its implications. Third, our claim is not that all the difficulties participants have with the MHD are due to the collider principle, there is clear evidence that other manipulations not related to causality can improve performance (e.g., Krauss & Wang, 2003). We did not attempt to reconcile all the existing evidence regarding the MHD, instead we asked if a more fundamental principle may contribute to why the MHD is particularly hard.

Experiment 1

In some contexts it appears to be easier to apply the collider principle. Competition may provide such a context because people often seem to see competitions in terms of causality (Lau & Russell, 1980; McAuley & Gross, 1983; White, 1993). Additionally, competitions may provide situations in which people have experienced the collider principle in action. For example, in a game between a competitor that we know has just defeated an opponent versus one who is yet to play a game, we may well favor the victor of the completed game. People may vary in how they explain such a preference, but to the extent that this general preference is correct (disregarding factors such as practice) it is due to the collider principle. The two independent variables of which competitor is best and which competitor did not compete in the first game both have an influence on which competitor loses the first game, yielding a collider structure:

(which competitor is best) → (competitor losing first game) ← (competitor left out)

By applying the collider principle (explicitly or implicitly), once we know who won the first game there is an association between which competitor was left out and which of the competitors is the best of the three. There are various ways to calculate that it is better to go with the winner of the first contest (e.g., by applying Bayes' theorem or by ordering the three competitors), but any of these calculations succeed because of the collider structure governing such a competition. If the best of the three competitors is *guaranteed* to win, then it is even possible to calculate exact winning probabilities for the two remaining competitors. Just as in the Doors version of the MHD, one third of the time the best competitor is initial chosen so who wins the first game is not constrained, but the other two-thirds of the time the two causes determine exactly which competitor wins. Thus switching wins two-thirds of the time because the first game's result reveals exactly who represents the prize two-thirds of the time. Competition may therefore provide us with a context in which people are able to better see how to apply the collider principle to an analog of the standard version of the MHD.

We designed a competition version of the MHD that instantiated the exact same rules that the host must follow when opening a door in the MHD. Three potential competitors replaced three doors, and instead of trying to pick which of the three doors conceals a prize, the contestant tried to win a prize by picking which of the three competitors is the one so good as to be guaranteed to beat either of the other two. An initially selected competitor is protected from elimination by not competing in the first match, just as the initially selected door cannot be opened by the host. Instead of the host following the rule "Do not open the door concealing the prize, but if neither do then choose randomly", the same rule can be implemented as "The one best competitor cannot lose the first match, but if neither is the best then who loses is random." The final choice becomes not "stay with the initial door selected, or switch to the door not opened by the host," but "stay with the initial competitor selected, or switch to the competitor who didn't lose the first match." Such a scenario contains a collider structure by Glymour's (2001) definition.

We created four isomorphic versions of the MHD: Three competition versions that used a competitive process to eliminate an option and a noncompetition version that was intended to be equivalent to a standard version of the MHD. By creating three different competition versions and comparing each to the noncompetition version we reduced the risk that a consistent competition effect could be due to a minor detail of a given scenario. The three competition versions also varied in terms of the fairness of the competition, which we expected would vary the success of the manipulation in increasing participants' performance. If people are better at applying the collider principle in competitive situations, then the easier it is to see a situation in terms of a true competition, and the more participants should decide to switch.

In the most competitive version we replaced the three doors with three boxers who would fight a pair of bouts. One of the three boxers was so good that he was *guaranteed* to win any bout. After a contestant selects one of the three boxers as his or her pick to be the best, the other two boxers fight. The winner of the first bout will then fight the boxer initially selected, and contestants win if they chose the winner of this second bout. However, after the first bout contestants are offered the choice: stay with their initial selection, or switch to the winner of the first bout. Table 1 shows that the *Boxers* version is formally isomorphic to the standard version of the MHD. The critical difference is the process implementing the rules eliminating one of the unchosen options. Both processes have collider structures, however they should differ in terms of how easily people understand the implications of their common causal structure.

The *Wrestlers* version was identical to the Boxers version, except that professional wrestlers replaced boxers, and it was pointed out that professional wrestling results are pre-determined. So although competition determines which of the unselected options is eliminated, it is not a fair competition. We expected this to reduce the likelihood of participants understanding the causal structure's implications and result in poorer performance than in the Boxers condition.

The Wrestlers and Boxers scenarios did not involve opening doors, which meant they differed from standard versions of the MHD because the prize was fixed to a person rather than a door. So the *Wrestlers-D* version had wrestlers defending doors, and the door concealing the grand prize (placed there before any matches started) was defended by the "best" wrestler. If a wrestler lost a match, then that wrestler's door was opened. Linking the wrestlers to the doors might further reduce the perception of this as a fair competition by emphasizing that no matter what the wrestlers do, the prize's location is fixed. So we predicted that participants would perform more poorly in the Wrestlers-D condition than in the Wrestlers condition.

A *Doors* version was similar to the Wrestlers-D version, except that now what determined which door was opened was not a competition between wrestlers, but instead was indicated by the promoters who knew where the prize was. The wrestlers just stood in front of doors and yelled at each other until a door was opened. The Doors version was designed to be equivalent to any other standard versions of the MHD, yet it retained the same fighting context as the other isomorphs. If the doors version is equivalent to other standard versions of the MHD then it should result in switch rates comparable to those for switching in other experiments. These switch rates however should be lower than in the three competition versions.

Switching appears to indicate that a participant has realized that the two options are not equivalent, which is what many researchers see as the critical insight necessary for solving the MHD. As an additional measure of performance we asked participants to give the exact probability of winning if they switched. Given that we know that people are poor at conditional probability problems in general (for reviews, see Kahneman, Slovic, & Tversky, 1982), we would expect most participants not to be able to give the precise probabilities even if they understood the problem correctly. However, if our manipulations both increased switching and led more participants to give the right probabilities, that would be good evidence that they encouraged the correct reasoning about the MHD.

Method

Participants

A total of 326 students in the Michigan State University subject pool participated in the experiment in partial fulfillment of class requirements.

Materials and Procedure

Participants were randomly assigned to one of the four versions of the MHD (summarized above[1]). Each version stated that a contestant is randomly selected at an event (either a boxing or wrestling night) and always presented with three options, one of which represents a substantial prize. Participants were told contestants had no knowledge that would help them make the initial selection. After an option is selected, one of the remaining two options is *always* shown not to be the critical one (prize or best competitor). The contestant then has to make a decision: stay with the first selection, or switch to the other remaining option. Participants were told that one night they were the randomly selected contestant, then a particular selection and result were described. They circled their choice, "stay" or "switch," and wrote down percentages for how likely they thought they were to win if they stayed, and how likely they would be to win if they switched.

Participants were also asked to write down why they made their choices. We originally intended to classify participants' explanations, but we discovered that these were very difficult to score objectively. Participants varied greatly in terms of the amount and precision of what they wrote. In addition they often repeated their answers or included elements of the scenario, therefore it was rarely possible for raters to score these explanations blind to participants' answers and conditions. Thus we decided to only use objective measures: participants' stay or switch choice, and their explicit estimates of how likely switching would lead to success. However in each experiment we asked participants to explain their answers in order to encourage them to think about their responses. Some participants changed their stay/switch choice (in both directions) after they started to write an explanation. In all experiments we also asked participants whether they were familiar with the problem they were given. Participants circled "yes" or "no" in response to this question and we eliminated participants who circled "yes."

Results & Discussion

Ten participants reported having seen the problem before and were dropped from all further analysis (interestingly, only 3 switched). Only 15% of participants in the Doors condition switched, a rate in the range found in other experiments using a standard version of the MHD. Thus the Doors version appears comparable to those used in earlier experiments, despite the text being longer and the context changing from a game show to a wrestling night.

Table 2 shows the percentage of participants in each condition choosing to switch. As predicted, significantly fewer participants in the Doors condition switched than did those in the Wrestlers-D condition, $X^2(1) = 10.94$, $p = .001$ ($\Phi = .26$), the Wrestlers condition, $X^2(1) = 5.49$, $p = .019$ ($\Phi = .18$), and the Boxers condition, $X^2(1) = 17.28$, $p < .001$ ($\Phi = .32$). (We report effect sizes using the phi [Φ] coefficient of association. Wickens [1989, chap. 9] discusses effect size measures for categorical data.)

