

Hindsight Bias: A By-Product of Knowledge Updating?

Ulrich Hoffrage, Ralph Hertwig, and Gerd Gigerenzer
Max Planck Institute for Human Development

With the benefit of feedback about the outcome of an event, people's recalled judgments are typically closer to the outcome of the event than their original judgments were. It has been suggested that this *hindsight bias* may be due to a reconstruction process of the prior judgment. A model of such a process is proposed that assumes that knowledge is updated after feedback and that reconstruction is based on the updated knowledge. Consistent with the model's predictions, the results of 2 studies show that knowledge after feedback is systematically shifted toward feedback, and that assisting retrieval of the knowledge prior to feedback reduces hindsight bias. In addition, the model accounts for about 75% of cases in which either hindsight bias or reversed hindsight bias occurred. The authors conclude that hindsight bias can be understood as a by-product of an adaptive process, namely the updating of knowledge after feedback.

In attempting to understand the past, historians have to deal with a number of methodological problems. The problem that concerns us here stems from what Leo Tolstoy (1869/1982) described in *War and Peace* as the "law of retrospectiveness, which makes all the past appear a preparation for events that occur subsequently" (p. 843). Tolstoy speculated that this law explains why Russian historians, writing after Napoleon's defeat, believed that the Russian generals deliberately lured Napoleon to Moscow (and defeat), although the evidence points to luck rather than deliberate planning. What Tolstoy called the law of retrospectiveness may also have inspired experimental psychologists. In fact, when Fischhoff (1975) began to study postevent memories, he referred to this methodological dilemma of historical research by quoting the historian Roberta Wohlstetter (1962):

It is much easier after the event to sort the relevant from the irrelevant signals. After the event, of course, a signal is always crystal clear. We can now see what disaster it was signaling since the disaster has occurred, but before the event it is obscure and pregnant with conflicting meanings. (p. 387)

If this historian's intuition is true, *hindsight judgments*, that is, judgments made with benefit of feedback about the

outcome of an event, should differ systematically from *foresight judgments*, that is, judgments made without knowledge of the outcome. Indeed, this is what Fischhoff (1975) found. He presented participants with historical scenarios, for instance, the 19th-century war between the British and the Gurkhas of Nepal. In the foresight condition, participants had to give confidence ratings for four possible outcomes, without knowing which of them had actually occurred. In the hindsight condition, participants were told the actual outcome and then asked to state their hypothetical confidence in all four possible outcomes, that is, the confidence they would have given had they not been told the actual outcome. Participants with hindsight were more confident about the actual outcome than those with foresight.

This phenomenon has been called *hindsight bias* or the "knew-it-all-along effect." It has been investigated in a number of studies, by using either a hypothetical or a memory design. With the hypothetical design (e.g., Fischhoff, 1975), two groups of participants are compared: One group has no outcome knowledge, and the other has outcome knowledge but is asked to ignore it. With the memory design, a comparison is made between the original and recalled answers of one group of participants. First, participants make a series of judgments about the outcome of certain events; second, they receive outcome information; and third, they have to recall their original answers. Note that the experimental condition of the hypothetical design approximates the situation of historians who normally write about an historical event without having given an assessment prior to its occurrence. In contrast, the experimental condition of the memory design approximates everyday situations in which individuals predict an event, receive feedback, and then eventually remember their judgment (e.g., elections, weather, etc.). Hertwig, Gigerenzer, and Hoffrage (1997) argued that the effects obtained in the two designs are systematically different, and they proposed reserving the term *hindsight bias* for an effect obtained in a memory design and the term *knew-it-all-along effect* for the hypothetical design. In this article, we adopt this distinction.

Ulrich Hoffrage, Ralph Hertwig, and Gerd Gigerenzer, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Berlin, Germany.

We are grateful to Hartmut Blank, Edgar Erdfelder, Klaus Fiedler, Reid Hastie, Peter Mueser, Rüdiger Pohl, members of the ABC Research Group, and two anonymous reviewers for helpful comments; to Heinz Mayringer for helping to collect the data; and to Anita Todd and Jill Vyse for editing the manuscript. We also thank the Deutsche Forschungsgemeinschaft (Grants Ho 1847/1 and SFB 504) for their financial support.

Correspondence concerning this article should be addressed to Ulrich Hoffrage, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. Electronic mail may be sent to hoffrage@mpib-berlin.mpg.de.

Hindsight bias and the knew-it-all-along effect have been identified in a wide range of task types, including confidence judgments in the outcome of events, choices between alternatives, and estimations of quantities as well as in a variety of domains, such as political events (Pennington, 1981), medical diagnosis (Arkes, Wortmann, Saville, & Harkness, 1981), outcomes of scientific experiments (Slovic & Fischhoff, 1977), economic decisions (Bukhszar & Connolly, 1988), autobiographical memory (Neisser, 1981), and general knowledge (Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988). Although the overall magnitude of the effects is small (according to a meta-analysis conducted by Christensen-Szalanski & Fobian Willham, 1991, $r = .17$, corrected for reliability, $r = .25$), the hindsight bias appears to be robust and difficult to eliminate (Fischhoff, 1982).

In their review, Hawkins and Hastie (1990) listed four general strategies for responding to the request for a hindsight judgment (for a discussion of these strategies, see Erdfelder & Buchner, 1998). Hawkins and Hastie concluded that the first 2 strategies—"direct recall of the original belief" and "anchor on the current belief and adjust to infer the original belief"—do not play an important role in explanations of results obtained in hindsight bias research. In contrast, the third and fourth strategy—"cognitive reconstruction" and "motivated self-presentation" (p. 320)—have been implicated in many of the findings they have reviewed.¹ Most promising, in Hawkins and Hastie's opinion, are those cognitive accounts where a hindsight judgment is seen as a "reconstruction of the prior judgment by 'rejudging' the outcome" (p. 321). In this view, hindsight bias emerges because of systematic differences between judging and rejudging the outcome. According to Stahlberg and Maass (1998), these differences are due to metacognitive processes; that is, participants who have forgotten their original estimates "are forced to guess and, in the presence of outcome information, are likely to utilize this information as an anchor, assuming that their estimates must have been somewhere in the proximity of the true outcome" (p. 110). However, it seems fair to say that neither this metacognition interpretation (inspired by McCloskey & Zaragoza, 1985) nor other interpretations of the cognitive reconstruction notion have specified a precise mechanism (for an exception, see Pohl & Eisenhauer, 1997). The model we propose offers such a mechanism: It allows us to explain at the level of individual responses (i.e., individual items for individual participants) why the effect occurred, did not occur, or even was reversed.

Previously (Hertwig et al., 1997), we proposed a model that assumed that observed hindsight bias results from the sum of true hindsight bias and the reiteration effect, that is, the phenomenon that mere repetition of an assertion increases confidence in the correctness of the assertion. This model accounts for the fact that observed hindsight bias is larger for assertions with "this assertion is true" feedback than for assertions with "this assertion is false" feedback, but it does not explain why there is true hindsight bias in the first place. The current model extends this previous work in two respects: It accounts for true hindsight bias, and it does so at the level of individual responses. In this article, we

outline the model, derive three predictions from it, and report two studies that tested these predictions.

Reconstruction After Feedback With Take The Best (RAFT)

We explain and subsequently test the model with a task in which an original response is made at Time 1, feedback about the correct answer is given at Time 2, and the original response has to be recalled at Time 3 (recalled response). In developing the present model, we were inspired by the theory of probabilistic mental models (PMM; see Gigerenzer, Hoffrage, & Kleinbölting, 1991). The PMM theory models the cognitive processes in tasks in which a choice is made between two objects in terms of a quantitative criterion, and a confidence judgment is made that the chosen object is correct. Here, we apply the PMM framework to a context in which a previous response (i.e., choice and confidence judgment) needs to be reconstructed after receiving feedback on whether the choice was correct. We refer to this model as the RAFT model, where RAFT stands for Recostruction After Feedback With Take The Best (Take The Best is a simple inferential heuristic that is described in the next section). The RAFT model makes three general assumptions. First, if (and only if) the original response cannot be retrieved from memory, it will be reconstructed by rejudging the problem. Second, the rejudgment involves a recall of the cues and cue values underlying the original choice. Third, knowledge, in particular uncertain knowledge, is automatically updated by feedback. According to the RAFT model, feedback does not directly affect the memory trace for the original response but indirectly by changing (i.e., updating) the knowledge that is used as input for the reconstruction process. Although the process of knowledge updating is adaptive because it enables individuals to improve their inferences over time, it has a by-product: the hindsight bias. We now specify the cognitive processes underlying the responses at Time 1 and Time 3.

