**Old Dominion University**
**ODU Digital Commons**

Psychology Faculty Publications

Psychology

2008

# Hindsight Bias in Insight and Mathematical Problem Solving: Evidence of Different Reconstruction Mechanisms for Metacognitive Versus Situational Judgments

Ivan K. Ash
*Old Dominion University*, iash@odu.edu

Jennifer Wiley

Follow this and additional works at: https://digitalcommons.odu.edu/psychology_fac_pubs

Part of the Cognitive Psychology Commons, and the Experimental Analysis of Behavior Commons

# Hindsight bias in insight and mathematical problem solving: Evidence of different reconstruction mechanisms for metacognitive versus situational judgments

**IVAN K. ASH**
*Old Dominion University, Norfolk, Virginia*

AND

**JENNIFER WILEY**
*University of Illinois, Chicago, Illinois*

This article presents two experiments that used insight and mathematical problems to investigate whether different factors would affect hindsight bias on metacognitive and situational judgments. In both studies, participants initially rated their likelihood of solving each problem within a certain amount of time (metacognitive judgments) and rated the importance of each component of the problem for finding the solution (situational judgments). Next, participants attempted to solve each problem. In Experiment 1, all participants were given solution feedback information, but in Experiment 2, participants were not given any solution feedback. After 1 week, participants were asked to recall their original judgments. Hindsight bias was assessed by comparing the initial with the final ratings. Insight problems and math problems showed different patterns of hindsight bias effects on the metacognitive and situational judgments. The results suggest that two competing models of hindsight effects are actually complementary explanations for judgment reconstruction on different types of judgment tasks.

There is an old saying that "hindsight is 20/20." The hindsight effect describes the observation that once people are aware of the outcome to a situation, they have a tendency to falsely believe that they would have predicted that outcome. This retrospective judgment bias is robust across a wide variety of domains and task environments (see Christensen-Szalanski & Willham, 1991; Guilbault, Bryant, Brockway, & Posavac, 2004; Hawkins & Hastie, 1990, for reviews). Although researchers have systematically studied this effect for over 30 years, there is still not a unified theory of the effect that can explain results across different judgment types and task domains.

Researchers have proposed two classes of cognitive reconstruction models to explain how retrospective judgment formation leads to the hindsight effect (Hawkins & Hastie, 1990). Anchor and adjust theories propose that people produce their retrospective judgments by starting with the given outcome and adjusting their judgment away from that anchor, based on metacognitive cues. One proposed adjustment mechanism is that people might attempt to recall how "surprising" they found the given outcome to be and adjust their judgment on the basis of this metacognitive information (Hoch & Loewenstein, 1989; Ofir & Mazursky, 1990, 1997). According to this theory, if a person recalls that the outcome was expected, he or she would get an "I would

have known that!" feeling. Based on this metacognitive information, a person will make a small adjustment from the given outcome, thereby overestimating his or her predictive accuracy. It is this underadjustment that leads to hindsight bias. On the other hand, if a person recalls that the outcome was surprising, he or she would get an "I would have never known that!" feeling. This metacognitive information would lead the person to make a larger retrospective adjustment, thereby reducing the likelihood of hindsight bias.

The second class of explanations for the hindsight effect can be considered updating and rejudging theories (Carli, 1999; Hasher, Attig, & Alba, 1981; Hawkins & Hastie, 1990; Pezzo, 2003; Roese & Olson, 1996; Wasserman, Lempert, & Hastie, 1991; Wiley & Trabasso, 2003). According to this perspective, exposure to outcome information affects the mental representation of the event or situation. Outcome knowledge can affect the memory representation by rendering outcome-supporting information more available in long-term memory or by reducing the weight given to the evidence supporting other possible outcomes (Hawkins & Hastie, 1990). According to this theory, hindsight bias occurs because people use this updated representation to reconstruct their predictive opinions. Because the representation of the event has changed to be more supportive of the actual event outcome, a per-

---

**I. K. Ash, iash@odu.edu**

---

son will overestimate his or her predictive accuracy. Thus, it is the representation change caused by acquiring outcome knowledge that leads to hindsight bias.

In the present studies, we investigate whether these different explanations of the hindsight bias should be viewed as competing accounts of the same effect, or whether these theories may represent explanations of different judgment phenomena that have been grouped under the umbrella term *hindsight bias*. There are two main paradigms that have been used to investigate hindsight effects: trivia tasks and case study tasks (Christensen-Szalanski & Willham, 1991; Guilbault et al., 2004; Hawkins & Hastie, 1990). These paradigms also differ in the kinds of judgment tasks they use for the bases of the hindsight measure. Below, we argue that anchor and adjust theories of the hindsight bias may provide a more appropriate explanation for hindsight effects that are traditionally found in trivia task paradigms, whereas updating and rejudging theories may provide a more appropriate explanation for the effects traditionally found in case study paradigms. We support this distinction in two experiments that use a problem solving paradigm to investigate the role of different types of judgment tasks on hindsight bias effects.

### Trivia Task Versus Case Study Paradigms

Trivia task paradigms present a general knowledge question as the target event or situation to be judged, such as, "What percentage of Germans live in cities with more than 100,000 inhabitants?" (Schwarz & Stahlberg, 2003, p. 399). People are generally asked to judge the likelihood of getting the answer correct. The feedback in these experiments consists of either being told the correct answer to the question or being given feedback about the accuracy of one's own response. In this paradigm, researchers have found hindsight bias effects on several different types of measures, including predictions of their own (or their peers') likelihood of producing the correct response to a question (e.g., Hoch & Loewenstein, 1989); numeric answers to problems or trivia questions (e.g., Schwarz & Stahlberg, 2003); and confidence ratings for answers to forced-choice trivia questions (e.g., Koriat, Lichtenstein, & Fischhoff, 1980). On these types of judgment tasks, the predictive and retrospective judgments are metacognitive in nature because people must use their beliefs about their own knowledge or abilities in the domain to formulate their judgments. Therefore, it seems reasonable that people may use their metacognitive reactions to outcome feedback (i.e., the "I would have known that!" feeling) as cues in reconstructing their previous metacognitive judgments. We propose that the judgment reconstruction processes described by anchor and adjust theories provide a plausible explanation for hindsight bias effects on these types of metacognitive judgment tasks.

Case study paradigms present a narrative story or description of a situation, such as a medical case history (Arkes, Faust, Guilmette, & Hart, 1988) or a story describing an upcoming event (e.g., Fischhoff, 1975). Participants are generally asked to predict the outcome of the event. The feedback in these experiments consists of a conclusion to the story or description of the "true" outcome. In this

paradigm, researchers have found hindsight bias effects on two different types of measures: outcome-likelihood predictions and evidence-importance judgments (see, e.g., Carli, 1999; Fischhoff, 1975; Wasserman et al., 1991). Outcome-likelihood predictions are produced when people are asked to judge the probability of different possible outcomes of the event or situation. Hindsight bias on this type of measure is observed as higher retrospective likelihood ratings for the given outcome. Evidence-importance judgments are produced when people are asked to rate the importance or relevance of each sentence or piece of evidence presented in the narrative or case history, in terms of determining the outcome. Hindsight bias on this type of measure is observed as higher retrospective importance ratings for outcome-supporting information. In this paradigm, the predictive and retrospective judgments that relate to the task are comprehension or representation based. This is because people must use their understanding of the evidence and how it relates to an outcome in order to produce their judgments. We propose that in these tasks, changes to the mental representation may lead to changes in retrospective judgments. Therefore, the judgment reconstruction processes described by updating and rejudging theories seem to provide a plausible explanation for hindsight bias for these types of situational judgment tasks.

### EXPERIMENT 1

In Experiment 1, we collected metacognitive judgments, such as those used in trivia tasks, and situational judgments, such as those used in case study tasks, using the same stimulus materials. The goal was to see whether patterns of hindsight bias depended on the nature of the judgment tasks. Multicomponent mathematical and insight problems where chosen as the stimulus materials because complex problem solving has both performance and representational aspects that can be assessed. The performance aspect involved in problem solving tasks makes these stimuli similar to those used in trivia task paradigms and allows for the assessment of predictive and retrospective metacognitive judgments. However, unlike most simple trivia or general knowledge tasks, complex problem solving also has a representational aspect that makes it comparable to case study paradigms. Most problem solving theories and computational models of problem solving include the formation of a mental representation of the problem as an initial phase (e.g., Anderson & Lebiere, 1998; Newell & Simon, 1972). Kintsch (1998) proposed that the formation of a mental problem representation relies on the same cognitive processes as the comprehension of events, situations, or texts. Therefore, complex problem solving tasks also lend themselves to situational judgments, such as those used in case study paradigms.