We also expected a progression of switching from the Doors to Boxers condition. The first part of the progression, Doors verses Wrestlers-D was significant. The second part of the progression was not as the Wrestlers-D versus Wrestlers condition did not differ, $X^2(1) = 1.0$, $p =$

.33. However, the last part of the progression was significant such that participants in the Boxers condition switched more than in the Wrestlers condition, $X^2(1) = 7.3$, $p = .007$ ($\Phi = .22$). Therefore, the proportions of participants switching largely matched our expectations: the fairer the competition, the more likely the correct choice. The similarity of Wrestlers and Wrestlers-D conditions suggests that fixing the prize's location was not critical.

To assess if participants could both switch and give the right probabilities of winning we scored participants as giving the correct probability if they: 1) switched, 2) wrote down 33% (or .33) as the chance of winning if one stayed, and 3) wrote down 66% or 67% (or .66 or .67) as the chance of winning if one switched. The results of applying these criteria are shown in Table 2. The proportion fitting these criteria in the Doors condition was less than that in the Wrestler, $X^2(1) = 4.75$, $p = .030$ ($\Phi = .17$), and in the Boxers condition, $X^2(1) = 8.32$, $p = .004$ ($\Phi = .23$).

Our results supported Hypothesis 1; putting the MHD into a context in which people find it easier to apply the collider principle led to better performance. Each of our three competition versions of the Monty Hall dilemma produced a greater number of switch decisions than did the standard version. Competition also increased the number of participants indicating the correct percentage chance of winning if they switched (though not for the Wrestlers-D condition). Thus presenting participants with versions of the MHD that emphasized the causal nature of the process eliminating an option led to better performance. Additionally, manipulating the fairness of the competition appeared to vary the size of the competition effect, as we expected. It should be noted however that we did not independently measure "fairness of competition."

An alternative explanation for Hypothesis 1 could be proposed. If competition implied an ordering of the three options, then a participant might be able to reason that it is better to switch to an option with no chance of having the worst ranking (i.e., the option not eliminated), rather than stay with one which could still be the one with the worst ranking (i.e., the initial selection).

An implied ordering is an interesting alternative way to calculate the answer to the MHD that does not require any understanding of the collider principle and could be applied to standard versions of the MHD. However in an unpublished experiment we explicitly ordered the quality of the options in Doors and Boxers version of the MHD. We found no effect of the ordering manipulation for either version but replicated the competition effect.

Experiment 2

Experiment 1 showed that scenarios involving competition led to improved performance on the MHD, but it did not demonstrate why. The competition manipulation was intended to improve performance in the MHD by leading participants to better understand the implications of the problem's causal structure. The collider structure applies to the *process* by which an option is eliminated, so participants in the competition condition should better understand that this process is critical. Participants have little reason to differentiate the options if they do not see the process as having importance beyond simply reducing the number of options from three to two. Therefore in Experiment 2 we tested Hypothesis 2 by directly asking participants about the elimination process and whether its nature influenced their answers.

Our competition manipulation was one way of improving understanding of the implications of the MHD's causal structure, but it should not be the only way. All that should be necessary is that the manipulation leads people to better understand the significance of an elimination process that has a collider structure. Any such manipulation should lead to improved performance.

A manipulation that can improve performance is increasing the number of initially identical options beyond three. A knowledgeable host then follows the same rules as in the MHD to eliminate every other option except for the one initially chosen, and one other. Then the choice is offered to stay with the first option or switch to the single option the host did not open. Empirically it has been found that increasing the number of initial options to at least ten can reliably increases the amount of switching relative to a three-option MHD version (Franco-Watkins et al., 2003; Hell & Heinrichs, 2000; Page, 1998; Sternberg & Ben-Zeev, 2001). Why the number of options manipulation is effective has not been explored, but, similar to competition, it should greatly increase the emphasis on the process by which options are eliminated. Thus this manipulation should increase the number of participants switching, and the number recognizing that the process through which an option was eliminated was important. Therefore in Experiment 2 we tested this by manipulating both whether the scenario was a form of competition or not, and the total number of options (3, 32, or 128).

If participants are more likely to switch when they understand the importance of the elimination process, then this should be true in all conditions. Hypothesis 4 proposes that switching should be associated with such understanding regardless of condition.

Method

Participants

Three hundred and seventy-nine introductory psychology students at Michigan State University participated in the experiment in partial fulfillment of course requirements. Nineteen were excluded from the analysis because they indicated that they had seen the problem before.

Materials and Procedure

The 3 (number of initial options: 3-options, 32-options, 128-options) x 2 (process: competition, noncompetition) design resulted in six conditions. Each participant was randomly assigned to one of these six conditions and received the relevant scenario.

In the competition conditions participants read that they could win a prize by selecting the winner of a tennis tournament. They were told they knew nothing about tennis, so they randomly picked a player, however before the tournament started their selected player was disqualified and the tournament was conducted without her. During the tournament players were eliminated when they lost a match, until just two players remained for the final. However, it was discovered during the final that the allegation leading to their selected player's disqualification was false. So it was decided to give that player the chance to win the tournament by playing the winner of the “final.” The text emphasized that the player's actual ability had nothing to do with this decision; it was done for reasons of basic fairness. Participants then indicated their preference: stay with their first choice or switch to the winner of the final?

In the noncompetition condition participants were told that they were given a chance to win a prize on a local television show. They had to try to pick the box containing a prize, and made an initial random choice. Boxes were then brought on stage two at a time, and one was opened by the host who knew where the prize was and never opened the box containing the prize if it was one of the two on stage (this two-boxes-at-a-time routine was designed to parallel a knockout tournament). Eventually just one unopened box remained, plus the original choice. Participants were then given the choice: stay with your first choice or switch to the last unopened box?

The initial number of boxes or players presented was 3, 32, or 128. The 3-options conditions were designed to be analogous to standard versions of the MHD. To present three options in the knockout tennis tournament some extra information was required. In the 3-option/competition conditions there were initially four players, but one withdrew with injury. After your player was disqualified, the tournament winner was supposed to be the winner of the match between the remaining two players. To emphasize that there were three possible tournament winners, it was stated that you calculated at the start that there was a 1/3 chance of your random selection winning the tournament. A similar statement was given in all conditions, specifying a 1/3, 1/32, or 1/128 chance. All participants chose to stay with their first option or switch to the only other remaining option, and specified the percentage chance of winning if they stayed or switched.

On a separate page a *process question* asked participants to choose between two different statements. The statement pairs differed slightly between conditions because opening boxes is not identical to playing tennis matches. In the Competition conditions participants were asked which of the following statements would they agree with more?

A. Now that there are just two players competing in a match, it doesn't make any difference how player Y got to the match against player X.

B. The process by which player Y (i.e., winning the final) got to the match against player X affected my judgment of how likely it was that player Y would win the match against Player X.

In the Noncompetition conditions participants chose between:

A. Now that there are just two boxes remaining, it doesn't make any difference how box Y got to be the one you considered together with box X.

B. The process by which box Y (i.e., not being opened by a person who knows where the prize is) got to be the one considered together with box X affected my judgment of how likely it was that box Y would contain the prize rather than box X.

Results & Discussion

Table 3 shows the number of participants who choose to stay or switch in each condition. Overall, more participants switched in the competition than in the random conditions, $X^2(1) = 24.47$, $p < .001$ ($\Phi = .26$), and when more options were presented, $X^2(2) = 37.42$, $p < .001$. There was a significant difference between the switch rates for 3-options and 32-options, $X^2(1) = 24.58$, $p < .001$ ($\Phi = .32$), but not between 32-options and 128-options, $X^2(1) = 0.54$, $p = .46$. There was no evidence of an interaction between competition and number of options as a loglinear analysis of logits found no significant interaction term, $G^2(2) = 0.70$, $p = .70$. For each number of options there was an effect of competition: 3-options, $X^2(1) = 6.50$, $p = .011$ ($\Phi = .23$); 32-options, $X^2(1) = 6.19$, $p = .013$ ($\Phi = .23$); 128-options, $X^2(1) = 12.94$, $p < .001$ ($\Phi = .32$).