¹ The simplest response strategy is to recall the old belief. It has been argued that outcome information could either destroy or disturb the memory trace of the original judgment (Fischhoff, 1975) or reduce its accessibility (Hell et al., 1988). This alone, however, cannot explain the occurrence of hindsight bias. The second strategy, anchoring and adjustment, would result in hindsight bias if the adjustment was not large enough. Although there is a large body of evidence suggesting that adjustment is usually insufficient (Tversky & Kahneman, 1974), Hawkins and Hastie (1990) and Hertwig et al. (1997) also pointed out problems with this explanation. For example, this explanation would lead us to predict the same effect for occurrences (anchor on 100%) and nonoccurrences (anchor on 0%), whereas occurrences actually do lead to larger effects. According to the motivational response-adjustment explanation, the bias is attributed to participants' motivation to appear intelligent, knowledgeable, or perspicacious. The empirical evidence supporting this view is scattered and weak. In line with most of the literature, Hawkins and Hastie concluded that motivational response adjustment appears to play either a minor role or no role in explaining hindsight bias.

Time 1: Original Response

Patricia, who is a visiting researcher from California, is concerned about eating a healthy diet. However, she has a sweet tooth, and at a restaurant she wants to order dessert. The menu provides her with the choice between chocolate fudge cake and pumpkin custard pie (Time 1). Because she wants to reduce her cholesterol consumption, she asks herself which of the two has more cholesterol (to choose the one having less). Not knowing the answer, she tries to infer it from what she knows about the two foods. According to PMM theory, she will construct a probabilistic mental model (PMM) to make this inference. Such a PMM consists of a reference class of foods, cues for cholesterol, and an inferential heuristic, as described below.

Knowledge about cues. In Patricia's case, the reference class might be foods in her local supermarket. According to PMM theory, knowledge about the objects in a reference class consists of probability cues and the values the objects have with respect to these cues. For example, saturated fat is such a cue: If one food item has more saturated fat than the other, then the one with more fat is also likely to have more cholesterol.

It is useful to think of knowledge stored in long-term memory as a matrix of Objects (e.g., food items) \times Cues (e.g., saturated fat), in which one can search for information. Whereas all the examples in Gigerenzer et al. (1991) and Gigerenzer and Goldstein (1996) involve binary cues, we extend PMM theory to continuous cues. For cues with continuous values, there are four possible relations among two objects with respect to any cue. The term *object relation* refers to the ordinal relation of objects with respect to a cue (rather than to the criterion). This relation can be larger (e.g., cake contains more saturated fat than pie), smaller, equal, or unknown (Table 1). The last is the case when entries are missing in the Object \times Cue matrix because of limited knowledge. The relations (i.e., $>$, $<$, $=$, $?$) can either be directly retrieved or deduced from knowledge about absolute cue values.

Table 1
Hindsight Bias at the Level of Confidence

Knowledge and choice	Time 1		Time 3
Saturated fat (80%)	cake ? pie	\rightarrow	cake $>$ pie
Calories (70%)	cake $>$ pie		cake $>$ pie
Protein (60%)	cake $>$ pie		cake $>$ pie
Choice Confidence	cake 70%		cake 80%

Note. The probabilistic mental model contains three cues ranked according to their validity (specified in parentheses). The symbols $>$ and $?$ denote the relations between objects on these cues. For example, in the Time 1 column, which describes the knowledge underlying the original response, the object relation on the saturated fat cue is unknown. As indicated by the arrow, this object relation changes after feedback that cake has more cholesterol than pie. The relation shifts toward feedback, that is, from $?$ to $>$ in the updated mental model (Time 3 column). As a consequence, hindsight bias occurs. Note that Take The Best searches only for the object relations that appear in boldface.

Some cues are better predictors than others. We call this predictive power *ecological validity*, which is defined as the relative frequency with which the cue correctly predicts which object scores higher on the criterion (in a defined reference class). Assume that Patricia's reference class consists of food items sampled from the local supermarket, with saturated fat, calories, and protein as quantitative cues for cholesterol. When we conducted a random sample in a Chicago supermarket, we found that in about 80% of the cases where one food item had more saturated fat (cue) than the other, it also had more cholesterol (criterion). This value is the ecological validity of the saturated fat cue. In this particular environment, saturated fat is the "best" cue, that is, it has the highest ecological validity. It is followed by the calorie (about 70%) and protein (about 60%) cues.

Inferential heuristic. How can Patricia infer whether cake has more cholesterol than pie? We account for her inference with a heuristic called "Take The Best," which is the core heuristic of the PMM framework (Gigerenzer & Goldstein, 1996). This lexicographic heuristic starts with a subjective rank order of cues according to their validities and makes the inference on the basis of the highest ranking (best) cue that discriminates. The steps of Take The Best (excluding the recognition heuristic, which is not relevant here) are:

1. **Search rule:** Choose the cue with the highest validity and retrieve the object's cue values from memory.
2. **Stopping rule:** If the best cue discriminates (object relation: $>$ or $<$), stop searching. If the best cue does not discriminate (object relation: $=$ or $?$), continue with the next best cue until a cue that discriminates is found.
3. **Decision rule:** Choose the object to which the cue points, that is, the object with the higher cue value (if criterion and cues are negatively correlated, then choose the object with the lower cue value). If no cue discriminates, then make a random choice (guess).
4. **Confidence rule:** Use the cue validity of the cue that discriminates as the confidence in the choice. If the choice was made at random, confidence is 50%.

A seemingly irrational feature of the heuristic is that it does not integrate all the available information but uses what we call *one-reason decision making*, where a decision (e.g., a choice) is based on only one cue. To illustrate this, Table 1 shows Take The Best applied to Patricia's knowledge. At Time 1, her choice is based solely on the calorie cue. Because the cake has more calories than the pie, the heuristic chooses the cake as the alternative with more cholesterol; confidence in the correctness of the decision is 70% (the validity of the calorie cue).² Take The Best is fast because it does not involve much computation, and it is frugal in the sense that it only searches for some of the available information.

The simplicity of Take The Best raises the suspicion that it might be highly inaccurate, compared with standard inferential algorithms that process and combine all available

² Note that the term *choice* has two meanings here. The first refers to an inference, for example, "which of two foods has more cholesterol"; this is how we use the term throughout this article. The second meaning relates to an actual selection, for example, Patricia's order, at the restaurant, of the food with less cholesterol.

predictors. Yet in 20 real-world environments, it was able to compete well with other, more complex algorithms, such as multiple regression (Czerlinski, Gigerenzer, & Goldstein, 1999) or Bayesian networks (Martignon & Laskey, 1999). Because Take The Best is not only accurate but also both fast and frugal, it is particularly suitable for situations in which time and/or knowledge is limited. This heuristic has successfully explained a number of phenomena in memory-based inference in a single framework (see Gigerenzer et al., 1991). We now show how the RAFT model applies Take The Best to a situation in which a past choice and a confidence judgment have to be reconstructed.

Times 2 and 3: Feedback and Reconstruction

Some weeks after having dinner at the restaurant, Patricia remembers her dessert dilemma and decides to check the nutrition labels at her local supermarket. She finds out that chocolate cake has more cholesterol than pumpkin pie (Time 2) and then asks herself what she actually chose at the restaurant (Time 3).

How is the original response recalled? The cognitive processes as assumed by the RAFT model can be seen in Figure 1. First, an attempt is made to access the original response directly from memory. The chance of doing this successfully depends on factors such as length of time between original and recalled response (e.g., Fischhoff & Beyth, 1975) and depth of encoding of the original response (Hell, Gigerenzer, Gauggel, Mall, & Müller, 1988).

If the original response cannot be retrieved, it will be reconstructed by repeating the steps taken at Time 1. A simple analogy may help to motivate this assumption: Imagine you are asked to multiply a two-digit by a three-digit number. A couple of days later you are asked to remember your result. If you cannot retrieve it from memory, you can compensate for this by performing the same calculation again; that is, lack of recall can be compensated for by recalculation. The same holds for a choice that has been made under uncertainty. To compensate for not being able to recall the choice, the probabilistic mental model used at Time 1 can be reconstructed at Time 3.