Insight and mathematical problems were chosen in order to differentiate between effects due to metacognitive reactions and those due to representational updating (see the Appendix for a list of problems). Insight problems have long been used to study how people overcome representational difficulty and find creative solutions (e.g., Duncker, 1945; Maier, 1931). The types of problems used in insight

research are designed to be unfamiliar and misleading, and typically lead people to form initial mental problem representations that are inappropriate for finding the correct solution path (Kaplan & Simon, 1990; Knoblich, Ohlsson, Haider, & Rhenius, 1999; Metcalfe, 1986a, 1986b; Metcalfe & Wiebe, 1987; Ohlsson, 1992; Wiley, 1998). Many researchers have proposed that people overcome the representational difficulties posed by these problems through cognitive processes that change or restructure the initial faulty problem representation (Ash & Wiley, 2006; Durso, Rea, & Dayton, 1994; Jones, 2003; Kaplan & Simon, 1990; Knoblich et al., 1999; Ohlsson, 1992). Multistep mathematical problems were chosen as a contrast class of problems for which most college students should have ample prior experience. On these problems, participants should have well-practiced strategies and analogous problems available in long-term memory that should help them avoid forming inappropriate representations (e.g., Ross & Kilbane, 1997). Therefore, the solution of mathematical problems should be highly unlikely to involve representational change or restructuring processes.[1]

These problems were used as the stimuli in a within-participants hindsight bias paradigm, such as those commonly used in trivia-based hindsight investigations (e.g., Pohl & Hell, 1996; Schwarz & Stahlberg, 2003). Participants were initially asked to make metacognitive judgments (likelihood of solving) and situational judgments (importance of each problem component) for each problem. Then they were given 3 min to attempt to solve each problem. Outcome feedback was given in the form of step-by-step explanations of each problem solution. A week later, participants returned and attempted to remember their initial ratings. Hindsight bias was assessed as the changes between initial ratings (predictions) and memory rating (retrospections).

Solutions to both types of problems used in this study involved a series of simple steps to complete. Therefore, we propose that exposure to the solutions to either type of problem should have been equally likely to elicit an "I would have known that!" reaction. Anchor and adjust theories would therefore propose that both problem types should be susceptible to hindsight bias effects. However, insight problems are more likely to involve representational change or restructuring than are mathematical problems. Therefore, updating and rejudging theories would predict that insight problems should be most susceptible to the hindsight bias. Specifically, this theory would predict that hindsight bias effects should be most prevalent on successfully solved insight problems, due to the spontaneous restructuring that occurs as part of the insightful problem solving process (Dominowski & Buyer, 2000).

## Method

**Participants**. Fifty-one introductory psychology students from the University of Illinois at Chicago Subject Pool completed both sessions of this experiment for partial fulfillment of a class research participation requirement.

**Materials**. The problem stimuli comprised four insight problems and four multistep math problems that were chosen or designed for this study (see the Appendix for problems and note 1 for the manipulation check). The elements of each problem were categorized as being relevant or irrelevant for solution. *Relevant components* were those parts of

the problem that needed to be used or manipulated to correctly solve the problem. For example in the "Camp Alaska" math problem (Appendix, Problem 4), the three people going on the camping trip were an adult male, an adult female, and a male teen. Therefore, the components in the chart that corresponded to these values were the relevant components. All other components in the chart were unnecessary in order to solve the problem. Therefore, these parts of the problem were *irrelevant components*. An insight problem example is the "Six Glasses" problem (Appendix, Problem 6). In order to solve the problem, one had to pour the liquid from the second glass into the fifth glass. The fifth glass was therefore the relevant component. The first, third, fourth, and sixth glasses were the irrelevant components. The second glass was excluded because it was implied to be important in the solution in the directions, and therefore any possible solution (even if incorrect) had to involve it. This component was therefore considered to be nondiagnostic in assessing the appropriateness of an individual's problem representation. The Appendix lists all insight and math problems used in this study. The relevant components for each problem are underlined.

The initial rating packet comprised eight different paper-and-pencil rating packets that presented the problems in a counterbalanced order, with the constraints that every other problem was either an insight or math problem and that each problem was the first and last problem in one of the orders. The cover page of the packet instructed participants to carefully read each problem without attempting to solve the problem. Instead, they were told that they would be asked a series of opinion questions about each problem. The directions stressed that we did not want them to begin solving the problems and that we were interested in their initial impressions of the problems. The directions also explained the rating procedure continuum for the questions described below.

The first page for each problem presented the entire problem below the sentence, "Please read this problem. Do NOT attempt to solve." Underneath the problem were general opinion questions about the problem, including the question, "How likely is it that you will complete this problem correctly in 3 minutes?" Below this question was a 7.3-cm continuum anchored with *not very likely* on the left and *very likely* on the right. This rating served as the predictive *metacognitive judgment*. On the following pages, each component of the problem was listed with the directions to "Please rate each component on how important it will be in solving the problem." Below each problem component was a 7.3-cm continuum with *not important* on the left and *very important* on the right. These ratings were combined (as detailed below) to serve as the predictive *situational judgment*.

The problem solving packet presented the same eight problems, each on its own page, in the same order as in the rating booklet. The directions were presented on the cover page of the booklet. The directions informed participants that they would be given 3 min to complete each problem. The directions asked participants to show all work, circle their final answer, and explain their solution in writing, if necessary. The directions explicitly stated that all problems had solutions and that answers such as *impossible* or *not solvable* would be counted as incorrect. Finally, the directions instructed participants to wait for the experimenter's signal before beginning each problem.

Following their solution attempts, we showed all participants the correct solution to each problem. For each problem, we created a script providing a step-by-step explanation of the correct solution procedure. Also, for each problem, we created a corresponding overhead projector slide that procedurally and/or graphically displayed the correct solution process. For example, the "Camp Alaska" slide had each of the three components used in solving the problem underlined and then showed each necessary mathematical calculation. The "Six Glasses" problem slide showed a graphical representation of the initial problem situation, a picture of the contents of the second glass being poured into the fifth glass, and a final solution representation, with the empty glass returned to its original position.

The rating memory packet was identical to the initial rating packet, except the directions asked participants to "Attempt to remember your original ratings from last week's session. In other words, do NOT rate these problems again. Try to remember the point in the scale that rep-

resents your original opinion of these problems." Furthermore, on each page of the booklet, "Please reproduce your original ratings from last week's session" appeared over each set of questions. Participants were asked to remember their rating for "How likely is it that you will complete this problem correctly in 3 minutes?" The rating from this question served as the retrospective metacognitive judgment. Likewise, the ratings from the component importance questions were combined to serve as the retrospective situational judgment.

**Procedure**. In Session 1, participants completed the study in groups of 3 to 12 and were randomly assigned to one of the eight problem orders. Participants sat so that those with the same-ordered packet were not placed next to each other. The session began with participants completing the informed consent form. Next, participants completed the initial rating packet. Upon completing the rating packet, participants attempted to solve each problem in the problem solving packet. Participants were given a maximum of 3 min to solve each problem. The experimenter kept time with a stopwatch and alerted the group when there were 30 sec left for solving. Before moving on to the next problem, the participants were reminded to circle their final answer, if they had reached one. After the problem solving phase was complete, the experimenter showed step-by-step solutions for each problem in a random order. The experimenter read a script that walked through each solution, and presented an overhead slide outlining the procedure. Each solution explanation took approximately 1 min. Finally, the experimenter asked participants not to discuss the experiment with anyone, including other members of the group.

In Session 2, participants returned to the same room exactly 1 week later and were seated in the same positions. The experimenter gave each participant a rating memory packet that matched the order of their Session 1 packets, as well as another problem solving packet. The experimenter read the directions for the rating memory packet while the participants followed along. The experimenter asked the participants to pay close attention to the directions because they differed from the previous week's directions. After completing the memory task, participants attempted to solve each problem again and were debriefed, reminded not to discuss the experiment, and dismissed.

**Data coding**. We coded the likelihood ratings by measuring the distance of the participant's mark on the 7.3-cm continuum with a ruler, so that lower scores indicated lower likelihood of solving and higher scores indicated higher likelihood of solving. We coded the component importance ratings in the same way. On these ratings, lower scores indicate lower importance ratings and higher scores indicate higher importance ratings. Since all problems have different ratios of relevant and irrelevant components, we averaged the ratings across relevant and irrelevant components within each problem. Next, the scores on the irrelevant components were reverse-coded, so that higher values on both types of components represented more appropriate problem representations. Finally, we combined average relevant and average reverse-coded irrelevant measures into a single component importance score by taking the unweighted average of the two indexes. The higher the component importance score, the closer the individual was to the correct problem representation for solution.