Table 3 also shows that the manipulations affected answers to the process question. Participants were more likely to indicate that the process was important when they were in the competitive conditions, $X^2(1) = 24.63$, $p < .001$ ($\Phi = .26$), and when there were more options, $X^2(2) = 21.85$, $p < .001$ (again the difference between 3-options and 32-options was significant, $X^2(1) = 10.76$, $p = .001$ [$\Phi = .21$], but not that between 32- and 128-options, $X^2(1) = 1.67$, $p < .20$). A loglinear analysis of logits found no significant interaction term between the options and competition factors, $G^2(2) = 0.20$, $p = .90$. The competition effect was again significant for each level of options: 3-options, $X^2(1) = 6.45$, $p = .011$ ($\Phi = .23$); 32-options, $X^2(1) = 10.52$, $p = .001$ ($\Phi = .30$); 128-options, $X^2(1) = 7.15$, $p < .007$ ($\Phi = .24$). Thus both competition and the number of options had the predicted effects, they increased (albeit imperfectly) the likelihood of participants seeing the process as important.

The process question was highly associated with whether a participant decided to stay or switch. Across conditions, 91% (157/172) of participants who switched said that the process was important, however only 19% (35/180) of participants who stayed said it was important, $X^2(1) = 183.1$, $p < .001$ ($\Phi = .72$). This relationship between the process question and switching held for both the competition (94% who switched agreed with the process question verse 23% who stayed, $X^2[1] = 93.2$, $p < .001$ [$\Phi = .74$]) and noncompetition conditions (86% who switched agreed with the process question verse 17% who stayed, $X^2[1] = 80.1$, $p < .001$ [$\Phi = .68$]). This finding shows that understanding the importance of the process by which the final two choices were derived was strongly associated with solving the MHD, regardless of the competition manipulation. Furthermore, it shows that most participants who switched did so for good reasons, not just because they were confused or picked randomly.

The 3-options analog

Standard versions of the MHD present only three options, so it could be argued that increasing the number of options changes the problem. However the results from the 3-options condition alone appear to replicate the competition effect from Experiment 1, despite a very different cover story. More participants in the competition condition switched than did so in the random condition. The number switching in the noncompetition condition was within the range found in previous experiments using other standard versions of the MHD.

Analysis of correct percentages (switching and indicating the probability of winning was 66-67%) again found that more participants in the competition condition (15%, 8/55) were correct than in the random condition (2%, 1/63), $X^2(1) = 7.00$, $p = .008$ ($\Phi = .24$). (Equivalent analysis for 32- and 128-option is harder to interpret because the correct answers are more difficult to calculate and are close to 100% [97% and 99%], which participants sometimes give even when there are only 3 options.) In the 3-option conditions it was found again that participants who switched (25/30) were more likely to indicate that the process was important than those who stayed (18/85), $X^2(1) = 36.6$, $p < .001$ ($\Phi = .56$).

Summary

Experiment 2 replicated the competition effect found in Experiment 1 with very different materials, thus Hypothesis 1 was further supported. In addition, we found that competition increased the proportion of participants recognizing that the process was important. Thus Hypothesis 2 was supported. Competition and number of options had independent effects on performance, but they both affected the same factor that our analysis suggests underlies their success: the likelihood of seeing the elimination process as important. Hypothesis 4 was supported because regardless of condition understanding the importance of the elimination process (i.e., the process governed by the collider principle) was associated with switching.

Experiment 3

Experiment 3 directly probed whether participants understood the implications of the causal structure of the MHD. Counterfactuals have been used to investigate participants' understanding of causality (see Spellman & Mandel, 1999), so we presented participants with a counterfactual: what if one of the critical causal links was removed? Would this change affect their choice to stay or switch? Such a counterfactual changes the probability of winning if they switch to 50%. For participants who made their choice on the basis of the implications of the causal structure, such a counterfactual should lead them to at least consider changing their answer. However for participants who made their choices on some basis other than the causal structure, the counterfactual should appear irrelevant. Therefore participants' reaction to this counterfactual should be an indicator of whether they had some understanding of the implications of the problem's causal structure (though it is no doubt an imperfect indicator) allowing us to test Hypothesis 3. If the competition manipulations led people to be more likely to switch because it led them to better understand the implications of the MHD's causal structure, then participants in a competition condition should be more likely than those in a noncompetition condition to indicate that the counterfactual might change their choice. Furthermore, if understanding the MHD's causal structure is important for solving the problem, then *irrespective* of condition participants who switched should be more likely than those who did not switch to indicate that the counterfactual may change their answers. Thus we could test Hypothesis 4.

A process question similar to Experiment 2 was posed but with a different format. A possible criticism of the previous process question was that it forced participants to make a choice between just two options. Perhaps those indicating that the process was important just disliked that answer less than the alternative. Therefore in Experiment 3 we provided more options. One option was equivalent to Option A in Experiment 2: the process was irrelevant to their choice. The remaining options distinguished between the process causing them to stay or to switch, and we gave participants the choice of not agreeing with any of the options. We expected to replicate

Experiment 2 thus participants in the competition condition should be less likely to agree that the process was irrelevant to their choice and more likely to agree that the nature of the process caused them to switch. Responding that the process was irrelevant should be more likely for participants who decided to stay, regardless of condition, as Hypothesis 4 suggests.

Failure to apply the collider principle could have two possible reasons: 1) participants may fail to understand that the process is causal because they think of an entity opening doors as just a random process; 2) they may understand that the process is not random, but fail to understand the implications of its causal structure. We have assumed that participants fail to understand the implications of the MHD’s causal structure, but it is possible that participants who chose to stay are not seeing the process as causal at all. To check this we asked participants to rate how random they thought the process was. If competition led participants to see the elimination process as more causal, then participants in the competition condition should rate the process as more nonrandom than those in the noncompetition condition.

Gigerenzer and Hoffrage (1995) argued that people reason better with frequencies than with percentages. This has been challenged (e.g., Kahneman & Tversky, 1996), but we decided to avoid a possible criticism by replacing the probability questions with frequency questions. Doing so also allowed us to ask about probabilities in a way that did not allow a valid "50%" answer. Forcing participants to choose an answer favoring one or the other option may be a more sensitive way to pick up any preference participants might have towards staying or switching.

Method

Participants

Six hundred and fourteen introductory psychology students at Michigan State University participated in the experiment in partial fulfillment of course requirements. A large sample ensured a number of participants in the noncompetition condition who chose to switch. Nineteen were excluded from the analysis because they indicated that they had seen a problem before.

Procedure and Materials

Participants were randomly assigned to one of the two conditions: *Competition* or *Noncompetition*, which utilized scenarios identical to the Boxers and Doors condition in Experiment 1, respectively. Participants chose to "stay" or "switch" then indicated the chance of winning if they switched. As explained above, we asked about the chance of winning if they switched using a frequency format. Participants answered the following two questions by writing a digit between 0 and 9 (of course, in the competition conditions the word "boxer" was substituted for "door"):

Imagine that over the next nine weeks, **nine** people were given the same scenario and choices (of course the prize will not always be behind the same door). Imagine that **all** choose to do the same thing. If they:

- **all** decided to ***SWITCH*** **to the other door**, how many of the 9 do you think would win?
- **all** decided to ***STAY*** **with their first choice**, how many of the 9 do you think would win?

Answers were in response to nine cases because this included the correct answer (6/9) but lacked a 50/50 point, forcing a response favoring switching or staying.

Other questions were presented on a separate page. First participants were presented with the *nonrandomness question*, "The process by which Door A [*Boxer A*] was eliminated is best described as…" They answered using a six-point scale from 1 (completely random) to 6 (completely nonrandom). Burns and Corpus (2004) used this scale successfully to distinguish between scenarios differing in terms of nonrandomness. Participants were then presented with the *counterfactual question*. In the Noncompetitive condition they were asked:

> What if you found out that the promoter who gave the signal to open Door A had got his notes mixed up at the last moment? So he actually had no idea where the prize was and had

> just opened a door completely randomly hoping it would not have the prize behind it? Fortunately, Door A did not have the prize behind it, but would the knowledge that which door was opened was random had any effect on how you reasoned about the problem, either about your choice whether to STAY or SWITCH or your estimate of how many times out of nine this choice would lead to a win?

For the Competitive condition they were asked:

> What if you found out that a bout of flu had been going around the boxers, such that it was actually completely random who would win any bout? So how good any of the boxers were had nothing to do with the fact that Boxer A lost the first bout. Fortunately, Boxer A was not the boxer who was guaranteed to beat any of the others boxers, but would the knowledge that who won the first bout was random had any effect on how you reasoned about the problem, either your choice about whether to STAY or SWITCH or your estimate of how many times out of nine this choice would lead to a win?