This process begins by retrieving the knowledge on which the choice at Time 1 was based, that is, by retrieving the original cues (in the original order) and the knowledge about those cues. In some cases, veridical retrieval may be possible; in others, memory of the cue values (object relations) may be vague or absent. RAFT's critical assumption is that feedback transforms some of the elusive relations into discriminating relations. This is due to the reversibility of the cue-criterion relationship: Because it is possible to draw inferences from a cue (e.g., saturated fat) to the criterion, the reverse is also possible—to draw inferences from the criterion to the cues. In other words, what used to be the distal variable (i.e., cholesterol) at Time 1 now turns into a proximal cue. This new proximal cue is used to infer what used to be a proximal cue at Time 1 (e.g., saturated fat) and what turns into a distal variable at Time 3, when an attempt is made to reconstruct the original PMM. Such a reversal between proximal cues and distal variables is

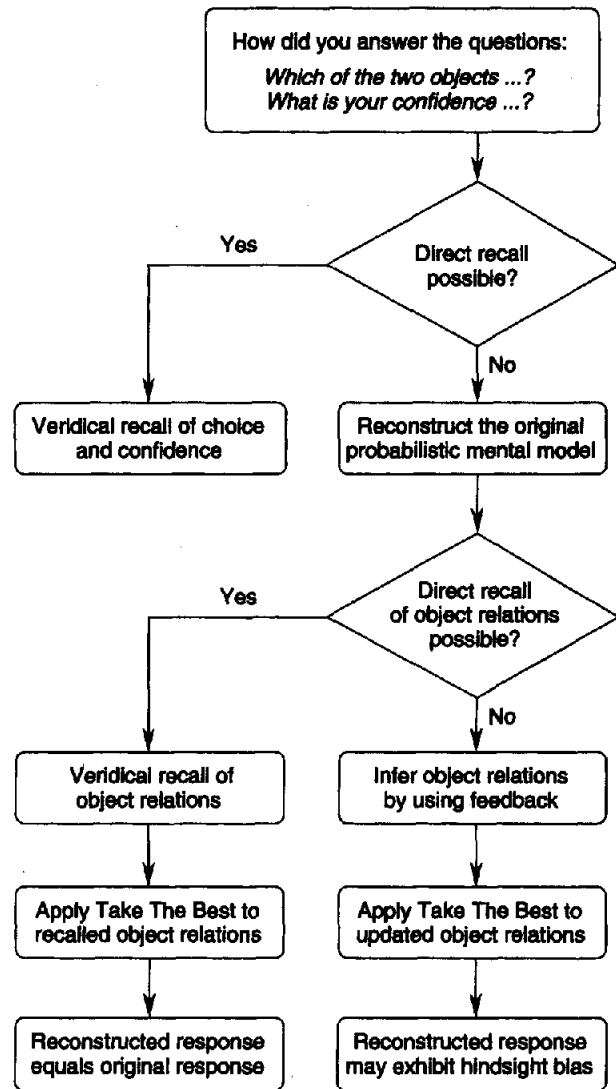


Figure 1. Cognitive processes at Time 3. The task is to remember the original response (choice and confidence) made at Time 1.

possible because cues and criterion are correlated with each other.

We assume that the process of updating knowledge is not restricted to reconstructions made in hindsight. Rather, we think of this updating as a general and continuous process. When new knowledge is acquired, it does not remain isolated but is automatically integrated into existing knowledge, which might involve adapting this new knowledge to existing knowledge or vice versa. The RAFT model stresses an assimilation of old knowledge to new available knowledge. This updating of old knowledge serves an adaptive function: It results not only in a more coherent corpus of knowledge but, if the new knowledge is valid, a more accurate one, as well.

Illustrations of the RAFT Model

Consider Patricia's dilemma again. Her question was, which of the two food items, the cake or the pie, has more

cholesterol. Saturated fat, calories, and protein were used as cues to infer the correct answer. As illustrated in Table 1, the most valid cue (saturated fat) did not discriminate at Time 1. Then at Time 2, she found out that the cake has more cholesterol than the pie. When she finally attempts, at Time 3, to reconstruct her original response, RAFT assumes that the new knowledge concerning cholesterol may be used to infer her (previous) cue values. As a result, the cue values (and their relations) are not veridically remembered but show systematic shifts toward feedback. For instance, the saturated fat cue discriminates now and points to the cake (Table 1, Time 3). If the same heuristic (here, Take The Best) is then applied to the updated knowledge base, the resulting choices and confidences will show systematic shifts toward feedback. In the example given, Patricia infers at Time 3 that she chose the cake as the food with more cholesterol. She also infers that her confidence in this choice was 80% (the validity of the saturated fat cue). Thus, her reconstructed choice is identical to her original choice. However, her reconstructed confidence increased relative to her original confidence, thereby exhibiting hindsight bias. More generally, hindsight bias at the level of confidence occurs if recalled confidence increases after receiving feedback that the originally selected alternative was correct (or decreases after receiving feedback that it was wrong).

Not only confidence, but even choice may change from Time 1 to Time 3. Hindsight bias at the level of choice occurs if the original choice was wrong and the recalled choice was correct. The RAFT model can also explain hindsight bias at this level. For instance, in a variant of Table 1, neither the saturated fat nor the calorie cue but only the protein cue discriminated at Time 1 (Table 2). This cue points to the pie. At Time 3, the saturated fat cue discriminates, now pointing to the cake. The result is hindsight bias at the level of choice.

It is also conceivable that hindsight bias can be reversed. Reversed hindsight bias at the level of confidence occurs in cases where recalled confidence decreases although feedback indicates that the originally selected alternative was correct (e.g., original choice: cake has more cholesterol, 70%; feedback: cake; recalled choice: cake, 60%) or increases, although feedback indicates that the originally selected alternative was wrong. Reversed hindsight bias at the level of choice occurs when the original choice was

correct and the recalled choice was wrong (e.g., original choice: cake; feedback: cake; recalled choice: pie). RAFT accounts for reversed hindsight bias by allowing for random shifts in the reconstructed object relations; that is, in addition to systematic shifts that are due to feedback, RAFT posits unsystematic shifts that are due to the imperfect reliability of one's memory of knowledge. Such random shifts are independent of feedback; this means they can either be manifested as hindsight bias (if they coincide with the direction of feedback) or as reversed hindsight bias (if they are counter to the direction of feedback).

Unless otherwise specified, we use the terms hindsight bias and reversed hindsight bias to refer to item-specific differences in original and recalled responses rather than to aggregated responses. Because RAFT specifies the conditions under which hindsight bias and reversed hindsight bias occur on an item-specific level, this theoretical precision allows us to apply the established terms to effects observed for individual items.

Predictions

Prediction 1 (Asymmetry in Shifts)

If feedback on the criterion is provided, then the object relations will shift asymmetrically, more often toward the correct alternative than away from it. If no feedback is provided, then both kinds of shift should be about equally prevalent.

This prediction is derived as follows. Feedback updates elusive or missing object relations. If, according to feedback, Object A scores higher (lower) on the criterion than Object B, it may be inferred that Object A probably also scores higher (lower) on the cue. In addition, there are random shifts, which will occur equally often toward and away from feedback. Across systematic and random shifts, more relations will change toward feedback than away from it.

Updating after feedback will be more likely when a cue did not discriminate at Time 1, compared with cases where it did discriminate. The rationale for this corollary of Prediction 1 is as follows. The fact that a cue discriminated at Time 1 indicates that some knowledge was available. The mere existence of such knowledge increases the likelihood that it can be accessed again at Time 3 and that the relation will be veridically retrieved even after feedback. What if, by contrast, a cue did not discriminate at Time 1, either because a cue value for one or both objects was unknown or because the cue values were equal? If the relation was unknown at Time 1, then feedback does not need to overcome preexisting knowledge to become manifest. A similar implication holds for equal relations. If both cue values in a pair of objects are equal at Time 1, a shift in one cue value is sufficient to change the relation. For a discriminating relation, in contrast, a shift in one cue value may reduce the difference between the two values but not necessarily cause a shift in the relation.

Table 2
Hindsight Bias at the Level of Choice

Knowledge and choice	Time 1		Time 3
Saturated fat (80%)	cake ? pie	→	cake < pie
Calories (70%)	cake = pie		cake = pie
Protein (60%)	cake > pie	→	cake ? pie
Choice	pie		cake
Confidence	60%		80%

Note. This table is a variant of Table 1, where the RAFT model predicts hindsight bias at the level of choice, not just confidence. RAFT = Reconstruction After Feedback With Take The Best.