For the problem solving phase, we had instructed participants to circle their final answer on each problem. Therefore, we coded any problem with a correct, circled final answer as a successfully solved problem. We coded all other problems as unsuccessfully solved problems. Within the unsuccessfully solved problems, we tallied two different types of failures: incorrect and unsolved. We coded problems as incorrectly solved if participants had circled a final answer that did not match the correct solution. We coded all problems without a circled final answer as unsolved problems.

## Results

**Quasi-experimental design and analysis strategy**. Our research question was not whether certain classes of people (i.e., *more intelligent*, *less intelligent*) would be more likely to show hindsight bias, but rather discerning when hindsight would be observed, keeping all other things

constant. Therefore, we needed to make within-participants comparisons of insight and math problems for both successful and unsuccessful solution attempts. Therefore, we selected participants who both correctly and incorrectly solved at least one of the insight and math problems. Otherwise, different groups of people would have contributed data in each condition, and any effects could be due to characteristics of the groups (i.e., good vs. poor problem solvers) rather than to the variables of interest. This participant selection procedure was based on Metcalfe's (Metcalfe, 1986a, 1986b; Metcalfe & Wiebe, 1987) work investigating metacognition and problem solving. Of the 51 participants who completed the study, 25 participants had data in each cell and were included in the analyses below.

For the within-participants sample, a dependent samples $t$ test comparing the mean proportion of successfully solved math and insight problems revealed no difference in solution success rates [math, $M = .49$, $SE = .034$; insight, $M = .45$, $SE = .038$; $t(24) = 0.70$, $p = .49$, $d = 0.14$]. Table 1 displays the number of successful, incorrect, and unsolved observations for each problem. As the table shows, there was considerable variability in success rates within the math and insight problems. Because of this variability among problems, we aggregated the rating data both by participants and by problems in order to examine hindsight effect as a function of problem type and solution success.

The by-participants analysis resulted in a 2 (judgment: predictive vs. retrospective) × 2 (problem type: insight vs. math) × 2 (solution success: successful vs. unsuccessful) repeated measures design with 25 observations per cell.[2] This analysis controlled for individual differences among problem solvers, but allowed individual problems to contribute more or less to each solution success category, depending on their solution rates. A planned comparison approach was adopted on the by-participants analysis in order to maximize power.[3] To investigate the effects of problem type and solution success on hindsight bias, we conducted retrospective–predictive contrasts at alpha level = .05 for each of the problem type × success conditions.

The by-problems analysis resulted in a 2 (judgment: predictive vs. retrospective) × 2 (problem type: insight vs.

**Table 1**
**Experiment 1 Solution Rates (Frequency and Percentage)**

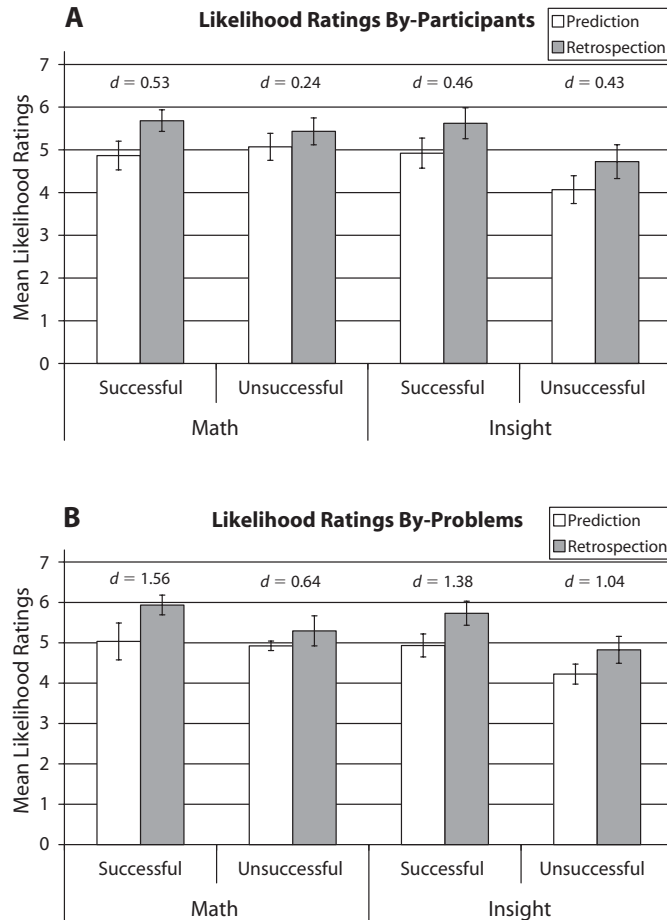|  | Successful | | Unsuccessful | | | |
|  |  | | Incorrect | | Unsolved | |
|  | Freq. | % | Freq. | % | Freq. | % |
|---|---|---|---|---|---|---|
| Math Problems |  |  |  |  |  |  |
| Solve for Y | 9 | 36 | 9 | 36 | 7 | 28 |
| Cash | 10 | 40 | 15 | 60 | 0 | 0 |
| Distance | 14 | 56 | 6 | 24 | 5 | 20 |
| Food | 16 | 64 | 6 | 24 | 3 | 12 |
| Total | 49 | 49 | 36 | 36 | 15 | 15 |
| Insight Problems |  |  |  |  |  |  |
| Triangle | 18 | 72 | 1 | 4 | 6 | 24 |
| Glasses | 15 | 60 | 8 | 32 | 2 | 8 |
| Match III | 5 | 20 | 13 | 52 | 7 | 28 |
| Match XV | 7 | 28 | 7 | 28 | 11 | 44 |
| Total | 45 | 45 | 29 | 29 | 26 | 26 |

Note—$N = 25$.

**Figure 1. Experiment 1 mean predictive and retrospective metacognitive judgments (solution likelihood ratings) as a function of problem type and solution. Error bars = standard error; *d* = Cohen's *d*.**

math) $\times$ 2 (solution success: successful vs. unsuccessful) mixed design, with problem type as a between-problems variable. Once again, we conducted a retrospective–predictive contrast for each of the problem type $\times$ success conditions. Since this analysis has only four observations per cell, this design provided very low power for detecting hindsight effects. However, the main point of this analysis was to examine whether the effects observed in the by-participants analysis were due to specific problems or were consistent across problems. Therefore, these effects will be examined descriptively, in order to assess whether they follow the same qualitative pattern as that for the by-participants analyses.

**Hindsight bias for metacognitive judgments.** Panel A of Figure 1 displays the mean predictive and retrospective likelihood ratings as a function of problem type and solution success for the by-participants analysis. Panel B of Figure 1 displays the same data when ratings are aggregated in a by-problems analysis. The size of the hindsight effect for each category was calculated using Cohen's *d* and is displayed over each pair of observations.

The results of the four predictive versus retrospective judgment contrasts are displayed in Table 2 for both the by-participants and by-problems analyses.

From the by-participants analysis, significant hindsight bias on likelihood ratings was detected on successfully solved math problems, successfully solved insight problems, and unsuccessfully solved insight problems. The effect for unsuccessfully solved math problems was in the same direction as that for the other conditions. However, this effect was approximately half the size of the effect in the other conditions and was not statistically significant. The by-problems analysis followed the exact same pattern as the by-participants analysis. This suggests that the effects observed in the by-participants analysis were consistent across problems and were not driven by any individual problem.

**Hindsight bias for situational judgments.** Panel A of Figure 2 displays the mean predictive and retrospective component importance scores as a function of problem type and solution success for the by-participants analysis. Panel B of Figure 2 displays the same data when scores are aggregated in a by-problems analysis. The size of the

**Table 2**
**Experiment 1 ANOVAs for Pairwise Contrasts of Predictive and**
**Retrospective Metacognitive Judgments (Likelihood Ratings),**
**As a Function of Problem Type and Solution Success**

| | | 95% Confidence Interval | | | |
|---|---|---|---|---|---|
| | Contrast | Upper Bound | Lower Bound | $F$ | $p$ |
| By-Participants Analysis | | | | | |
| Math | | | | | |
| Successful | 0.82 | 1.34 | 0.29 | 10.40 | .004 |
| Unsuccessful | 0.36 | 0.88 | −0.16 | 2.03 | .167 |
| Insight | | | | | |
| Successful | 0.70 | 1.22 | 0.18 | 7.62 | .011 |
| Unsuccessful | 0.66 | 1.18 | 0.13 | 6.71 | .016 |
| By-Problems Analysis | | | | | |
| Math | | | | | |
| Successful | 0.90 | 1.54 | 0.27 | 12.07 | .013 |
| Unsuccessful | 0.37 | 1.01 | −0.27 | 2.03 | .204 |
| Insight | | | | | |
| Successful | 0.80 | 1.43 | 0.16 | 9.44 | .021 |
| Unsuccessful | 0.60 | 1.24 | −0.04 | 5.34 | .060 |