In both conditions participants answered by circling "YES," "NO" or "MAYBE." The "maybe" option was included to capture participants who recognized that the situation may have changed, but may not be confident in calculating the consequences. Such participants should contrast with those who think the counterfactual has changed nothing.

The *process question* asked participants to choose which of the following statements they most agreed with (the relevant term for the Noncompetition condition is in brackets):

> A. The actions leading up to the elimination of Boxer A [*Door A*] had **nothing** to do with whether I was more likely to win if I decided to STAY or SWITCH.
>
> B. The actions leading up to the elimination of Boxer A [*Door A*] **caused** me to be more likely to win if I decided to SWITCH.
>
> C. The actions leading up to the elimination of Boxer A [*Door A*] **caused** me to be more

likely to win if I decided to STAY.

D. None of the above statements are correct. Please explain:

Very few participants gave Option D so we did not attempt to classify their responses.

Results & Discussion

As in Experiments 1 and 2, a larger proportion of participants switched in the Competition (Boxers) condition (122/299, 41%) than in the Noncompetition (Doors) condition (38/296, 13%), $X^2(1) = 59.2$, $p < .001$ ($\Phi = .32$). More participants in the Competition conditions (40/299, 14%) switched and indicated the correct chance of winning (6/9) than in the Noncompetition condition (5/296, 2%), $X^2(1) = 29.2$, $p < .001$ ($\Phi = .22$). Therefore we supported Hypothesis 1 again, but using a frequency question did not appear to improve performance.

Comparisons to Krauss & Wang's results for correct reasoning

Krauss and Wang (2003, p. 14) classified 32% and 38% of their participants as indicating a two-thirds chance of winning and using the correct mathematical reasoning in their most effective conditions. They pointed out that this was much more than the number reported as giving the correct percentages in an earlier report of Experiment 1 (Burns & Wieth, 2000). However, it is hard to directly compare our data because we asked participants to indicate the expected frequency of winning out of *nine* cases whereas Krauss and Wang's participants indicate a frequency out of only *three* cases. Only 5% of participants in Experiment 3 gave the extreme answers of 0/9 or 9/9, probably because most participants realized that the result is not certain. Assuming Krauss and Wang's participants also avoided the extremes, in effect most of them were faced with a dichotomous choice between 1/3 and 2/3. Presumably participants who favored switching for any reason would be more likely to indicate 2/3 than 1/3. In our experiment participants who favored switching, but did not know what the exact likelihood of winning was, had four options other than 100% (5/9, 6/9, 7/9, 8/9). Thus those who answered 6/9 probably had a good reason to choose exactly this response, as did those who answered exactly 66% or 67% in Experiment 1.

Krauss and Wang (2003, p. 11) state that in order to classify a participant as reasoning correctly both the 2/3 answer and a written responses indicating a comprehensively derivation of the answer was required. However no detailed criterion are given for this classification (what constitutes a valid mathematical derivation of the answer has been disputed, see Morgan, Chaganty, Dahiya, & Doviak, 1991). Thus it is hard to evaluate their criterion.

In an attempt to make our findings more equivalent to Kraus and Wang's (2003) findings we assumed that participants who answered 5/9, 6/9, or 7/9 probably would have answered 2/3 if the only real options were 1/3 or 2/3. In the Competition condition 34% gave one of these three responses and switched, whereas only 6% did so in the Noncompetition condition, $X^2(1) = 68.2$, $p < .001$ ($\Phi = .48$). Regardless of condition, 75% of participants who choose to switch gave one of these three responses, whereas only 17% of participants who decided to stay did so, $X^2(1) = 173.8$, $p < .001$ ($\Phi = .54$).

We do not expect that a different question or scoring would have changed Krauss and Wang's (2003) patterns of result. Our point is only that it is possible that they overestimated the true proportion of their participants reasoning mathematically well enough to give the correct probability of winning if they switched. Ultimately though we are addressing different questions using different scenarios that are difficult to directly compare to Krauss and Wang's.

Counterfactual Responses

Table 4 presents participants' responses to the counterfactual question. As predicted participants in the Competition condition were more likely to indicate that if the elimination process was random they would (or might) change their answers, $X^2(2) = 9.53$, $p = .009$. Furthermore, participants who switched were more likely to indicate that the counterfactual situation would (or might) affect their answer in both the Competition, $X^2(2) = 46.2$, $p < .001$, and Noncompetition conditions, $X^2(2) = 11.6$, $p = .003$. Dichotomizing between participants thinking the counterfactual at least might affect their decision ("Yes" or "Maybe" answers) versus those thinking it definitely would not ("No" answers), we found that 75% of participants in the Competition condition who switched indicated that the counterfactual at least might affect their answer, as opposed to 36% of participants in that condition who stayed, $X^2(1) = 43.4$, $p < .001$ ($\Phi = .38$). In the Noncompetition condition, 63% of participants who switched indicated that the counterfactual at least might affect their answer, as opposed to 37% of participants in that condition who stayed, $X^2(1) = 9.75$, $p < .001$ ($\Phi = .18$).

Process Responses

Table 5 presents participants' process question responses. Participants in the Competition condition (31%) were more likely than those in the Noncompetition condition (9%) to say that the process caused them to switch, $X^2(1) = 44.9$, $p < .001$ ($\Phi = .28$). However it is unclear how to interpret the responses of participants who chose to stay and indicated that the process caused them to make that decision. More directly comparable between Experiments 2 and 3 are participants who gave *Response A* indicating explicitly that the process had nothing to do with their choice. Participants were more likely to give the *nothing* response in the Noncompetition (63%) than the Competition (47%) condition, $X^2(1) = 15.0$, $p < .001$ ($\Phi = .16$). Again, across conditions answering that the process had *nothing* to do with their choice had a clear association with staying (67%) rather than switching (26%), $X^2(1) = 77.9$, $p < .001$ ($\Phi = .36$). Participants who stayed were also more likely than those who switched to give the *nothing* response, whether they were in the Competition (66% vs. 21%), $X^2(1) = 60.3$, $p < .001$ ($\Phi = .45$), or the Noncompetition condition (67% vs. 42%), $X^2(1) = 8.32$, $p = .004$ ($\Phi = .17$).

Thus we again found evidence for Hypothesis 2, as the competition manipulation affected whether participants thought the process was critical. Thinking this way was strongly associated with switching, further supporting Hypothesis 4. The relationship between switching and viewing the process as important was weaker than in Experiment 2 perhaps due to the increased range of answers offered. However, finding the same effects with a very different question format adds to our confidence in Hypotheses 2 and 4.

Nonrandomness Responses

To not see the process as causal, in either scenario, would be an error that indicates that participants misunderstood the causal structure of the problem. Only if participants understood the problem's causal structure, can the counterfactual question be informative regarding whether participants understood the implications of that structure (i.e., the collider principle). The nonrandomness scale provided information regarding whether the Noncompetition scenario led participants to think that the elimination process was random to start with (which would render the counterfactual scenario moot). Participants in the Noncompetition condition saw the process as more nonrandom ($M = 4.2$, $SD = 1.5$) than those in the Competition condition ($M = 3.7$, $SD = 1.5$), $t(592) = 3.9$, $p < .001$ ($\omega^2 = .02$). (To provide some context to this measure, Burns & Corpus [2004] report that participants gave a mean nonrandomness rating of 3.6 to a competition between sales people, and a mean 2.2 rating to a coin toss.) Why participants saw the Noncompetition version as slightly less random than the Competition version is unclear. It could be that participants tend to regard any competition as involving an element of uncertainty whereas in the Noncompetition version which door was opened was governed by a clear set of rules. However critical to our thesis was that participants did not see the process in the Doors version as more random than in the Boxers version, suggesting that the competition manipulation did not simply lead participants to better understand that the links are causal.

To test whether participants' answers to the nonrandomness question were associated with their answers to the other questions, we performed an ANOVA on response to the nonrandom question with factors of scenario, stay/switch choice, and counterfactual answer. As well as the scenario effect, participants who switched rated the process more nonrandom (*M* = 4.22, *SD* = 1.5) than those who stayed (*M* = 3.8, *SD* = 1.5), *F*(1,578) = 11.9, *p* = .001 (ω^2 = .02). However there was no effect of counterfactual answer, *F*(2,578) = 0.12, *p* = .88, and no interactions approached significance. The lack of association between the nonrandomness question and the counterfactual question suggests that they tap into different aspects of participants' understanding of the task: whether the process is causal verse what implications that causality has.