Prediction 2 (Contingency of Hindsight Bias on Recalled Cue Values)

On the basis of a participant's recalled object relations for a particular item, RAFT is able to account for observed outcomes (hindsight bias, reversed hindsight bias, or veridical recall).

This prediction is derived as follows. In the RAFT model, feedback is not considered to have a direct impact on recalled choice and confidence but rather causes systematic shifts in the cue values (and, thus, in the object relations). These systematic shifts, in turn, can lead to biased recollections of choice and confidence: If the original response cannot directly be recalled from memory, it will be reconstructed by applying Take The Best to the (updated) object relations. By comparing the reconstructed response to the original response, RAFT is able to predict whether hindsight bias, reversed hindsight bias, or veridical "recall" occurs.

Prediction 3 (Reduction of Hindsight Bias)

If recall of the cue values is assisted, then hindsight bias will be reduced.

The rationale behind this prediction is as follows. Because hindsight bias is attributed to systematic changes between the cue values at Time 1 and at Time 3, experimental manipulations that reduce the likelihood of these changes should also reduce hindsight bias. This likelihood can be reduced by assisting the recall of cue values. In the studies reported herein, we used three ways of assisting the recall: (a) Participants' memory of cue values at Time 3 was refreshed by repeating the learning phase in which they were taught these cue values (Study 1); (b) the retention interval between Time 1 and 3 responses was shortened (Study 1); and (c) the cue values as recalled after the learning phase were presented to the participants before asking them to recall their original response (Study 2).

Note that to the extent that other than the modeled reconstruction process underlies hindsight bias (Erdfelder & Buchner, 1998; Hawkins & Hastie, 1990), assisting the recall of cue values should reduce but not eliminate hindsight bias. By providing one mechanism (in our view, a crucial one), RAFT does not invalidate other processes, such as metacognition or motivational response adjustments. There is one study that allowed us to derive a rough estimate of the size of reduction in hindsight bias that was due to assisting recall. In a hypothetical design, Davies (1987, Experiment 1) asked participants to read descriptions of four psychological experiments and to write down comments on the clarity of the instructions, appropriateness of the methods, and reasons why the experiment may turn out one way or the other. Two weeks later, they had to judge the likelihood of various experimental outcomes. Before participants made these judgments, researchers told one group the actual outcomes and asked them to ignore them; a control group did not receive this outcome information. In addition, half of the participants in each of these conditions were given the notes they had made in the first session. Among those participants who did not have the opportunity to look

at their own notes, the mean likelihood ratings in the reported outcomes were 15.6 percentage points higher than those made by participants who had no outcome information, demonstrating the knew-it-all-along effect. Moreover, consistent with Prediction 3 of the RAFT model, among the participants who were shown their notes, this difference decreased to 7.5 percentage points; that is, the effect had been reduced by about half.

We conducted two studies. Study 1 was designed to test Predictions 1 through 3 for a two-alternative choice and confidence task with quantitative cues. Study 2 was designed to replicate the tests of Predictions 1 through 3 with binary cues.

Study 1

Method

Participants. Eighty students from the University of Chicago took part in the experiment. They were paid volunteers, recruited by advertisement from a broad spectrum of disciplines, and tested in groups of up to four people.

Design and procedure. A topic of significance for many people is nutrition: In the United States, cholesterol, in particular, has become a major concern. To provide participants with a context for the present study, we informed them of the physiological mechanism that explains why cholesterol is one of the main risk factors for heart disease. We then informed them that cholesterol tends to covary with three substances—saturated fat, calories, and protein—and that the amount of cholesterol can be inferred from the amounts of these substances.

Despite the potential significance of nutrition, most people do not have much specific knowledge about it. Therefore, the experiment started with a learning phase in which the participants learned the actual saturated fat, calorie, and protein values of 36 food items. They were instructed to read over the list several times and to learn the objects and cue values by heart. They were informed that this information would be instrumental in solving the task that followed. After each of three learning trials (10 min per trial), we checked whether the participants had actually acquired the information (5 min per test trial).

Immediately after the learning phase, participants were given a list of 18 food pairs (constructed from the pool of 36 items) and asked two questions about each pair: "Which food do you think has more cholesterol?" and "How confident are you that your choice is correct?" If the participants were absolutely certain, they were instructed to give 100% as their confidence. If their answer was simply a guess, they should give 50% as their confidence. In all other cases, they were asked to provide values between 50% and 100%, in 10-point increments. After they had given their choice and confidence rating, we instructed the participants to recall the amounts of saturated fat, calories, and protein they had learned for each food item in the learning phase or to indicate for each food pair the object relations with respect to each cue (knowledge before feedback).

After 1 day, 40 participants—and after 1 week, the other 40—attended the second session and were randomly assigned to one of three conditions. In the feedback condition, participants ($n = 40$; 20 after 1 day, 20 after 1 week) first received feedback for each of the 18 questions they had answered previously (e.g., "The cholesterol values for chocolate fudge cake and pumpkin custard pie are 44 mg and 31 mg per 100 g, respectively"). To ensure that they paid attention to the feedback given, the participants had to enter the cholesterol values for each food pair in a graph. Then they

were asked to recall which food they had originally chosen as having more cholesterol and how confident they were that their choice was correct. Afterwards, in a new questionnaire, they were asked to recall the saturated fat, calorie, and protein values they had learned in the learning phase (knowledge after feedback). The recall of cue values was necessary to test Predictions 1 and 2. In the no-feedback condition, the procedure and tasks were identical to those in the feedback condition except that participants ($n = 20$; 10 after 1 day, 10 after 1 week) received no feedback. In the relearning condition, the procedure and tasks were identical to those in the feedback condition except that before receiving feedback, participants ($n = 20$) refreshed their memory of the cue values by studying the information they had originally learned for another 10 min (followed by a test of whether they had acquired the information).

Materials. We used a set of 36 food items selected from a supermarket near the University of Chicago. In the learning phase, participants learned 62 of the 108 cue values (36 food items \times 3 cues); the remaining 46 cue values were not specified. The cue validities that participants were taught (80%, 70%, and 60% for the saturated fat, calorie, and protein cues, respectively) corresponded closely to the actual validities (83%, 69%, and 62%) in the chosen set of food items. From the 36 food items, we constructed 18 pairs: In 9 pairs, the most valid cue that discriminated (on the basis of the information received in the learning phase) was saturated fat; in 6 pairs, it was the calorie cue; and in 3 pairs, it was the protein cue.

To control for possible sequencing effects, we used a different random order of food pairs in each session. In addition, for both sessions, we randomly determined the positions of food items for each food pair (i.e., which of the two objects was presented on the left and which on the right).

Results

Did we obtain aggregated hindsight bias? Following Winman, Juslin, and Björkman (1998), we mapped original and recalled confidence judgments to a full-range confidence scale, thereby recoding the confidence judgments for those food pairs where a wrong choice was made (e.g., a confidence judgment of 70% that the wrong alternative was the correct one was recoded as 30%). This way all confidence judgments were conditioned on the correct alternative, and hindsight bias should become manifest in an increase of confidence. We first computed the difference between the original and the recalled confidences for each participant separately by averaging across items. Figure 2 illustrates these differences, averaged across participants. Confidence increased in the feedback condition by an average of 3.7 percentage points ($n = 39$, $SD = 9.4$, $SE = 1.5$), whereas in the no-feedback condition it decreased by 1.1 percentage points ($n = 19$, $SD = 5.4$, $SE = 1.2$).³ The effect size for the difference ($\Delta = 4.8$), $t(56) = 2.04$, $p = .023$, was $d = 0.56$ (see Cohen, 1988, p. 20, Formula 2.2.1). According to Cohen's (1988) classification, this hindsight bias is a medium effect and, thus, larger than the average effect size reported in Christensen-Szalanski and Fobian Willham's (1991) meta-analysis.