Note—$MS_e$ for the rating × problem × success interaction was used for the calculation of confidence intervals and $F$ ratios (by-participants $MS_e$ = .802; by-problems $MS_e$ = .135). For the by-participants analysis, $df$ = 1,24; for the by-problems analysis, $df$ = 1,6.
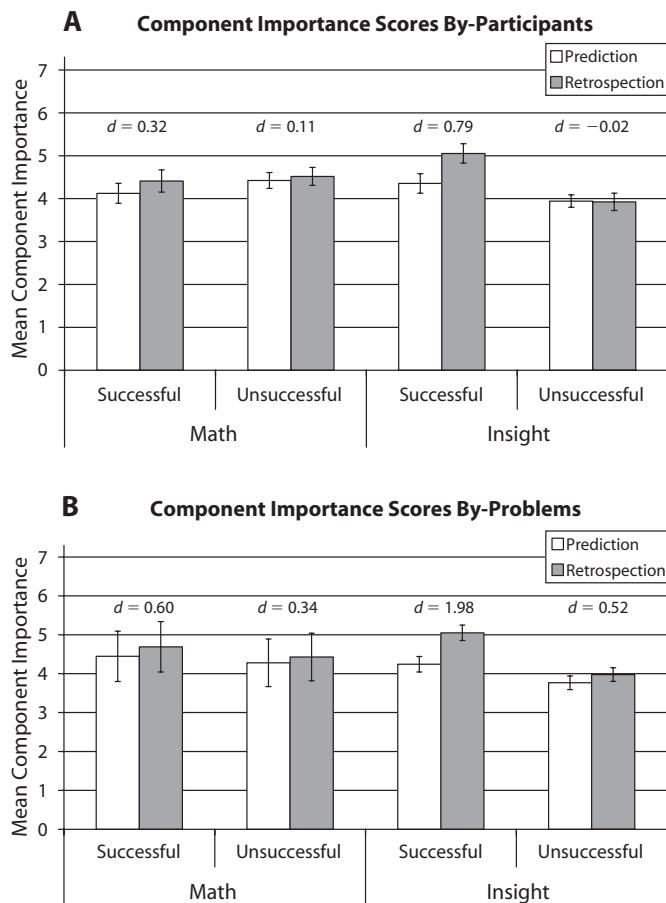


Figure 2. Experiment 1 mean predictive and retrospective situational judgments (component importance scores) as a function of problem type and solution success. Error bars = standard error; $d$ = Cohen's $d$.

**Table 3**
**Experiment 1 ANOVAs for Pairwise Contrasts of Predictive and**
**Retrospective Situational Judgments (Component Importance Scores),**
**As a Function of Problem Type and Solution Success**

| | | 95% Confidence Interval | | | |
| --- | --- | --- | --- | --- | --- |
| | Contrast | Upper Bound | Lower Bound | $F$ | $p$ |
| By-Participants Analysis | | | | | |
| Math | | | | | |
| Successful | 0.29 | 0.72 | −0.15 | 1.83 | .189 |
| Unsuccessful | 0.10 | 0.53 | −0.34 | 0.20 | .659 |
| Insight | | | | | |
| Successful | 0.70 | 1.14 | 0.26 | 10.85 | .003 |
| Unsuccessful | −0.02 | 0.42 | −0.46 | 0.01 | .921 |
| By-Problems Analysis | | | | | |
| Math | | | | | |
| Successful | 0.24 | 1.02 | −0.53 | 0.59 | .472 |
| Unsuccessful | 0.15 | 0.92 | −0.62 | 0.22 | .656 |
| Insight | | | | | |
| Successful | 0.81 | 1.58 | 0.03 | 6.52 | .043 |
| Unsuccessful | 0.21 | 0.98 | −0.56 | 0.44 | .532 |

Note—$MS_e$ for the rating × problem × success interaction was used for the calcula-
tion of confidence intervals and $F$ ratios (by-participants $MS_e$ = .562; by-problems
$MS_e$ = .200). For the by-participants analysis, $df$ = 1,24; for the by-problems analy-
sis, $df$ = 1,6.

hindsight effect for each category was calculated using Cohen's $d$ and is displayed over each pair of observations. The results of the four predictive versus retrospective judgment contrasts are displayed in Table 3 for both the by-participants and by-problems analyses.

From the by-participants analysis, significant hindsight bias on component importance judgments was detected only on successfully solved insight problems. The effects for all other problem type and solution conditions were substantially smaller and failed to approach statistical significance. The by-problems analysis followed the exact same pattern as did the by-participants analysis. This suggests that the effect observed on successfully solved insight problems was consistent across problems and was not driven by any individual problem.

## Discussion

In Experiment 1, the hindsight patterns observed on the two types of judgments qualitatively differed. On the metacognitive judgments (solution likelihood ratings), hindsight bias was observed on both math and insight problems. However, on situational judgments (component importance scores), hindsight bias was only observed on successfully solved insight problems. This pattern supports the notion that hindsight bias for these different types of judgments represents different psychological phenomena and requires different theoretical explanations. The pattern of results for metacognitive judgments in this experiment seems most consistent with the basic premises of the anchor and adjust theories that suggest that people use metacognitive responses to outcome information in order to reconstruct their judgments. On the other hand, the pattern of hindsight bias on the component importance ratings is more consistent with the updating and rejudging explanation of hindsight bias.

One interesting aspect of the results of this experiment was that the hindsight effects differed as a function of both problem type and solution success. On the metacognitive judgments, significant hindsight bias was detected in all conditions, except for unsuccessfully solved math problems. In the introduction to this experiment, we proposed that exposure to correct solutions should lead to an "I would have known that!" reaction, regardless of problem type. However, the observed pattern suggests that exposure to the solutions to unsuccessfully solved math problems did not evoke the same metacognitive reaction as it did in the other conditions.

Problem type and success also affected hindsight bias in the situational ratings. Although a substantial hindsight effect was detected on correctly solved insight problems, we found no evidence of hindsight bias on unsuccessfully solved insight problems. This moderating effect of solution success was not completely unexpected; it is consistent with evidence from prior insight studies that have shown that restructuring may only occur when a person actually solves an insight problem, and not simply as the result of being told the solution (Dominowski & Buyer, 2000).

Theories of the hindsight effect that describe it as an outcome knowledge effect would suggest that any knowledge of an outcome should be enough to produce the effect. However, the results of Experiment 1 suggest that there may be different sources of feedback that may lead to different kinds of hindsight effects: internal feedback that comes from self-generated solution attempts and outcome knowledge from external feedback, such as being told the correct solutions. In Experiment 2, we seek to better understand the role that internally generated outcome knowledge may play in producing hindsight effects.

## EXPERIMENT 2

The effects of solution success on the hindsight bias we found in Experiment 1 suggest that the externally provided outcome feedback was not the only factor leading to

changes in retrospective judgments. We propose that the act of actively attempting to solve a problem may provide internally generated feedback that may have independent effects on retrospective judgment formation. However, we were unable to find any studies that have specifically looked for hindsight bias effects on successfully and unsuccessfully solved problems in the absence of externally provided feedback. Therefore, in this study, we replicated the procedure of Experiment 1. However, after the participants attempted to solve the problems, we did not provide them with any solution feedback. There were two goals of this experiment. The first goal was to investigate the role of internally generated feedback in hindsight effects on metacognitive and situational judgments. The second goal was to see whether the effects caused by internally generated feedback would replicate the disassociations between metacognitive and situational judgments that were observed in Experiment 1. We predicted that hindsight bias on metacognitive judgments should follow a pattern consistent with anchor and adjust explanations of hindsight bias, and hindsight bias on situational judgments should follow a pattern consistent with updating and rejudging theories of hindsight bias. Therefore, we again predicted similar hindsight bias patterns on metacognitive judgments for both math and insight problems. However, we predicted hindsight bias on situational judgments only for successfully solved insight problems.

## Method

**Participants**. Forty-eight introductory psychology students from the University of Illinois at Chicago Subject Pool completed both sessions of this experiment for partial fulfillment of a class research participation requirement.

**Materials and Procedure**. All of the materials and the procedure used in this experiment were identical to those in Experiment 1. The only difference occurred at the end of Session 1. In this experiment, participants were dismissed from Session 1 directly after the problem solving phase of the experiment. Therefore, no one in this study was shown the step-by-step solutions, and no one received any external feedback about the accuracy of their solutions.