Summary

Apart from replicating the competition effects on performance (supporting Hypothesis 1), Experiment 3 again showed that the competition manipulation increased participants' likelihood of seeing the elimination process as important (supporting Hypothesis 2). Answers to the process question were again associated with switching regardless of conditions (supporting Hypothesis 4). New in the experiment was a counterfactual question that focused on a causal link that Glymour's (2001) analysis suggests should be critical, the causal relationship from which option represents the prize to which option is eliminated. Without this link, the collider principle does not apply to the scenario. Participants in the competition condition were more likely to recognize that making the elimination of an option random – by removing the causal link – could affect their choice. Regardless of whether participants were in the Competition or Noncompetition conditions, those who switched were more likely to recognize that the elimination of this causal link might change their answer, whereas most participants who selected the "stay" choice indicated that making this link random would have no impact on their answers. Thus Hypothesis 4 was supported. The results from the randomness scale suggest that this was not because participants in the Noncompetition condition already viewed the process as random. These findings support our argument that an important reason why people decide to stay in the MHD is that they fail to understand the implication of its causal structure.

Experiment 4

So far, we have shown that we can improve participants' performance through modifying the MHD by putting it into a context in which people find it easier to apply the collider principle. This implies that appropriate training on the collider principle should improve participants' performance on standard versions of the MHD, as Hypothesis 5 suggests.

Experiment 4 presented a Transfer group with training on the collider principle before they were given a standard version of the MHD. We wanted to demonstrate not just better solution rates but also conceptual transfer measured using the counterfactual question from Experiment 3. The transfer literature has demonstrated the difficulty in obtaining conceptual transfer (see Detterman & Sternberg, 1993), however conceptual transfer can be improved by presenting multiple sources (Gick & Holyoak, 1983; Spiro, Feltovich, Coulson, & Anderson, 1989). Thus our training had more than one task, because we did not expect that a simple presentation of the collider principle would be effective given how difficult people find both the collider principle and the MHD. Participants in the Transfer group were given three tasks that should all help with understanding how the collider principle applies to the MHD.

Method

Participants

Two hundred and eighty introductory psychology students at Michigan State University participated in the experiment in partial fulfillment of course requirements.

Procedure and Materials

Participants were randomly assigned to one of the two conditions: *Transfer* or *Nontransfer*. Participants in the Transfer condition were first presented with the Boxer version used in

Experiment 3 then they answered the stay/switch question and estimated the frequency of winning in both cases. They were then given the following text and question:

> From the point of view of maximizing your chances of picking the best boxer, you should SWITCH to Boxer C, the one who just won the first bout. **It is not certain that Boxer C will turn out to be the best, but there is a 2/3 chance of him being the best boxer** (if the best boxer, whoever it is, is absolutely guaranteed to win against either of the other two)
>
> People often find it hard to understand why it makes any difference whether they stay or switch (though if you try this as an experiment at home, you will quickly discover that switching wins more often). The essential reason is that you have learnt **something about Boxer C** because he had the chance to be eliminated, but wasn't, whereas you have learnt **nothing about Boxer B** who was waiting for the second bout. Therefore, on balance, it's likely that Boxer C is the better one.
>
> Do you really believe (be honest!) that it is better to switch to Boxer C than stay with Boxer B? YES NO

Transfer participants were then presented with another example of the collider principle that made explicit the causal structure and the principle. The example was a modification of the graduate-school example discussed earlier but presented in a way that maps better to the MHD:

> *Who is a good musician?*
>
> Why does knowing about Boxer C tell you something you didn't know before? In one sense, the answer is obvious, Boxer C won. But the underlying reason is that the pattern of causes present in the scenario: there are two causes (which boxer you initially chose, and who is best) of a single outcome (who wins the first bout). This is called the *collider principle*. We will try to explain why with a different example.
>
> Imagine a college called Beethoven College, which is very selective so the only way to

> get in is to have really high SATs or be really good at music. In the general population there is no correlations between grades and musical ability, but at Beethoven College SATs and musical ability are *negatively* correlated, that is, students with high SATs tend to have average/poor musical ability and students with high musical ability tend to have low/average SATs (some students are high on both, but they are rare, and there is no one who is low on both). This isn't because high SATs cause you to be bad at music but because this is an example of the *collider principle*: there are two causes (high SATs and/or high musical ability) of one effect (getting into Beethoven College). So knowing about your SATs tells me something about your likely musical ability, if I know already that you are a student at Beethoven College.
>
> Imagine that you were looking for a really good musician and didn't have time to hold auditions. There are three candidates but for some reason you know the SAT scores of **two** of them. If the group were all MSU students this would not help you at all, but imagine instead that the potential musicians were students at Beethoven College. If all you knew was which of these two students had the higher SAT score, who should you choose: (please circle one answer)
>
> A. The student who had the **higher** of the two SAT scores.
>
> B. The student who had the **lower** of the two SAT scores.
>
> C. The student whose SAT scores is **unknown**.
>
> D. They are **all equally** likely to be good musicians.
>
> Does this situation seem similar to the Three Boxer problem? YES NO

(Note that SAT is a standardized test commonly used in US college admissions.) In order to allow an easier mapping between this scenario and a standard version of the MHD the above scenario introduced three students that map onto the three doors. To do this required a scenario that utilizes the collider principle *twice*. Its first application is what we initially gave as how the graduate-school example involves the collider principle, that knowing an SAT score is informative about musical ability, conditional on knowing that a student has been accepted at Beethoven College. This collider structure creates a relationship between SAT and musical ability, which are causal with regard to the outcome of being accepted, and thus gives significance to SAT scores. The second collider structure has an outcome at its apex: which of the *known* SAT scores is the higher one. There are two causal factors influencing this outcome: which of the three students' SAT score is unknown, and which of the two *known* scores is higher. (Note that no information is given about the absolute level of the scores compared.) If you know that out of Students B and C it is Student B who has the higher SAT (and thus is equivalent to the door opened), then that tells you that Student A (the one left out of the comparison, just as the door initially chosen is left out) has the better chance of having the higher SAT score of the two remaining students. Thus due to the collider structure that creates a relationship between SAT scores and musical abilities, Student C is likely to be a better musician than Student A.

Both the Transfer and Nontransfer groups were then presented with the Doors version of the MHD used in Experiment 3. After answering the stay/switch and frequency questions, both groups answered the counterfactual question given in Experiment 3. Although participants in the Nontransfer condition were not presented with anything equivalent to the Boxer version before doing the Doors version, both groups had completed an unrelated set of written tasks before being presented with the materials for this experiment.

Results and Discussion

Seventy-four participants reported having seen either the Doors or Boxers versions of the MHD, which was a much larger proportion of participants than in the previous experiments. It was discovered that this was because the MHD had been presented and explained to one of the classes participating in the subject pool (44 out of the 74 switched). Some participants may have failed to report they had seen the problem, but due to random assignment this should raise switch rates equally across conditions. However, only 12% of analyzed participants switched when given the Doors version first (i.e., the Nontransfer group), which is indistinguishable from the Doors condition switch rates in Experiments 1 and 3.

Table 6 presents the percentage of participants giving the switch responses to the Doors version in both conditions, and the Boxer version in the Transfer condition. This shows that the competition effect was again replicated: 12% of participants switched when given the Doors version first but 44% switched when given the Boxers version, $X^2(1) = 26.1$, $p < .001$ ($\Phi = .36$). More importantly for this experiment, more participants switched in the Doors version (36%) in the Transfer than the Nontransfer condition, $X^2(1) = 16.7$, $p < .001$ ($\Phi = .28$).

Frequency responses

For the Doors version, more participants (13/103) in the transfer condition indicated the correct frequency of winning if they switched (6/9 if switched and 3/9 if stayed) than in the Nontransfer condition (6/103), but this difference was not quite statistically significant, $X^2(1) = 2.84$, $p = .092$ ($\Phi = .12$). However, if we used the criterion of whether a participant indicated that switching would lead to a win 5, 6 or 7 times out of 9, then giving this response and switching was more likely for participants in the transfer condition (33/103) than the nontransfer condition (10/103), $X^2(1) = 15.5$, $p < .001$ ($\Phi = .27$).