We also determined the percentage of cases in which either the recalled choice or the recalled confidence differed from the original choice or confidence. We combined the two response modes (choice and confidence) as follows. First, we compared the original and recalled choices for each

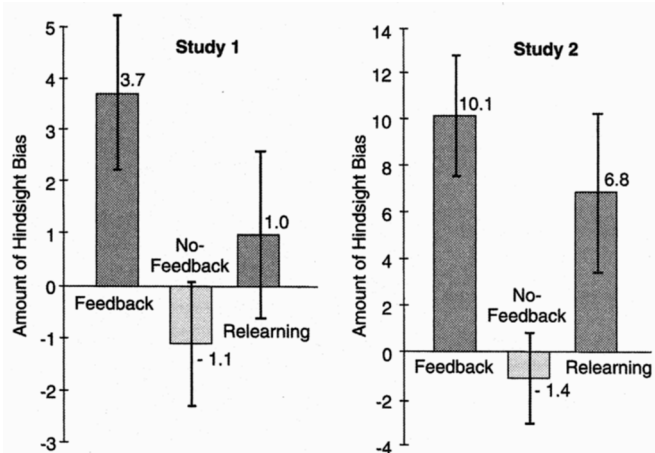


Figure 2. Amount of hindsight bias: (reflected) original confidence judgments minus (reflected) recalled confidence judgments. A positive difference indicates hindsight bias, and a negative difference indicates reversed hindsight bias. The bars denote standard errors.

item. If they were different, we classified this item as showing either hindsight bias or reversed hindsight bias. If they were identical, we compared the original and recalled confidences and classified this item as showing hindsight bias, reversed hindsight bias, or veridical recollection. When averaged across all participants in the feedback condition, cases showing hindsight bias exceeded cases of reversed hindsight bias by 9.4 percentage points (34.5% vs. 25.1%). In the no-feedback condition, this difference was -7.9 percentage points (29.7% vs. 37.6%). If we considered only choices, the corresponding differences were 2.6 percentage points (11.0% vs. 8.4%) and -2.1 percentage points (8.0% vs. 10.1%) for the feedback and the no-feedback conditions, respectively.

The proportion of veridical recalled choices was almost identical in the feedback (80.6%) and in the no-feedback conditions (81.9%). Moreover, the proportion of cases in which both choice and confidence were veridically recalled was even slightly higher in the feedback than in the no-feedback condition. These findings are consistent with the biased reconstruction hypothesis (Stahlberg & Maass, 1998) but cannot be accounted for by the memory impairment hypothesis, which assumes that feedback changes existing memory traces. Likewise, Dehn and Erdfelder (1998), who used a multinomial model approach, concluded that they "failed to find any evidence for memory impairment hypotheses of the hindsight bias . . . in none of our

³ Two participants were excluded from the analysis because they apparently misunderstood the instructions: One was excluded in the feedback condition, because in the second session (after 1 day) he recalled exclusively 0% and 100% confidence judgments, whereas his original confidences were distributed across all confidence categories; the other was excluded in the no-feedback condition (after 1 week), because almost half of her confidence judgments were below 50%.

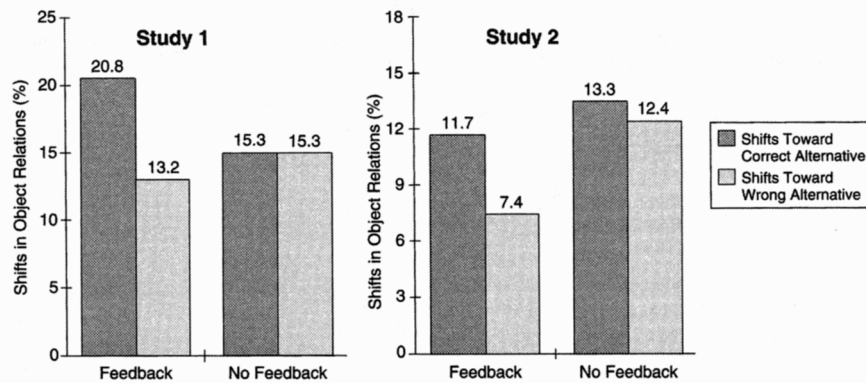


Figure 3. Percentages of shifts of object relations toward and away from the correct alternative in the feedback and no-feedback conditions. Veridical recollections are not included.

experimental conditions does the probability of recollecting the original answer depend on whether feedback information is provided or not" (p. 144).

The fact that hindsight bias was reversed in the no-feedback condition can be attributed to a base rate effect:⁴ Across all items and participants in the feedback (no-feedback) condition at Time 1, 67.3% (67.0%) of all the choices were correct. Thus, in about two thirds of the cases, feedback was supportive (i.e., initial choice was *a*, feedback was *a*) and, for those cases, only an identical recollection (recalled choice is *a*) or reversed hindsight bias (recalled choice is *b*) on the level of choice could occur. Accordingly, for the remaining 32.7% (33.0%) of cases where the initial choice was wrong, the only possible outcomes are a veridical recollection or hindsight bias. Thus, random guessing at Time 3 would lead not only to 50% veridical recollections but also to twice as many cases of reversed hindsight bias as cases of hindsight bias. Because the same percentage of cases was also correct at Time 1 in the feedback condition, such a base rate effect could be expected there as well. Thus, hindsight bias was not favored, but it nevertheless occurred (and it was even larger than the average bias observed by Christensen-Szalanski and Fobian Willham, 1991). In fact, 34% of the cases where feedback contradicted the original choice resulted in hindsight bias, whereas only 12% of the cases where feedback supported the original choice showed reversed hindsight bias. In the no-feedback condition, the corresponding percentages were 24% and 18%, respectively. We next turn to Prediction 1.

Is there an asymmetric shift in object relations? To reiterate, Prediction 1 states that if feedback on the criterion is provided, then the object relations will shift asymmetrically, more often toward the correct alternative than away from it. (If no feedback is provided, then both kinds of shift should be about equally prevalent.) For each item and cue, we determined whether the recollection of object relations was veridical or whether a shift toward or away from the correct alternative occurred. A shift toward the correct alternative included (a) relations that pointed to the smaller object at Time 1 but were recalled either as unknown or even reversed at Time 3 and (b) relations that were unknown at

Time 1 but pointed to the larger object at Time 3. Shifts away from the correct alternative included all cases with shifts in the opposite direction, that is, where "smaller" and "larger" in (a) and (b) were exchanged. An object relation was classified as unknown if a participant did not specify the relation by entering either a relation symbol or the values for the two objects.

As can be seen in Figure 3, in the feedback condition, 20.8% of the cases shifted toward the correct alternative, and 13.2% shifted away from it ($n = 431$ and $n = 273$ of 2,076, respectively). In the no-feedback condition, the two kinds of shift occurred equally often (15.3%, or $n = 150$ of 981 for both kinds of shift), as predicted. The percentages of veridical recollections of the object relations were 66.1% (1,372 of 2,076) for the feedback condition and 69.4% (681 of 981) for the no-feedback condition.

A corollary of Prediction 1 is that cue validities that are based on participants' recollections of cue values after feedback should be higher than those before it. Cue validity is defined by the proportion of correct inferences that are based on a cue. If object relations systematically shift toward feedback, the proportion of correct inferences should increase. As depicted in Figure 4, across all cues and recalled cue values, the average validity indeed increased by 7.2 percentage points (from 58.3% to 65.5%) in the feedback condition, but only by 0.4 percentage points (from 59.2% to 59.6%) in the no-feedback condition.

Is the impact of feedback greater when a cue does not discriminate at Time 1 (either because a cue value for one or both objects is unknown, or because the values are equal)? In 710 of 2,076 responses in the feedback condition, a cue did not discriminate at Time 1. After feedback, 38.7% of these cases shifted to discriminating object relations, with 27.7% now pointing to the larger object and 11% to the smaller one ($\Delta = 16.7\%$). If, however, a cue did discrimi-

⁴ It is noteworthy that with numerical judgment tasks (e.g., "How high is the Eiffel Tower in Paris?") in the no-feedback conditions, the opposite result, namely hindsight bias, is usually obtained. This outcome has been explained as a regression-toward-the-mean phenomenon (see Erdfelder & Buchner, 1998, footnote 7).