## Results

**Quasi-experimental design and analysis strategy**. Once again, we created a fully within-participants design by selecting only participants who correctly and incorrectly solved at least one of each of the problem types. Out of the 48 participants who completed the study, 30 had solution patterns that conformed to these criteria. A dependent samples *t* test comparing the mean proportion of successfully solved math and insight problems revealed no overall difference in solution success rates [math, $M = .45$, $SE = .035$; insight, $M = .43$, $SE = .032$; $t(29) = 0.65$, $p = .52$, $d = 0.12$]. Table 4 displays the number of successful, incorrect, and unsolved observations for each problem. Once again, there was considerable variability in success rates within the math and insight problems. Therefore, we aggregated the rating data both by participants and by problems in order to be able to assess whether hindsight bias patterns were consistent across participants and problems.

The by-participants analysis resulted in a 2 (judgment: predictive vs. retrospective) × 2 (problem type: insight

**Table 4**
**Experiment 2 Solution Rates (Frequency and Percentage)**

| | | | Unsuccessful | | | |
| | Successful | | Incorrect | | Unsolved | |
| | Freq. | % | Freq. | % | Freq. | % |
|---|---|---|---|---|---|---|
| **Math Problems** | | | | | | |
| Solve for Y | 8 | 27 | 10 | 33 | 12 | 40 |
| Cash | 11 | 37 | 16 | 53 | 3 | 10 |
| Distance | 22 | 73 | 5 | 17 | 3 | 10 |
| Food | 13 | 43 | 16 | 53 | 1 | 3 |
| Total | 54 | 45 | 47 | 39 | 19 | 16 |
| **Insight Problems** | | | | | | |
| Triangle | 12 | 40 | 10 | 33 | 8 | 27 |
| Glasses | 22 | 73 | 5 | 17 | 3 | 10 |
| Match III | 7 | 23 | 10 | 33 | 13 | 43 |
| Match XV | 10 | 33 | 6 | 20 | 14 | 47 |
| Total | 51 | 43 | 31 | 26 | 38 | 32 |

Note—$N = 30$.

vs. math) × 2 (solution success: successful vs. unsuccessful) repeated measures design with 30 observations per cell. The by-problems analysis resulted in a 2 (judgment: predictive vs. retrospective) × 2 (problem type: insight vs. math) × 2 (solution success: successful vs. unsuccessful) mixed design with problem type as a between-problems variable and four observations per cell. In order to investigate the effects of problem type and solution success on hindsight bias, we conducted a retrospective–predictive judgment contrast at alpha level = .05 for each of the problem type × solution success conditions.

**Hindsight bias for metacognitive judgments**. Panel A of Figure 3 displays the mean predictive and retrospective likelihood ratings as a function of problem type and solution success for the by-participants analysis. Panel B of Figure 3 displays the same data when ratings are aggregated in a by-problems analysis. The size of the hindsight effect for each category was calculated using Cohen's *d* and is displayed over each pair of observations. The results of the four predictive versus retrospective judgment contrasts are displayed in Table 5 for both the by-participants and by-problems analyses.

The by-participants analysis revealed a large hindsight bias on successfully solved math problems and a moderate hindsight bias on successfully solved insight problems. In both of these conditions, participants' mean retrospective likelihood ratings were higher than their presolution-attempt predictive ratings. This pattern replicates the Experiment 1 finding of hindsight bias effects for successfully solved math and insight problems. As in Experiment 1, there was no significant hindsight bias effect detected on unsuccessfully solved math problems. However, unsuccessfully solved insight problems followed a different pattern from that seen in Experiment 1. In this experiment, participants showed a reverse hindsight bias, producing significantly lower retrospective likelihood ratings after unsuccessful solution attempts on insight problems. The by-problems analysis revealed the same overall pattern of results as did the by-participants analysis, although in this analysis, only the large hindsight effect observed on successfully solved math problems reached statistical significance.
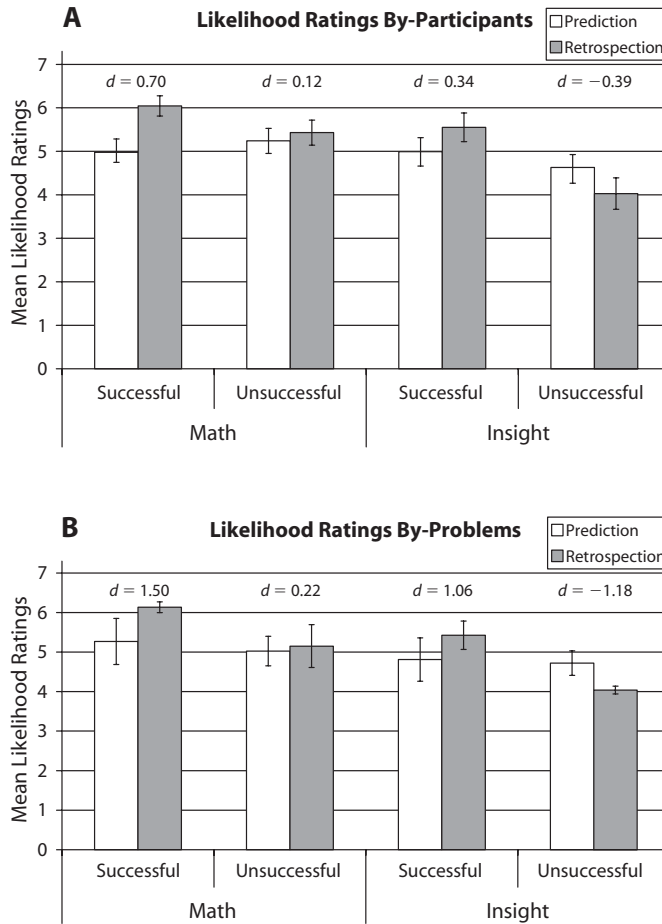
Figure 3. Experiment 2 mean predictive and retrospective metacognitive judgments (solution likelihood ratings) as a function of problem type and solution success. Error bars = standard error; *d* = Cohen's *d*.

**Table 5**
**Experiment 2 ANOVAs for Pairwise Contrasts of Predictive**
**and Retrospective Metacognitive Judgments (Likelihood Ratings),**
**As a Function of Problem Type and Solution Success**

|  | Contrast | 95% Confidence Interval | | *F* | *p* |
|---|---|---|---|---|---|
|  |  | Upper Bound | Lower Bound |  |  |
| By-Participants Analysis | | | | | |
| Math | | | | | |
|   Successful | 1.07 | 1.53 | 0.60 | 21.85 | .000 |
|   Unsuccessful | 0.19 | 0.66 | −0.28 | 0.70 | .410 |
| Insight | | | | | |
|   Successful | 0.56 | 1.03 | 0.09 | 6.00 | .021 |
|   Unsuccessful | −0.60 | −0.13 | −1.07 | 6.92 | .014 |
| By-Problems Analysis | | | | | |
| Math | | | | | |
|   Successful | 0.87 | 1.71 | 0.03 | 6.35 | .045 |
|   Unsuccessful | 0.12 | 0.97 | −0.72 | 0.13 | .731 |
| Insight | | | | | |
|   Successful | 0.61 | 1.46 | −0.23 | 3.19 | .124 |
|   Unsuccessful | −0.68 | 0.16 | −1.53 | 3.94 | .094 |

Note—$MS_e$ for the rating × problem × success interaction was used for the calculation of confidence intervals and *F* ratios (by-participants $MS_e$ = .782; by-problems $MS_e$ = .237). For the by-participants analysis, *df* = 1,29; for the by-problems analysis, *df* = 1,6.

**Hindsight bias for situational judgments**. Panel A of Figure 4 displays the mean predictive and retrospective component importance scores as a function of problem type and solution success for the by-participants analysis. Panel B of Figure 4 displays the same data when scores are aggregated in a by-problems analysis. The size of the hindsight effect for each category was calculated using Cohen's *d* and is displayed over each pair of observations. The results of the four predictive versus retrospective judgment contrasts are displayed in Table 6 for both the by-participants and by-problems analyses.

As in Experiment 1, the by-participants analysis detected no evidence of hindsight bias on component importance scores for successfully or unsuccessfully solved math problems. Likewise, a significant hindsight effect was detected on successfully solved insight problems. After successfully solving insight problems, participants' retrospective component importance scores were indicative of a more appropriate problem representation than that indicated by their initial predictive scores. The opposite pattern was observed on unsuccessfully solved insight problems. In this condi-

tion, participants' retrospective component importance scores were significantly lower than their initial scores. Therefore, participants' retrospective judgments suggested a less appropriate problem representation on unsuccessfully solved insight problems. The by-problems analysis revealed the same pattern as the by-participants analysis. There was one notable difference, though. In the by-problems analysis, the magnitude of the hindsight effect for successfully solved insight problems appeared larger than the effect for unsuccessfully solved problems. However, when we aggregated the data across problems, the magnitude of the effects was similar. This suggests that the overall pattern is consistent across problems. However, there may have been some differences in the magnitude of hindsight effects among insight problems, which led to the differences in effect size that were observed in the by-participants analysis.