It could be argued that some Transfer participants may have switched simply because they were told that switching was correct for the Boxer version, so by making switching part of the criteria for the correct frequency for the Doors version we may have overestimated the performance of the transfer group relative to the nontransfer group. Therefore we calculated the number of participants giving the 5/9, 6/9 or 7/9 answers regardless of whether they stayed or switched. (Some participants who stayed may give such answers either because they think it a 50% chance so 5/9 is as likely as 4/9, or simply because they are confused.) These answers were still more likely for participants in the Transfer condition (43/103) than the Nontransfer condition (19/103), $X^2(1) = 13.3$, $p < .001$ ($\Phi = .25$).

Counterfactual responses

As in Experiment 3, answering "yes" or "maybe" to the counterfactual question was associated with switching in a Doors version of the MHD. Of the total deciding to stay 47% gave one of these answers, whereas of those who decided to switch 76% gave one of these answers, $X^2(1) = 12.1$, $p < .001$ ($\Phi = .24$). Thus Hypothesis 4 was again supported.

Table 7 presents the frequencies of each counterfactual answer for the Transfer and Nontransfer conditions. Participants in the Transfer condition were more likely to give the "yes" or "maybe" response to the counterfactual questions (61%) than were those in the Nontransfer condition (47%), $X^2(1) = 4.40$, $p = .036$ ($\Phi = .15$). Unlike the stay/switch question which participants had been given already in the Boxers version, nothing like the counterfactual question had been presented in the training procedure. This suggests that better performance on the counterfactual question is due to conceptual transfer.

Individual training tasks

Participants in the Transfer condition were presented with three training tasks, each related to helping participants understand the collider principle. Our question was not which task would be most effective at helping participants solve the Doors version of the MHD. Nonetheless each task had a question associated with it that tested understanding, so we investigated whether success with particular tasks differentially predicted switching in the Doors version of the MHD.

Table 8 reports how well Transfer participants answered the training task questions and the correlations of these answers with switching in the Doors version of the MHD, and with each other. All three questions correlated positively with switching, but none was statistically significantly. However, if we added up how many of the three questions participants answered correctly, we found a significant correlation with switching in the Doors version, $r(101) = .23$, $p = .019$. Thus we had no evidence that one training task was more effective than the others at leading to switching, even though completing the whole training procedure increased switching.

Summary

Experiment 4 indicated that giving participants training on the collider principle resulted in better performance in a standard version of the MHD. Transfer participants were not only more likely to switch, but also showed conceptual transfer by correctly answering a counterfactual question that required them to understand the implications of the MHD's causal structure. Thus this experiment supported Hypothesis 5 and provided converging evidence that one reason for people's poor performance is that they fail to understand the implications of its causal structure.

The finding that participants could answer the more conceptual counterfactual question after being given the training procedure is good evidence that training helped participants understand the MHD's causal structure. However the large rise in the number of Transfer participants deciding to switch in the Doors version need not be dismissed as a trivial example of surface transfer. Although we told participants that switching was best for the Boxer version, simply telling participants that they should switch and explaining why, only raised the switch rate from 44% to 62%. This is consistent with the observation that the MHD is a stubborn problem that is somewhat impervious to statements of the correct answer. However this experiment did not give a clear indication as to which of the three training tasks was most effective at increasing the rate at which participants answered the Doors version of the MHD correctly.

General Discussion

The growing number of empirical studies of the MHD seems predicated on the assumption that something interesting must explain people's failures. In many reasoning tasks people make consistent errors, but normally once it is explained to them people will accept that they have made an error. Piattelli-Palmarini (1994, p. 161) regards the MHD as the most expressive example of a "cognitive illusion" into which even knowledgeable people get trapped. The power of the MHD to stir strong disagreement whenever it is presented to a skeptical public, or even a knowledgeable one, makes it unusual. (The closest comparison may be the strong reaction to Gilovich, Vallone & Tversky's [1985] debunking of the hot hand in basketball, but see Burns, 2004.) Previous studies have shown that participants misrepresent the MHD and if given clues to the correct representation they will do better. However these studies have provided relatively little evidence regarding what general principles might explain why it is particularly hard to correctly represent the MHD. Glymour (2001) put the MHD into the broader framework of causal reasoning (Burns & Wieth [2000] also made this suggestion) which allows the MHD to be used to test an interesting claim: that people find it difficult to reason with the collider principle. Our experiments provide the first empirical evidence that at the heart of the MHD is a causal reasoning problem by supporting the five hypotheses we derived from the claim that the MHD is difficult because it requires people to understand the implications of the collider principle. Thus the results of the four experiments provide support for Glymour's (2001) speculation that the MHD is hard because it requires understanding the implications of its causal structure.

Interpretation of Results

When presenting earlier versions of this research we have encountered various queries to our interpretation of the data. It is worthwhile specifying why we think our interpretation holds up and how the set of experiments addresses these queries.

Why is competition effective? With any complex written material it is impossible to determine exactly how participants interpret it. Thus it could be asked whether our competition effect is due to something unrelated to it invoking the collider principle (explicitly or implicitly). However, our consistent finding of a competition effect with each one of the four different competition scenarios suggests that the improved performance is not just due to some quirk of an individual scenario. The competition scenario in Experiment 2 in particular differs greatly from the others.

It has been suggested that it is "obvious" that people should favor the winner of the first competition, but this begs the questions of why is it obvious that the winner of one game is likely to be better than another competitor who was not involved in that game? When we question people about why they think it is obvious they usually do not get further than stating that the winner is better, which is repeating the fact that they favor that competitor to win. Excluding factors due to having competed (e.g., fatigue, practice), the winner of the first game should be favored against a competitor who has not yet competed only because this is an application of the collider principle (or a generalization of it). We have not determined whether people are better at applying the collider principle to competition because they think of it in terms of causality, or because it is a specific situation in which they have experience with the collider principle.

It was important to test predictions derived from the causal analysis for manipulations other than competition. Experiment 2 showed that manipulating the number of options improved performance on the MHD also increased participants' recognition of the importance of the elimination process governed by the collider phenomenon. Experiment 4 showed that training relevant to the collider principle increased performance on a standard version of the MHD.

Effect Sizes for Switching. Labeling the MHD as a cognitive illusion seems to imply that a single factor should provoke an "Aha!" response and make the illusion disappear. However even if solving the MHD were purely a matter of giving people the right representation, then this would not necessarily imply that it will easily disappear. Weisberg and Alba (1981) showed that having the right representation of a problem is a prerequisite to solving an insight problem (the nine-dot problem), not a guarantee of success. Krauss and Wang (2003) and Tubau and Alonso (2003) showed that manipulating participants' representation did not make the MHD easy, which is consistent with our claim that there is a further reason (i.e., failure to understand the implications of its causal structure) underlying the difficulty in correctly representing the MHD.

The number of participants switching in our experiments was often around 50%, but this should not be interpreted as evidence that participants were simply responding randomly because the bias against switching is a strong one. Switching was strongly associated, regardless of condition, with the process questions in Experiments 2 and 3, and the counterfactual question in Experiments 3 and 4. This suggests that participants who switch are not just confused or acting randomly, instead they usually switch for meaningful reasons that are consistent with our explanation of what is necessary for understanding the MHD. Conversely some people may be able to give the right answer to the MHD without an understanding of what underlies the right answer. They however may be like someone who can apply a statistical formula without understanding, and therefore fails to recognize when it produces a ridiculous result.

The Dangers of Failures to apply the Collider Principle

This paper is the first empirical support for Glymour's (2001) analysis of the difficulty of the MHD as due to it involving the collider principle. However there is no reason to think the MHD is unique as a failure to apply the collider principle, even if it is a strong case. It is possible that other unexplained reasoning errors may be due to failures to apply the collider principle.

An example of the collider principle in epidemiology is Berkson's (1946) paradox: two diseases may be found to be correlated amongst hospital patients but be unrelated in the general population. Many studies of diseases focus on hospital admissions, thus this paradox is a serious problem as it can lead to health researchers wasting resources on investigating what look like promising leads. There is no statistical fix to this problem, as it is fundamental if the sample is selected from hospital populations. Thus over 750 citations can be found of Berkson's paper since 1970. Yet Schwartzbaum, Ahlbom, and Feychting (2003) still felt the need to publish a review of Berkson's paper to accompany Sadetzki, Bensal, Novikov, Modan (2003) paper on the consequences of using hospital controls to establish cancer etiology. Unlike these papers, Pearl (2000, p. 17) points out that Berkson's paradox arises due to a causal structure in which two independent causes (the two diseases) can both produce a single outcome (hospital admission). He uses different terms but states the collider principle as clearly as Glymour (2001) does.