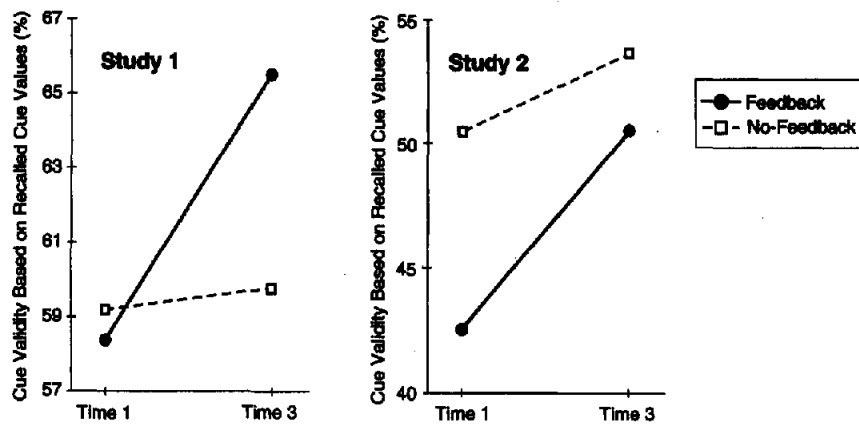


Figure 4. Cue validities based on participants' recollection of cue values before and after feedback (averaged across all cues and all recalled object relations).

nate at Time 1, feedback had almost no impact: Here 32.4% of the object relations shifted (429 of 1,366), but shifts were almost symmetrical ($\Delta = 2.8\%$). Consistent with Prediction 1, in the no-feedback condition, there was symmetry, both when cues did not discriminate at Time 1 ($\Delta = 1.7\%$) and when they did ($\Delta = -1.5\%$).

To summarize, consistent with the RAFT mechanism of updating knowledge after feedback, more object relations shifted toward feedback than away from it. This difference was most pronounced when the original relations were nondiscriminating.

Is hindsight bias contingent on recalled object relations?

To reiterate, Prediction 2 states that, on the basis of a participant's recalled object relations for a particular item, RAFT is able to account for observed outcomes (hindsight bias, reversed hindsight bias, or veridical recall). We tested this prediction at the level of choices only and at the level of choice and confidence combined.

For the test on choices, we determined, for each participant and food pair, the observed outcome at the level of choice (hindsight bias, reversed hindsight bias, or veridical recall). Next, we applied Take The Best to the updated object relations at Time 3 and compared the choice (inferred by Take The Best) with the original choice given at Time 1. This comparison determined whether RAFT would lead us to predict hindsight bias, reversed hindsight bias, or no hindsight bias (if Take The Best was forced to guess, RAFT's possible predictions were treated as equally likely). If one defines hindsight bias as the percentage of choices exhibiting hindsight bias minus the percentage of choices exhibiting reversed hindsight bias across all responses, the observed difference in the resulting amount of hindsight bias between the feedback and the no-feedback condition was 4.7 percentage points. The predicted difference was 6.7 percentage points.

Thus, the predicted hindsight bias was of the same magnitude. However, these numbers do not provide a strict test of Prediction 2, because this prediction relates to the match between the observed and the predicted outcome at the level of individual responses. To determine this match,

RAFT's predicted outcome was compared with the observed outcome, and, for each participant, we determined the percentage of correct predictions across all items. Across all participants in the feedback and no-feedback conditions, the averaged percentage of correct predictions was 83.5% (see Figure 5).

By what benchmark can this value be measured? Comparing the performance of the RAFT model with chance is especially important, because the two outcomes—observed and predicted—had the same reference point, namely, the original response. Thus, they are related, and chance performance might well be above 50%. Across all participants in the feedback and the no-feedback conditions, RAFT's performance was 26.7 percentage points better than chance performance (56.8%, see Figure 5), $t(57) = 15.4$, $p = .001$.⁵

At Time 3, about 80% of the choice coincided with those of Time 1. Some of this high percentage of identical choices may be caused by direct recall rather than reconstruction (in which case RAFT is not applicable). For this reason, we conducted another test. In this test, we used a very strict operationalization of direct recall and excluded all cases in which the original and recalled choices were identical. Here

⁵ Chance outcome was generated by using participants' actual knowledge distributions, rather than by assuming ignorance about cue values or a uniform probability distribution of cue values. To derive the prediction of chance for a specific item, we predicted, by using Take The Best, the outcome for this item on the basis of knowledge about the cue values for another item. Thus, for each participant in Study 1 and Study 2, we arrived at 17 and 5 predictions for each item, respectively. Then, we compared the original response for a specific item with each of the predicted responses and determined the percentages of correct predictions (first, within each participant and across all items and predictions, then averaged across participants). As can be seen in Figures 5 and 6, this procedure provided a benchmark that was much higher than the simple and unwarranted assumption that chance performance would be 50% (such a simple chance model ignores that, because of scale-end-effects, the probability of a match between observed and predicted outcome is larger than 50% for original responses with extreme confidence).

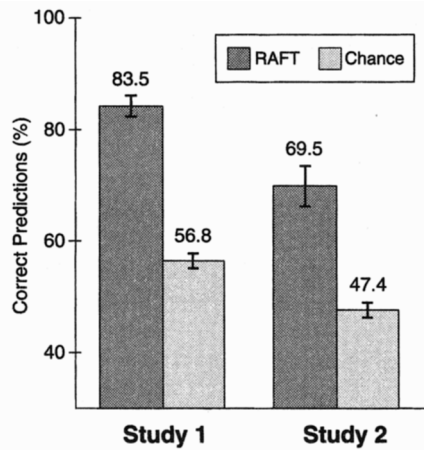


Figure 5. Match between observed outcomes at the level of choice (hindsight bias, reversed hindsight bias, or veridical recall) and outcomes as predicted by the RAFT model and by a chance model, respectively. Cases in which original choice and recalled choice were identical are included. The bars denote standard errors. RAFT = Reconstruction After Feedback With Take The Best.

RAFT's performance was still 19.1 percentage points better than that of chance (72.9% vs. 53.9%), $t(53) = 4.6$, $p = .001$.

How good is RAFT's predictive performance on the level of choices combined with confidence? Again, we excluded cases of direct recall (this time, those relatively rare cases in which both choice and confidence at Time 1 and 3 were identical). As Figure 6 shows, RAFT correctly predicted 76.3% of the observed outcomes. In contrast, the performance of the chance model (67.9%) was 8.4 percentage points worse, $t(57) = 5.0$, $p = .001$. The fact that RAFT fared better with predicting hindsight bias on the level of choices rather than hindsight bias on the level of choice and confidence is consistent with the common observation that it is more difficult to model confidences than choices (e.g., Hoffrage, 1995).

Could hindsight bias be reduced? To reiterate, Prediction 3 states that if recall of the cue values is assisted, then hindsight bias will be reduced. In the relearning condition, we tried to assist recall of the cue values by repeating the learning phase before giving feedback. The average increase in (reflected) original to recalled confidence in this condition was 1.0% ($n = 20$, $SD = 7.3$, $SE = 1.63$; see Figure 2), which is 2.7 percentage points less than in the standard feedback condition (3.7%) at 100% and in the no-feedback condition (-1.1%) at 0%, then hindsight bias in the relearning condition amounted to 56%. In other words, relearning the cue values before giving feedback reduced the difference between the feedback and no-feedback conditions by about half (44%), which is consistent with Davies's (1987) finding.

We also manipulated the recall of the cue values indirectly by comparing a 1-day ("short") and 1-week ("long") interval between the original and recalled responses. On the basis of the plausible assumption (independent of the RAFT model) that memory traces become less accessible over

time, we expected less hindsight bias for the short interval. The results were mixed: Consistent with this assumption, we found more veridical recollections of both choice and confidence after the short rather than the long interval (feedback condition: 46.2% and 34.7% for the short and long intervals, respectively; no-feedback condition: 33.3% and 32.0%). The same was true for choice only (feedback condition: 83.4% and 78.0%; no-feedback condition: 82.8% and 81.0%). Inconsistent with the assumption, hindsight bias was larger after the short interval than the long interval. Confidence in the correct alternative increased in the feedback condition by an average of 5.6 and 1.8 percentage points, whereas in the no-feedback condition, it decreased by 1.2 and 0.9 percentage points for the short and long intervals, respectively.

We can only offer a partial explanation for the unexpected result of the larger hindsight bias after the short rather than the long interval. Participants in the feedback condition, who came to their second session after 1 day, had in their first session 3.3 percentage points fewer correct choices than those after the 1-week condition (65.7% versus 69%). Because of fewer correct choices at Time 1, hindsight bias had a better chance to occur in the 1-day condition.