## Discussion

The results of this experiment showed that hindsight effects are not confined to situations in which individuals receive external outcome feedback. In this experiment,



**A   Component Importance Scores By-Participants**
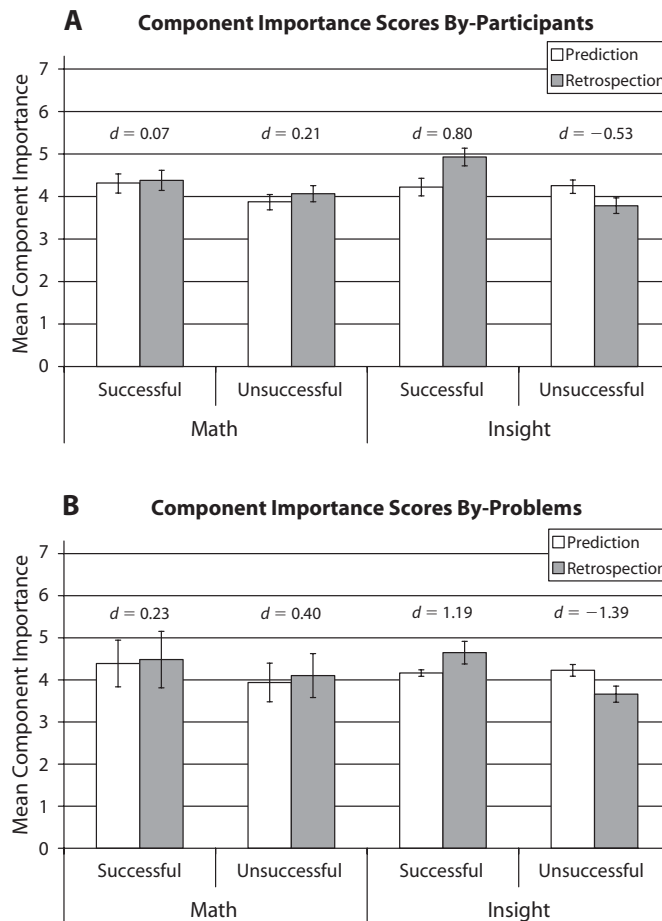
**B   Component Importance Scores By-Problems**

Figure 4. Experiment 2 mean predictive and retrospective situational judgments (component importance scores) as a function of problem type and solution success. Error bars = standard error; *d* = Cohen's *d*.

**Table 6**
**Experiment 2 ANOVAs for Pairwise Contrasts of Predictive and**
**Retrospective Situational Judgments (Component Importance Scores),**
**As a Function of Problem Type and Solution Success**

| | | 95% Confidence Interval | | | |
| --- | --- | --- | --- | --- | --- |
| | Contrast | Upper Bound | Lower Bound | $F$ | $p$ |
| By-Participants Analysis | | | | | |
| Math | | | | | |
|   Successful | 0.06 | 0.39 | −0.26 | 0.15 | .701 |
|   Unsuccessful | 0.19 | 0.52 | −0.14 | 1.42 | .243 |
| Insight | | | | | |
|   Successful | 0.71 | 1.03 | 0.38 | 19.66 | .000 |
|   Unsuccessful | −0.47 | −0.15 | −0.80 | 8.82 | .006 |
| By-Problems Analysis | | | | | |
| Math | | | | | |
|   Successful | 0.09 | 0.58 | −0.40 | 0.21 | .663 |
|   Unsuccessful | 0.16 | 0.65 | −0.33 | 0.67 | .444 |
| Insight | | | | | |
|   Successful | 0.48 | 0.97 | −0.01 | 5.84 | .052 |
|   Unsuccessful | −0.57 | −0.08 | −1.06 | 8.02 | .030 |

Note—$MS_e$ for the rating × problem × success interaction was used for the calculation of confidence intervals and $F$ ratios (by-participants $MS_e$ = .381; by-problems $MS_e$ = .080). For the by-participants analysis, $df$ = 1,29; for the by-problems analysis, $df$ = 1,6.

participants were not shown the correct solution to the problems. However, the internally generated feedback that resulted from successfully solving math or insight problems resulted in hindsight bias on metacognitive judgments. In these conditions, the participants' retrospective ratings overestimated their predictive likelihood ratings. Effect sizes suggested that this hindsight bias was larger on successfully solved math problems than for successfully solved insight problems. On unsuccessfully solved insight problems, retrospective ratings underestimated original likelihood predictions, resulting in a reverse hindsight bias. Therefore, in these three conditions, hindsight bias on metacognitive judgments was in the direction of the internally generated feedback produced by the successful or unsuccessful solving attempts. Once again, we propose that these results are consistent with anchor and adjust theories of retrospective judgment formation. The results on successfully solved problems suggest that the internally generated feedback that comes with accepting a successful final solution can produce the "I would have known that!" feeling which, according to this theory, leads to the hindsight bias (Hoch & Loewenstein, 1989; Ofir & Mazursky, 1990, 1997).

The larger hindsight bias effect on math versus insight problems suggested by the effect sizes (see Figure 2) was also informative. In this experiment, we selected math problems as stimulus materials to represent a familiar problem solving domain. It may be that the familiar nature and procedural solution process for these problems allowed for more salient, or *certain*, internally generated feedback. However, the insight problems were selected to represent unfamiliar and misleading problem situations for which individuals do not possess algorithmic strategies for verifying the veracity of their solutions. Therefore, the internally generated feedback in these situations may be less salient or lead to less certainty that the correct solution had been obtained. These observations are consistent with Schwarz and Stahlberg's (2003) suggestion that familiarity or exper-

tise in a domain may lead to greater hindsight bias effects on these types of metacognitive judgments.

It appears that the saliency of internally generated feedback may be reversed for unsuccessfully solved problems. For math problems, failure cues may not be as salient as for insight problems. Failing to solve math problems may happen for mundane reasons, such as miscalculations, that may not even be noticed by the solver. Thus, salient cues may not exist when one fails to solve a math problem. This would explain why no hindsight bias effects were detected on unsuccessfully solved math problems in either experiment. However, for unsuccessfully solved insight problems, the internal cues associated with failure may be very apparent. In fact, this internal recognition of "being stuck" or reaching an impasse plays a central role in some theories of the insightful problem solving process (e.g., Knoblich, Ohlsson, & Raney, 2001). This recognition of being at an impasse could lead to an "I never would have known that!" feeling, which would explain the direction of the hindsight effect we observed on unsuccessfully solved insight problems. As in Experiment 1, the hindsight bias patterns on these metacognitive judgments are best explained by the anchor and adjust theory of hindsight effects.

The situational judgment results followed a qualitatively different pattern. On these measures, individuals exhibited hindsight effects only on insight problems. On successfully solved insight problems, retrospective ratings were biased toward a more appropriate problem representation than were predictive ratings. On unsuccessfully solved insight problems, retrospective ratings were biased toward a less appropriate problem representation. The pattern is more consistent with the update and rejudge explanation of retrospective judgment formation. Representational change, or *restructuring*, has been proposed as playing a central role in the insightful solution process (Ash & Wiley, 2006; Jones, 2003; Kaplan & Simon, 1990; Knoblich et al., 1999; Ohlsson, 1992). Therefore, it is par-

ticularly on insight problems that we would expect solvers to have different presolution and postsolution representations. The best explanation for the hindsight bias pattern on these judgments is that individuals were using their final problem representations to reconstruct their initial predictive judgments. In the case of successfully solved problems, the participants' post solution representations were more appropriate for solving the problem, because the solvers had restructured their initial representations as they solved the problem. In the case of unsuccessfully solved problems, the postsolution representations were less appropriate for solving problems.

The same effects were not seen for the situational judgments on math problems. Because the difficulty on these math problems should not stem from inappropriate representations, we would not expect the participants' postsolution representations for math problems to be structurally different from their initial representations. According to the updating and rejudging theory, if there is no representational change between the time of the predictive and the retrospective judgments, then the judgments reconstructed from the current representation should not be systematically biased, because little has been updated. On the basis of these results, we propose that updating and rejudging theories, once again, provide the most plausible explanation for the hindsight effect patterns observed on situational judgments.

## GENERAL DISCUSSION

In the present experiments, we investigated whether hindsight effects observed on different types of judgment tasks could be attributed to a unitary process or, alternatively, whether hindsight effects could be better explained by judgment-specific cognitive reconstruction theories. Both experiments found distinct patterns of hindsight for different problem types and judgment tasks. The results of both experiments supported the notion that anchor and adjust theories provide the most plausible explanation of hindsight bias for metacognitive judgments and that updating and rejudging accounts provide the most plausible explanation of hindsight bias for situational judgments. Thus, the results offer evidence that two competing models of hindsight bias are actually complementary explanations for judgment reconstruction on different types of tasks.