Berkson's paradox does not only apply to physical diseases. Clarkin and Kendall (1992) defined *comorbidity* as the "occurrence at one point in time of two or more *DSM-III-R* disorders" (p. 904) and suggest that it is "an emerging and essential concern for both theoretical understanding of psychological disturbances and treatment planning." (p. 904) A decade later Jensen (2003) echoed this sentiment. However Lilienfeld (2003) pointed out that care must be taken because of the purely mathematical consequences of the fact that an individual with two disorders can seek treatment for either disorder. Thus Berkson's paradox could produce what looks like evidence of comorbidity. Therefore Lilienfeld suggested it is important when seeking evidence of comorbidity to take into account biases in how people ended up in the sample.

Collider phenomena may be at the heart of other reasoning problems. For example, the collider principle could be seen as at the core of a fundamental dispute over how to conduct scientific research: what can be concluded about a hypothesis from the rejection of an alternative hypothesis? Oaksford and Chater (1994, 1996) point out that Popper (1959) argued that experiments could only falsify, never confirm, general laws. Thus scientists should focus on falsification, because no matter how often an alternative hypothesis is falsified this does not support the original hypothesis. Oaksford and Chater point out that contemporary philosophers of science reject falsification as inaccurate history of science and unworkable. Instead they apply

Bayesian approaches to confirmation to derive a new understanding of the Wason selection task (though their approach has been criticized, e.g., Laming, 1996). Oaksford and Chater's approach suggests that causal reasoning may be important in a range of reasoning tasks. Furthermore the collider principle may apply to scientific reasoning, as there is an analogy to the MHD.

Imagine three hypotheses (A, B, C) regarding a phenomenon. Assume that the three are mutually exclusive but one can be shown to be correct (or at least correct in the sense that only the other two will be rejected). Further assume that only two hypotheses can be tested in a single experiment but that any such experiment can definitively reject one of the two hypotheses tested. This situation is isomorphic to the MHD. If Experiment 1 tests Hypotheses A and B, and it rejects B, then Experiment 2 is more likely to reject Hypothesis C than Hypothesis A. This is because the situation has a collider structure:

(Which hypothesis is correct) → (Hypothesis rejected in Exp. 1) ← (Hypotheses tested)

Of course usually there are more than three hypotheses, and those hypotheses may not be mutually exclusive. However this is analogous to increasing the number of doors and giving the doors nonuniform values: it still pays to switch.

Generalization of collider phenomena makes clear why participants failed to see the elimination process as more causal in the Competition than in the Noncompetition version of the MHD in Experiment 3. Participants do not perform better in the Competition version because they see it as causal; they do so because the context helps them see the implications of the causal structure. When epidemiologist cite Berkson's paradox they are not suggesting that doctors do not understand what diseases put people in hospital, they are pointing out the perils of not understanding the implications of there being two unrelated diseases that can do this.

Implications

Pearl (2000, p. xiii) points out that making proper inferences about causality is the central aim of the physical, behavioral, social, and biological sciences. However, if even simple collider structures can lead to difficulties for experts in statistics, then it may not be surprising that Pearl also observes that we have difficulty understanding causality. Collider phenomenon may be difficult to reason about because they relate to a fundamental question: under what circumstances can information about one entity tell us something about another entity? The answer depends on the causal structure of the situation, which is not always easy to determine. Thus the explanation underlying why the MHD is so hard may be an important principle whose influence on reasoning has not been explored by psychology. An implication of this analysis is that selection biases (e.g., Berkson's paradox) can be much more dangerous than is generally thought, because they exploit a blind-spot in people's understanding of the implications of causality.

This paper is the first experimental investigation of failure on a problem due, at least partly, to a failure to understand the implications of the collider principle. In general, more examination of reasoning in terms of causality may be fruitful, and focusing on causality may open up interesting new avenues for future research into reasoning.

References

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

Bar-Hillel, M., & Neter, E. (1996). Why are people reluctant to exchange lottery tickets? *Journal of Personality & Social Psychology, 70*, 17-27.

Berkson, J. (1946). Limitations of the application of fourfold table analysis to hospital data. *Biometrics Bulletin, 2*, 47-53.

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

Burns, B. D. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand". *Cognitive Psychology*, *48*, 295-331.

Burns, B. D., & Corpus, B. (2004). Randomness and inductions from streaks: "Gambler's fallacy" versus "hot hand." *Psychonomic Bulletin & Review, 11,* 179-184.

Burns, B. D., & Wieth, M. (2000, November). *The Monty Hall Dilemma: A causal explanation for a cognitive illusion.* Poster session presented at the Forty-First Annual Meeting of the Psychonomic Society, New Orleans, LA.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271-280.

Clarkin, J. F., & Kendall, P. C. (1992). Comorbidity and treatment planning: Summary and future directions. *Journal of Consulting and Clinical Psychology, 60,* 904-908.

Detterman, D. K., & Sternberg, R. J. (Eds.) (1993). *Transfer on trial: Intelligence, cognition, and instruction.* Norwood, NJ: Ablex Publishing.

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*, 67-90.

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

Geiger, M. A. (1997). Educators' warnings about changing examination answers: Effects on student perceptions and performance. *College Student Journal, 31*, 429-432.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

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

Gilovich, T., Medvec, V. H., & Chen, S. (1995). Commission, omission, and dissonance reduction: Coping with regret in the "Monty Hall" problem*.* *Personality and Social Psychology Bulletin, 21*, 185-190.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*, 295-314.

Glymour, C. N. (2001). *The mind's arrow: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT Press.

Granberg, D. (1999). Cross-cultural comparison of responses to the Monty Hall Dilemma. *Social Behavior and Personality, 27*, 431-438.

Granberg, D., & Brown, T. A. (1995). The Monty Hall dilemma. *Personality and 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.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*, 1372-1388.

Hell, W., & Heinrichs, C. (2000). Das Monty-Hall-Dilemma (Ziegenproblem) mit 30 Türen [The Monty Hall Dilemma (goat problem) with 30 doors]. In D. Vorberg, A. Fuchs, T. Futterer, A. Heinecke, U. Heinrich, U. Mattler & S. Töllner (Eds.), *Experimentelle Psychologie* [Experimental Psychology] (p. 82). Lengerich, Germany: Pabst.

Jensen, P. S. (2003). Comorbidity and child psychopathology: Recommendations for the next decade. *Journal of Abnormal Child Psychology, 31,* 293-300.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naive probability: A mental model theory of extensional reasoning. *Psychological Review, 106*, 62-88.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582-591.

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

Laming, D. (1996). On the analysis of irrational data selection: A critique of Oaksford and Chater (1994). *Psychological Review, 103,* 364-373.

Lau, R. R., & Russell, D. (1980). Attributions in the sports pages. *Journal of Personality & Social Psychology, 39*, 29-38.

Lilienfeld, S. O. (2003). Comorbidity between and within childhood externalizing and internalizing disorders: Reflections and directions. *Journal of Abnormal Child Psychology, 31,* 285-291.

Maier, N. R. F. (1931). Reasoning in humans. II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology, 12,* 181-194.

Mathews, C. O. (1929). Erroneous first impressions on objective tests. *Journal of Educational Psychology, 20*, 280-286.

McAuley, E., & Gross, J. B. (1983). Perceptions of causality in sport: An application of the Causal Dimension Scale. *Journal of Sport Psychology, 5*, 72-76.

Morgan, J. P., Chaganty, N. R., Dahiya, R. C., & Doviak, M. J. (1991). Let's make a deal: The player's dilemma. *The American Statistician, 45*, 284-287.

Nickerson, R. S. (1996). Ambiguities and unstated assumptions in probabilistic reasoning. *Psychological Bulletin, 120*, 410-433.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review, 101*, 608-631.

Oaksford, M., & Chater, N. (1996). Rational explanation of the selection task. *Psychological Review, 103,* 381-391.

Page, S. E. (1998). Let's make a deal. *Economics Letters, 61*, 175-180.

Pearl, J. (2000). *Causality: Models, reasoning, and inference.* Cambridge, UK: Cambridge University Press.