Study 2

RAFT can be applied to continuous and binary cues. In Study 1, we obtained evidence that the model performed well with continuous cues. In Study 2, we tested its performance with binary cues, using material unknown to our participants, which had the advantage of giving us better control over the participants' knowledge of cues. In Study 1, the participants might already have had some knowledge about the criterion (i.e., cholesterol) or might have used other information than the three cues we taught them. Another difference between the studies was the manipula-

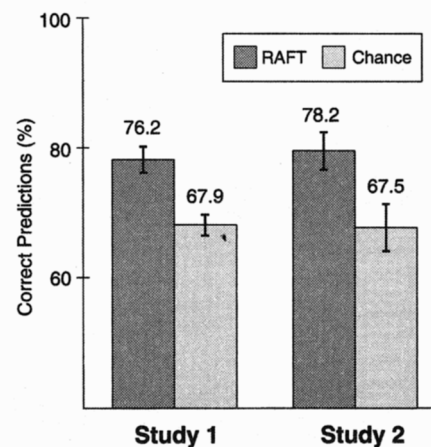


Figure 6. Match between observed outcomes at the level of choice and confidence (hindsight bias, reversed hindsight bias, or veridical recall) and outcomes as predicted by the RAFT model and by a chance model, respectively. Cases in which both choice and confidence were identical at Time 1 and Time 3 are excluded. The bars denote standard errors. RAFT = Reconstruction After Feedback With Take The Best.

tion chosen to test Prediction 3. In Study 2, we adopted a method used by Davies (1987), who provided participants at the time of recall with the notes they had made in arriving at their original responses. Similarly, during the recollection phase (Time 3), we presented the participants with the cue values indicated by them at Time 1. Aside from these variations, in Study 2 we attempted to replicate the results obtained in Study 1.

Method

Participants. Fifty-five participants from the University of Salzburg (most of them psychology students) were paid for taking part in this experiment. They were divided into small groups, with a maximum of 5 members and a mean number of 2.3 members.

Design and procedure. The participants were asked to put themselves into the role of a health insurance company employee. The first task consisted of learning some facts about 12 fictional individuals: whether they "have (or had) parents with hypertension," "are overweight," and "are smokers." This learning phase lasted 18 min. The participants were then told that these people had submitted applications to purchase health insurance. We explained that the cost of health insurance depends on certain criteria, including risk factors such as high blood pressure, and that these 12 applicants for health insurance had not yet indicated their values for blood pressure.

For the following choice task, the applicants were paired (six pairs), and the participants were asked to decide for each pair, "Which of these two applicants has higher blood pressure?" and to express their confidence in having answered correctly on a scale of 50% to 100%, in 10-point increments. We told the participants that the three variables—parents' hypertension, overweight, and smoking—were cues for high blood pressure. Then, after explaining the concept of cue validity to our participants, they learned that the validities were as follows: 80% for parents with hypertension, 70% for overweight, and 60% for smokers.

We tested participants' recall of the cue values in the following way. They were provided with four categories: "+" (cue value is positive), "-" (cue value is negative), "0" (no information was given about that person and that cue), and "?" (I have forgotten whether there was any information, or I have forgotten the information that was given about this applicant). Participants then had to (a) state a choice and a confidence and (b) recall the cue values. The sequence of tasks (a) and (b) was varied; one third of the participants performed all the pair comparisons first for (a), another third started with (b), and the last third had to compare each pair for both tasks at the same time. This first session lasted about 1 hr altogether.

One week later, the participants were presented with the same six pairs of applicants. As in Study 1, the participants had to carry out the tasks under one of three conditions: (a) the feedback condition, where the participants were told which of the applicants had higher blood pressure; (b) the no-feedback condition; or (c) the relearning condition, where the participants not only received feedback but could also refresh their memories of their original knowledge base. Unlike in Study 1, we did not repeat the learning phase but showed each participant how he or she had previously (i.e., in the first session) recalled the cue values. The participants' task was to recall their original responses, as well as the cue values they had given in Session 1. (The participants in the relearning condition did not have to recall the cue values.)

Materials. The names of the 12 applicants (6 women, 6 men) were randomly drawn from the local telephone book. In the learning phase, the participants received information about the

applicants on the three cues (i.e., parents with hypertension, overweight, smoking). We provided 24 pieces of information; for three of the applicants we provided information on all three variables, for six on two variables, and for the last three on only one variable. For each of the six single-sex pairs, only one of the three cues discriminated—each cue discriminated twice, for one item followed by feedback that could have been predicted from the cue values, and for the other item with surprising feedback. For the second session, the left-right position of half the applicant pairs was reversed.

Results

Did we obtain aggregated hindsight bias? Confidence increased after feedback by an average of 10.1 percentage points ($n = 18$, $SD = 10.9$, $SE = 2.6$); in the no-feedback condition, confidence decreased by 1.4 percentage points ($n = 19$, $SD = 8.5$, $SE = 1.9$); the effect size for the difference was $d = 1.15$; see also Figure 2). Across all responses in the feedback condition, the difference between the percentage of cases with hindsight bias and reversed hindsight bias was 26.3 percentage points (48.5% vs. 22.3%; the remaining 29.1% were veridical recollections of both choice and confidence). For the no-feedback condition, the difference was -5.3 percentage points (34.5% vs. 39.8%). For choice only, the difference was 10.7 percentage points (19.4% vs. 8.7%) for the feedback condition and -2.7 percentage points (10.6% vs. 13.3%) for the no-feedback condition. Thus, in the feedback condition, cases of hindsight bias outnumbered cases of reversed hindsight bias, whereas in the no-feedback condition, we found the same small incidences of reversed hindsight bias as in Study 1. Study 2 also replicated the finding that only the difference between cases of hindsight bias and reversed hindsight bias was effected by feedback, whereas the proportion of cases of veridical recollections did not systematically differ (both choice and confidence: 29.1% and 25.7%; choice only: 71.9% and 76.1%, for feedback and no-feedback conditions, respectively).

Prediction 1: Is there an asymmetric shift in object relations? As shown in Figure 3, in the feedback condition, 11.7% of the object relations shifted toward the correct alternative and 7.4% shifted away from it. (The remaining 80.9% were veridically recalled.) In the no-feedback condition, the corresponding percentages were 13.3, 12.4, and 74.3, respectively. The difference between the two conditions (4.3% vs. 0.9%) is not as large as in Study 1 but, again, points in the predicted direction.

Prediction 2: Is hindsight bias contingent on recalled object relations? As in Study 1, for each participant, we used the cue values recalled at Time 3 to predict choice and confidence for each item (a cue discriminated either if the value for one applicant was + and for the other it was -, or if it was + or - for one applicant and unknown for the other). The observed and predicted differences in hindsight bias—the difference between the feedback and the no-feedback conditions with respect to cases of hindsight bias minus cases of reversed hindsight bias—were 13.3 and 13.5 percentage points, respectively.

As argued earlier, the more interesting test is the match between observed and predicted outcomes on the level of

individual responses as shown in Figure 5. RAFT's performance on the level of choice was again much higher than chance (69.5% vs. 47.4%), $t(36) = 5.3$, $p = .001$. As in Study 1, we tested RAFT's performance under more difficult conditions, that is, by also taking confidences into account and by excluding all cases in which the original and recalled response (both choice and confidence) were identical. Still, RAFT's performance (78.2%) was 10.7 percentage points better than that of the corresponding chance model (see Figure 6), $t(36) = 2.9$, $p = .003$.

Prediction 3: Could hindsight bias be reduced? In Study 1, we assisted the recall of cue values by repeating the learning phase in the relearning condition. In Study 2, we presented each participant with the cue values that he or she had indicated in the first session. When we compared the average original and recalled confidences in this condition, we obtained an increase of 6.8 percentage points ($n = 18$, $SD = 15.0$, $SE = 3.5$; Figure 2). As in Study 1, the extent of hindsight bias in Study 2 was between that of the no-feedback condition (-1.4%) and the feedback condition (10.1%). If we give the no-feedback (feedback) conditions values of 0 (100) percent, then hindsight bias is reduced to 59%. This reduction is comparable with the one Davies (1987) observed in his Experiment 1 (48%).