A novel result from this study came from the manipulation of externally provided outcome knowledge across the two experiments. In most studies of hindsight effects, participants are given outcome knowledge, such as the ending of a story, the answer to the trivia questions, or feedback on whether they had guessed or answered correctly. The general finding is that hindsight judgments are biased in the direction of the given outcome. However, most hindsight studies that include a performance aspect, such as trivia question paradigms, simply examine the effects of solution feedback. Therefore, the analyses in these studies often collapse hindsight results across successful and unsuccessful solvers. The present study shows that successful solution attempts lead to the same hindsight effects as externally provided feedback. Therefore, this calls into question whether hindsight effects should be viewed as outcome

knowledge effects. We propose that hindsight effects are best conceptualized as being a by-product of judgment reconstruction processes and are not necessarily an outcome-knowledge driven phenomenon (see Schwarz & Stahlberg, 2003, for a similar discussion of this issue). Furthermore, the differences between hindsight effects on incorrectly solved insight problems in the two experiments suggest that a failure to detect hindsight effects does not necessarily suggest that externally provided outcome knowledge has no effect on retrospective judgments. On unsuccessfully solved insight problems for which participants were later shown the solutions, participants' retrospective situational judgments were similar to their initial judgments. This seems to suggest that outcome knowledge had no effect on the retrospective judgment process. However, when we consider that participants who were not given feedback on these problems showed significantly lower retrospective situational judgments, it seems that solution feedback did indeed have an effect in Experiment 1.

The implications of our discovery of internally generated hindsight effects can be illustrated by comparing our results to those of a previous study, which used insight and incremental problems to investigate the effects of outcome feedback on hindsight effect. Hoch and Loewenstein (1989, Experiment 5) used a between-participants paradigm to investigate differences in hindsight bias on insight and incremental problems.[4] In their study, participants were presented with two problems designed to require insightful solution processes and two problems designed to require incremental solution processes. All participants attempted to solve the problems for 3 min. Afterwards, half of the participants were immediately asked to predict the percentage of their peers who would correctly solve the problem (no-feedback, or *predictive* condition), and the other half were shown the correct solutions to the problems, and then asked to make the same judgment (feedback, or *retrospective* condition). Their results showed a clear hindsight bias effect on the insight problems and a reverse hindsight bias effect on the incremental problems.

At face value, these results seem to be inconsistent with the metacognitive judgment results found in the present Experiment 1, which showed positive hindsight bias effects on metacognitive judgments for both types of problems. However, we propose that the Hoch and Loewenstein (1989) results were actually just an artifact caused by the confounding of internal and external feedback in their design. In their investigation, successful and unsuccessful solvers were grouped together in the no-feedback group. In Experiment 2, we showed that internally generated feedback led to a large hindsight effect on metacognitive judgments for successfully solved math problems, as well as to a negative hindsight bias effect on unsuccessfully solved insight problems. Therefore, the pattern of hindsight bias effects observed by Hoch and Loewenstein's investigation may have been driven by the effects of internally generated feedback in their between-participants feedback groups. This example illustrates the importance of controlling for solution success and manipulating external feedback when investigating hindsight effects in performance-based paradigms, such as

problem solving or trivia tasks. Furthermore, this comparison suggests that results from between-participants hindsight studies should be interpreted with caution when the design allows for the possibility of uncontrolled internally generated feedback.

## Conclusions

The present study supports the general explanation that the hindsight bias is caused by people's attempts to reconstruct their original judgments on the basis of cues that are affected by outcome feedback. Although the preference for parsimony may lead us to search for a unified explanation of all hindsight bias effects, the empirical evidence suggests that hindsight bias may be best thought of as a taxonomic category describing a set of retrospective judgment phenomena. Here we have identified two classes of judgments, metacognitive and situational, which require different judgment reconstruction processes in order to explain the observed patterns of hindsight bias. Hence, we conclude that the anchor and adjust and update and rejudge theories should be viewed as complementary explanations of hindsight bias on different types of judgment tasks, not as competing accounts of the same general retrospective judgment bias.

### REFERENCES

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Arkes, H. R., Faust, D., Guilmette, T. J., & Hart, K. (1988). Eliminating the hindsight bias. *Journal of Applied Psychology*, **73**, 305-307.

Ash, I. K., & Wiley, J. (2006). The nature of restructuring in insight: An individual-differences approach. *Psychonomic Bulletin & Review*, **13**, 66-73.

Carli, L. L. (1999). Cognitive reconstruction, hindsight, and reactions to victims and perpetrators. *Personality & Social Psychology Bulletin*, **25**, 966-979.

Christensen-Szalanski, J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior & Human Decision Processes*, **48**, 147-168.

Dominowski, R. L., & Buyer, L. S. (2000). Retention of problem solutions: The re-solution effect. *American Journal of Psychology*, **113**, 249-274.

Duncker, K. (1945). On problem-solving. *Psychological Monographs*, **58**(No. 270).

Durso, F. T., Rea, C. B., & Dayton, T. (1994). Graph-theoretic confirmation of restructuring during insight. *Psychological Science*, **5**, 94-98.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under certainty. *Journal of Experimental Psychology: Human Perception & Performance*, **1**, 288-299.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once-future things. *Organizational Behavior & Human Performance*, **13**, 1-16.

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic & Applied Social Psychology*, **26**, 103-117.

Hasher, L., Attig, M. S., & Alba, J. W. (1981). I knew it all along: Or, did I? *Journal of Verbal Learning & Verbal Behavior*, **20**, 86-96.

Hawkins, S. A., & Hastie, R. (1990). Hindsight biased judgments of past events after the outcomes are known. *Psychological Bulletin*, **107**, 311-327.

Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. *Psychological Review*, **104**, 194-202.

Hoch, S. J., & Loewenstein, G. F. (1989). Outcome feedback: Hindsight and information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 605-619.

Jones, G. (2003). Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 1017-1027.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, **22**, 374-419.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York: Cambridge University Press.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 1534-1555.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & Cognition*, **29**, 1000-1009.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 107-118.

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.

Metcalfe, J. (1986a). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 288-294.

Metcalfe, J. (1986b). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 623-634.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, **15**, 238-246.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Ofir, C., & Mazursky, D. (1990). "I could never have expected it to happen": The reversal of the hindsight bias. *Organizational Behavior & Human Decision Processes*, **46**, 20-23.

Ofir, C., & Mazursky, D. (1997). Does a surprising outcome reinforce or reverse the hindsight bias? *Organizational Behavior & Human Decision Processes*, **69**, 51-57.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomenon. In M. Keane & K. Gilhooley (Eds.), *Advances in the psychology of thinking* (pp. 1-44). London: Harvester Wheatsheaf.

Pezzo, M. V. (2003). Surprise, defence, or making sense: What removes hindsight bias? *Memory*, **11**, 421-441.

Pohl, R. F., & Hell, W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational Behavior & Human Decision Processes*, **67**, 49-58.

Roese, N. J., & Olson, J. M. (1996). Counterfactuals, causal attributions, and the hindsight bias: A conceptual integration. *Journal of Experimental Social Psychology*, **32**, 197-227.

Ross, B. H., & Kilbane, M. C. (1997). Effects of principle explanation and superficial similarity on analogical mapping in problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 427-440.

Schwarz, S., & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of meta-cognitions. *Memory*, **11**, 395-410.

Stahlberg, D., & Maass, A. (1998). Hindsight bias: Impaired memory

or biased reconstruction? In W. Stroebe & M. Hewstone (Eds.), *European Review of Social Psychology* (Vol. 8, pp. 105-132). Chichester, U.K.: Wiley.

WASSERMAN, D., LEMPERT, R. O., & HASTIE, R. (1991). Hindsight and causality. *Personality & Social Psychology Bulletin, 17*, 30-35.

WERTH, L., STRACK, F., & FÖRSTER, J. (2002). Certainty and uncertainty: The two faces of the hindsight bias. *Organizational Behavior & Human Decision Processes, 87*, 323-341.

WILEY, J. (1998). Expertise as mental set: The effects of domain knowledge in creative problem solving. *Memory & Cognition, 26*, 716-730.

WILEY, J., & TRABASSO, T. (2003, June). *Hindsight bias as a function of memory updating during story comprehension.* Paper presented at the 13th Annual Meeting of the Society for Text and Discourse, Madrid, Spain.