Pennington, N., & Hasties, R. (1988). Explanation-based decision making: Effects of memory structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 521-533.

Piattelli-Palmarini, M. (1994). *Inevitable illusions: How mistakes of reason rule our minds* (M.

Piattelli-Palmarini & K. Botsford, Trans.). New York: John Wiley & Sons.

Popper, K. R. (1959). *The logic of scientific discovery.* London: Hutchinson.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science, 27,* 709-748.

Sadetzki, S., Bensal, D., Novikov, I., & Modan, B. (2003). The limitations of using hospital controls in cancer etiology - one more example for Berkson's bias. *European Journal of Epidemiology, 18*, 1127-1131.

Schwartzbaum, J., Ahlbom, A., & Feychting, M. (2003). Berkson's bias reviewed. *European Journal of Epidemiology, 18*, 1109-1112.

Schechter, B. (1998). *My Brain is open: The mathematical journeys of Paul Erdős.* New York: Simon & Schuster.

Selvin, S. (1975a). A problem in probability [Letter to the editor]. *The American Statistician, 29*, 67.

Selvin, S. (1975b). On the Monty Hall problem [Letter to the editor]. *The American Statistician, 29*, 134.

Shanks, D. R., Medin, D. L., & Holyoak, K. J. (Eds.) (1996). *The psychology of learning and motivation: Vol. 34. Causal learning*. San Diego, CA: Academic Press.

Shaughnessy, J. M., & Dick, T. (1991). Monty's dilemma: Should you stick or switch? *Mathematics Teacher, 84*, 252-256.

Spellman, B. A., & Mandel, D. R. (1999). When possibility informs reality: Counterfactual thinking as a cue to causality. *Current Directions in Psychological Science, 8*, 120-123.

Sperber, D., Premack, D., & Premack, A. J. (Eds.) (1995). *Causal cognition: A multidisciplinary debate.* Oxford, UK: Oxford University Press.

Spiro, R. J., Feltovich, P. J., Coulson, R. L, & Anderson, D. K. (1989). Multiple analogies for complex concepts: Antidotes for analogy-induced misconception in advanced knowledge acquisition. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* (pp. 498-531). New York: Cambridge University Press.

Sternberg, R. J., & Ben-Zeev, T. (2001). *Complex cognition: the psychology of human thought*. New York: Oxford University Press.

Tor, A., & Bazerman, M. H. (2003). Focusing failures in competitive environments: Explaining decision errors in the Monty Hall game, the Acquiring a Company problem, and multiparty ultimatums. *Journal of Behavioral Decision Making, 16*, 353-374.

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

Tversky, A., & Kahneman, D. (1980). Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology.* Hillsdale, NJ: Erlbaum.

vos Savant, M. (1997). *The power of logical thinking.* New York: St Martin's Press.

Weisberg, R. W., & Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General, 110*, 169-192.

White, S. A. (1993). The effect of gender and age on causal attribution in softball players. *International Journal of Sport Psychology, 24*, 49-58.

Wickens, T. D. (1989). *Multiway contingency tables analysis for the social sciences.* Hillsdale, NJ: Erlbaum.

Author Note

Bruce D. Burns, Department of Psychology; Mareike Wieth, Department of Psychology.

We would like to thank Fernanda Ferreira, Stefan Krauss, Raymond Nickerson, David Shanks, Regina Vollmeyer and three anonymous reviewers for helpful comments on earlier drafts of this paper. A preliminary report of the results of Experiment 1 was presented at the Forty-First Annual Meeting of the Psychonomic Society, New Orleans, LA, November 2000.

Correspondence concerning this article should be addressed to Bruce Burns, Department of Psychology, Michigan State University, East Lansing, MI 48824-1117. Electronic mail may be sent via Internet to `burnsbr@msu.edu`.

Footnotes

[1] All materials are available from the authors by request.

Table 1.

Illustration of the isomorphism between the standard (Doors) version of the MHD, and the competition (Boxers) version.

Doors version	**Boxers version**
1. A decision maker tries to win a prize by picking which of three doors conceals a prize.	1. A decision maker tries to win a prize by picking which of three boxers is unbeatable.
2. Choose one door randomly.	2. Choose one boxer randomly.
3. One of the other two doors is opened	3. One of the other two boxers loses a bout between them.
4. The door concealing the prize cannot be opened in line 3.	4. The one unbeatable boxer cannot lose the bout in line 3.
5. If neither of the two doors in line 3 conceals the prize, then which door is opened is random.	5. If neither of the two boxers in line 3 is unbeatable then who loses is random.
6. Two doors remain, the one chosen in line 2 and the door remaining after line 3. Soon the prize will be revealed. One door must conceal the prize.	6. Two boxers remain, the one chosen in line 2 and the winner of the bout in line 3. They will now fight to reveal who is the best. One of them must be unbeatable.
7. Decision maker stays with the door first chosen, or switches to the door not opened in line 3.	7. Decision maker stays with the boxer first chosen, or switches to winner of the bout in line 3.
8. Correct choice in line 7 wins the prize.	8. Correct choice in line 7 wins the prize.

Table 2

Percentages of participants choosing to switch, and percentages correctly stating the probability of wining if they switch for each condition of Experiment 1. Exact proportions are given in parentheses based on the number of participants giving the answer out of the total number of participants in a condition.

	Door	Wrestlers-D	Wrestlers	Boxers
Switch rates	15% (13/88)	37% (28/75)	30% (23/77)	51% (39/76)
State correct probability of winning	2% (2/88)	4% (3/75)	10% (8/77)	14% (11/76)

Table 3

Percentage of participants choosing to switch, and percentage saying "yes" to the process question (indicating that the process was important) in each condition of Experiment 2 (exact proportions are given in parentheses). Note that some participants did not answer the process question.

	3-options	32-options	128-options	Total
Switching responses				
Competition	36%	68%	77%	62%
	(20/55)	(41/60)	(48/62)	(109/177)
Noncompetition	15%	46%	46%	36%
	(10/63)	(27/59)	(28/61)	(65/183)
"yes" to process question				
Competition	50%	73%	78%	68%
	(26/52)	(44/60)	(47/60)	(117/172)
Noncompetition	27%	44%	55%	42%
	(17/63)	(26/59)	(32/58)	(75/180)

Table 4

Frequencies of responses in Experiment 3 to the counterfactual question by stay/switch decision in each condition.

		stay	switch	Total
Competition condition				
Counterfactual	YES	25	48	73
Response	MAYBE	38	42	80
	NO	112	30	142
Total		175	120	
Noncompetition condition				
Counterfactual	YES	35	12	47
Response	MAYBE	59	12	71
	NO	163	14	177
Total		257	38	

Table 5

Frequencies of responses in Experiment 3 to the process question by stay/switch decision in each condition.

		stay	switch	Total
Competition Condition				
Causal	None	117	25	142
Response	Switch	10	83	93
	Stay	36	6	42
	other	13	6	19
Total		176	120	
Noncompetition Condition				
Causal	None	171	16	187
Response	Switch	9	18	27
	Stay	64	2	66
	other	13	2	15
Total		257	38	

Table 6

Percentage of participants choosing to switch in each group for each version of the MHD in Experiment 4 (exact proportions given in parentheses). Only the Transfer group attempted the Boxers version.

Nontransfer group	*Transfer group*	
Doors	Boxers	Doors
12% (12/103)	44% (44/101)	36% (37/103)

Table 7

Frequencies of responses in Experiment 4 to the counterfactual question in each condition.

		Transfer	Nontransfer
Counterfactual	YES	26	14
Response	MAYBE	37	34
	NO	40	52

Table 8

Correlations in Experiment 4 for correct responses to training task questions and whether Transfer condition participants switched together with statistical significance levels in parentheses. Percent answering each question correctly (and what the correct response was) is also reported.

	Boxer switch	Really believe you should switch in Boxers problem	Beethoven College problem correct	Musician problem like doors problem
Doors switch	.08 (*p*=.43)	.17 (*p*=.08)	.14 (*p*=.16)	.15 (*p*=.14)
Boxer switch		.57 (*p*=.43)	.11 (*p*=.29)	.18 (*p*=.07)
Really believe you should switch in Boxers problem			.14 (*p*=.17)	.11 (*p*=.26)
Beethoven College problem correct				.21 (*p*=.03)
Percent correct answer	44% "switch"	62% "yes"	73% "lower SAT"	72% "yes"