### NOTES

1. In a separate study, 47 introductory psychology students completed all eight problems (four insight, four math) while performing a think-aloud protocol procedure. Video recordings were coded for instances of impasse by two independent coders (kappa = .92). Analysis of the percentage of problems categorized as displaying an impasse revealed that participants were 2.77 times more likely to exhibit signs of impasse on the insight ($M = 61$, $SD = 21$) than on the math ($M = 22\%$, $SD = 29$) problems [$t(46) = 7.55$, $p < .001$, $\eta^2 < .55$]. Since impasse during problem solving has been associated with representation difficulties and restructuring (e.g., Knoblich, Ohlsson, & Raney, 2001), we considered this to be evidence of the validity of our problem-type manipulation.

2. Because of the low number of incorrect and unsolved observations across the different problems, it was impossible to examine hindsight effects as a function of these two solution types. Therefore, we collapsed across unsuccessful solutions in all hindsight effect analyses.

3. Meta-analytic reviews of the literature have provided ample evidence that hindsight bias is a ubiquitous and reliable effect that is extremely difficult to eliminate (Christensen-Szalanski & Willham, 1991; Guilbault, Bryant, Brockway, & Posavac, 2004). Because of this, we concluded that any significant hindsight bias effects detected in this study were highly unlikely to be Type I errors. The primary diagnostic prediction of our theory is that hindsight bias should be less likely on situational judgments for the math problems. Therefore, the probability of Type II error (i.e., failing to detect a hindsight effect for situational judgments on math problems when it exists in the population) was deemed more of a threat to the internal validity of this study than was the probability of Type I error. This is why we chose the more powerful contrast analyses, using alpha level = .05.

4. In within-participants paradigms, preoutcome predictive and postoutcome retrospective judgments are collected from the same individuals, and hindsight effects are operationalized as the difference between these judgments (e.g., Fischhoff & Beyth, 1975). Other studies in the hindsight literature have used between-participants or *hypothetical* designs (see Hawkins & Hastie, 1990). In such designs, one group of participants is asked to make predictive judgments before receiving outcome information, and a second group is asked to make retrospective judgments after receiving outcome information and being asked to ignore the information. In these designs, hindsight effects are operationalized as the difference in judgments between the no-feedback and outcome feedback groups (for more discussions of differences between within-participants and between-participants hindsight bias designs, see Hertwig, Gigerenzer, & Hoffrage, 1997; Stahlberg & Maass, 1998; Werth, Strack, & Förster, 2002).

### APPENDIX

For each problem, the individual elements below the instructions were used in the component importance ratings. Elements that are underlined in each problem were coded as *relevant components*. All other elements were coded as *irrelevant components*. No elements were underlined in the problems that were presented to participants.

**Math Problems**

1. Solve for Y. Find the exact number that the variable Y equals by using only the necessary equations from the set of equations below.

$$\underline{3Z * 3 = 27} \qquad \underline{2C - 9 = Z} \qquad P - C = 2\,D$$

$$\underline{5Z - 11 = M} \qquad 2X = 56 + A \qquad \underline{8M - C = Y}$$

$$3Y + 14 = X$$

2. Bob left home with an ATM card, a credit card, a checkbook, and $70 in his wallet. Below is a list of the purchases that he made during the day. How much money did Bob have in his wallet at the end of the day?

| Location | Type of Purchase | Amount ($) | Tax ($) | Tip ($) |
|---|---|---|---|---|
| Corner Diner | Credit | 15.72 | 1.26 | 2.50 |
| Quick Mart | Credit | 9.75 | 0.78 | * |
| Rapid Transit | Cash | 1.50 | * | * |
| Sports Inc. | Check | 75.00 | 6.00 | * |
| Fast Food Hut | Cash | 8.72 | 0.69 | * |
| Movie Plaza | Credit | 11.00 | * | * |
| Movie Concession | Cash | 15.20 | 1.22 | * |
| Rapid Transit | Cash | 1.50 | * | * |
| Mickey's Pub | Credit | 22.50 | 1.80 | 4.50 |

**APPENDIX (Continued)**

3. Jane and Frank are driving from San Francisco to New York City. They have a chart that lists distances between cities and maximum speed limits on the connecting highways. Jane wants to use this chart to figure out how far it is from San Francisco to New York City. What is the distance they must drive in order to get from San Francisco to New York City?

| From | To | Number of Miles | Speed Limit |
|---|---|---|---|
| Austin, TX | Kansas City, MO | 445 | 55 MPH |
| Boston, MA | New York, NY | 280 | 65 MPH |
| Cheyenne, WY | Omaha, NE | 500 | 70 MPH |
| Chicago, IL | Toledo, OH | 300 | 70 MPH |
| Detroit, MI | Chicago, IL | 455 | 65 MPH |
| Davenport, IA | Chicago, IL | 225 | 55 MPH |
| Omaha, NE | Davenport, IA | 325 | 65 MPH |
| San Francisco, CA | Cheyenne, WY | 450 | 65 MPH |
| San Francisco, CA | L.A., CA | 255 | 45 MPH |
| Toledo, OH | New York, NY | 565 | 65 MPH |
| Jacksonville, FL | Washington, D.C. | 660 | 55 MPH |

4. Larry (age 35), his wife June (age 34) and his son Kenny (age 15) are going on a three-month camping trip in Alaska. Larry has heard stories of people getting snowed in during camping trips and not having enough food to survive. He wanted to be sure to send enough supplies to the cabin before he leaves on the trip. Larry got a list of the amount of supplies needed per-day by people of different age groups. In total, exactly how many pounds of food supplies will his family need per day?

| Age and Gender | Amount Needed per Day (in Pounds) | | | |
| | Bread | Vegetables | Meat | Water |
|---|---|---|---|---|
| Male child | 0.9 | 1.0 | 1.2 | 1.9 |
| Female senior citizen | 4.3 | 3.2 | 3.6 | 6.5 |
| Male baby | 0.5 | 0.8 | 0.4 | 1.0 |
| Male teen | 2.5 | 3.0 | 3.5 | 4.2 |
| Female teen | 2.5 | 3.5 | 3.0 | 4.4 |
| Female adult | 4.0 | 3.7 | 3.5 | 5.5 |
| Male senior citizen | 4.4 | 4.5 | 2.0 | 6.7 |
| Female baby | 0.5 | 0.8 | 0.6 | 1.1 |
| Male adult | 3.5 | 4.0 | 3.3 | 6.0 |
| Female child | 0.9 | 1.4 | 1.0 | 2.0 |

**Insight Problems**

5. The triangle shown below points to the top of the page. Show how you can move 3 circles to get the triangle to point to the bottom of the page.
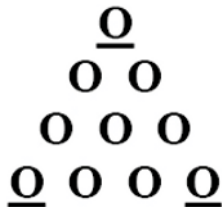
**APPENDIX (Continued)**

6. The picture below is of six glasses. The first three contain liquid. Describe how you could make it so no two glasses containing liquid are next to each other, while keeping three of the six glasses full. To do this, you are only allowed to move **one** glass.

Note—The second cup is used in both the correct solution and the most common incorrect solution (moving the second cup to the end). Therefore, this component was not used in the calculation of relevant or irrelevant components.

7. Below is a picture of an equation in which the Roman numerals are constructed using matchsticks. Notice that both sides of the equation are not mathematically equal. Describe how you could make both sides equal by moving **only one matchstick**. The rules are that: (a) only one stick is to be moved; (b) a stick cannot be discarded; that is, it can only be moved from one position in the equation to another; (c) a slanted stick cannot be interpreted as a vertical matchstick; and (d) the result must be a correct arithmetic statement.
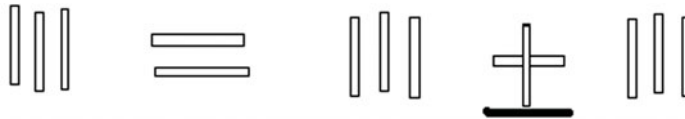
8. Below is a picture of an equation in which the Roman numerals are constructed using matchsticks. Notice that both sides of the equation are not mathematically equal. Describe how you could make both sides equal by moving **only one matchstick**. The rules are that: (a) only one stick is to be moved; (b) a stick cannot be discarded; that is, it can only be moved from one position in the equation to another; (c) a slanted stick cannot be interpreted as a vertical matchstick; and (d) the result must be a correct arithmetic statement.
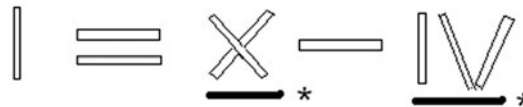
Note—This problem has two different solutions. Relevant components were calculated using the X term if participants used this term in solving. The participants in Experiment 1 were given the answer that involved the IV term. Therefore, for all other participants, the IV term was used as the key component. Analyses using either or both of these components of the problem as relevant components did not change the overall pattern of results.