Summary of Studies 1 and 2

RAFT specifies cognitive processes underlying hindsight bias: If people fail to recall their original response, this response will be reconstructed. Consistent with Prediction 1, we found in both studies that feedback on the criterion systematically influenced participants' recollections of their knowledge about cues. Updating of cue values toward feedback is an adaptive process and can lead to hindsight bias as a by-product. In fact, we were able to replicate the hindsight bias in both studies. Moreover, RAFT can explain why hindsight bias occurs, does not occur, or is reversed for individual responses. Consistent with Prediction 2, about 76% (Study 1) and 78% (Study 2) of all the cases in which either hindsight bias or reversed hindsight bias occurred were accurately predicted by RAFT. Consistent with Prediction 3, supporting the process of reconstruction by assisting the recall of cue values did reduce hindsight bias (as measured against the no-feedback condition)—in Study 1, during which the learning phase at the beginning of Session 2 was repeated, by 44% and in Study 2, during which each participant was presented with the cue values that he or she had indicated in the first session, by 41%. This reduction is particularly noteworthy when compared with the various attempts to reduce hindsight bias. In his review of debiasing strategies, Fischhoff (1982) concluded that "few of these techniques have successfully reduced the hindsight bias; none has eliminated it" (p. 428). The RAFT model provides a straightforward way to reduce hindsight bias by half.

General Discussion

The RAFT model integrates theoretical concepts proposed by Frederic Bartlett, Egon Brunswik, and Herbert

Simon. Remembering is seen as a process of reconstruction (Bartlett) that involves cue-based inferences (Brunswik) in a "satisficing" way (Simon). In his seminal book *Remembering*, Bartlett (1932/1995) concluded that

remembering is not the re-excitation of innumerable fixed, lifeless and fragmentary traces. It is an imaginative reconstruction, or construction, built out of the relation of our attitude towards a whole active mass of organized past reactions or experience, and to a little outstanding detail which commonly appears in image or in language form. (p. 213)

However, Bartlett (1932/1995) did not specify how this (re)construction functions, that is, how exactly it is "built out of . . . our attitude towards a whole active mass." We suggest that, consistent with Brunswik's (1943, 1952) framework, this (re)construction is based on uncertain cues. Note that the framework of cue-based inferences inspired the RAFT model in a threefold way. First, cues in the original probabilistic mental model have been used to derive the original response; second, the reconstructed probabilistic mental model has been used to infer what this original response was; and third, feedback on the criterion served as a cue to update elusive cue values in the original probabilistic mental model. Rather than remaining vague as Bartlett did, or following the neo-Brunswikian idea that cues are weighted and integrated by multiple regression (Cooksey, 1996; Doherty, 1996; Hammond, 1955), we propose that the nature of the inferential mechanism is satisficing, following Simon (1982).

In the following, we discuss why we use Take The Best as a model for the inferential process involved in the reconstruction process and how RAFT relates to both other explanations of the hindsight bias and to similar phenomena. We conclude with a functional view of human memory.

Fast and Frugal Inferences

The Take The Best heuristic is fast and frugal: It is computationally simple compared with, for instance, multiple regression, and it does not search for all of the available information. Nevertheless, it can correctly infer real-world states as accurately as more complex algorithms (Gigerenzer & Goldstein, 1996; Gigerenzer, Todd, & the ABC Research Group, 1999). Aside from Take The Best, other heuristics have been proposed that are also quite simple, astonishingly accurate, and thus psychologically plausible, such as Elimination-By-Aspects (EBA; Tversky, 1972), Weighted Pros (Huber, 1979), Lexicographic (LEX; Fishburn, 1974), Lexicographic semiorder (Luce, 1956), and QuickEst (Hertwig, Hoffrage, & Martignon, 1999). Take The Best shares commonalities with these heuristics: For instance, similar to Take The Best, most of these heuristics can be characterized by a search rule, a stopping rule, and by the processing of the information in a noncompensatory fashion. However, there are also differences. EBA, for instance, eliminates alternatives, depending on an absolute threshold against which the cue value of a particular alternative is compared, whereas Take The Best does not require such a threshold but selects the alternative solely based on the relation between the cue values. Moreover, EBA checks cues in a probabilistic order,

whereas Take The Best consistently uses the order determined by cue validity (for a more comprehensive list of simple heuristics, as well as the relation of Take The Best to these other heuristics, see Rieskamp & Hoffrage, 1999).

How dependent is RAFT on the proposed inferential mechanism? To check the robustness of RAFT, we reanalyzed the data and tested Prediction 2 with several other heuristics, such as a unit-weight linear model, a linear model with cue validities as the weights, or naïve Bayes. None of the alternative heuristics modeled human judgment better than Take The Best; they all performed similarly well. The reason is that there were only three cues in our experiments; for most constellations of cue values, the various heuristics made the same inference (for the problem of separability of heuristics, see also Hoffrage, Martignon, & Hertwig, 1997). Thus, the results reported here seem to be robust across various candidate heuristics.

Although the proportion of explained judgments does not allow us to discriminate between RAFT and other more complex strategies, we favor RAFT over other strategies. Why? The reason is that the processes underlying RAFT are psychologically more plausible. First, RAFT is more frugal than any of the other heuristics; that is, it requires less information to draw an inference. Second, by relying on only one cue, namely the one which discriminates between the two alternatives, it has a very simple stopping rule for search, does not integrate information, and is thus computationally simple. Third, there is now a growing number of studies—specifically designed to discriminate between various strategies—that show that people in fact use these simple heuristics. For instance, Payne, Bettman, and Johnson (1988, 1993) observed that people select their strategies according to various conditions such as time pressure, memory load, or difficulty of the task. Less time, higher memory load, and more difficult tasks seem to favor strategies that rely on less information. Rieskamp and Hoffrage (1999) provided further support that people's choices, in particular those under time pressure, can best be modeled by heuristics that only process some of the available information, such as Take The Best. Bröder (in press, Experiments 3 and 4) showed that when information is costly, more than 60% of participants were classified as using Take The Best, whereas none were classified as using a compensatory unit-weight linear model. Although our experimental design did not involve explicit time pressure, it was a task that was (a) difficult (i.e., people had to retrieve numerous choice and confidence judgments that they had made in the most extreme case a week earlier), (b) it involved high memory load (all the strategies assume that the choices are based on cue values retrieved from memory), and (c) the long sequence of memory judgments was likely to have encouraged participants to search for few pieces of information and to respond quickly.

RAFT's performance ranges between 70% and 84%, depending on whether hindsight bias is measured on the level of choices and confidence combined or on the level of choice only and on whether cases of veridical recall were excluded from the analysis. Although in each possible test, RAFT's performance is significantly better than chance, it is

by no means perfect. In assessing RAFT's performance, however, we should not forget that the modeling was based on the knowledge the participants stated. For the purpose of modeling, we assumed that the recalled response was based on this knowledge. This is, of course a simplifying assumption. Other processes are likely to have occurred as well; for example, participants may have constructed a local mental model (Gigerenzer et al., 1991); that is, they may have used direct knowledge on the criterion, rather than using any cues at all. Or they may have used cues other than those we had taught or checked them in another order than the one we used when modeling their responses. The fact that none of the other simulated strategies outperformed RAFT indicates that the less than perfect performance is not due to the Take The Best module within RAFT.

Biased Reconstruction

RAFT is a candidate mechanism for a broader class of cognitive explanations in which a "reconstruction of the prior judgment by 'rejudging' the outcome" (Hawkins & Hastie, 1990, p. 321) is postulated and where hindsight bias is seen as a result of systematic differences between judging and rejudging the outcome. The causes of these differences may be located at various stages of the reconstruction. In their review of the explanations suggested so far, Hawkins and Hastie considered three subtasks that are probably involved in (re)judgment: sampling of evidence, interpretation of evidence, and integration of the implications of evidence.

At the level of the first subtask (sampling of evidence), hindsight bias could result from selective loss or suppression of evidence contradicting feedback (Dellarosa & Bourne, 1984). Assume that a participant has to judge whether the British or the Gurkhas of Nepal won the colonial war in the 19th century. She might have considered three pieces of evidence favoring the British and two favoring the Gurkhas, and perhaps she expressed some confidence in a British victory. If this participant were told that the British had won and then showed an increase in recalled confidence, this could be explained as a reconstruction on the basis of three recalled pieces of evidence favoring the British and only one favoring the Gurkhas. Pohl and Eisenhauer (1997) recently proposed a computational model of the hindsight bias (SARA: Selective Activation, Reconstruction, and Anchoring) that is based on the selective-loss hypothesis.

Hindsight bias could also result if the interpretation of a piece of evidence—the second subtask—depends on outcome feedback. For instance, heavy rain during a battle could be viewed as favoring the Gurkhas (e.g., because they are more likely to be used to such weather), whereas, with outcome feedback of a British victory, it could be viewed as favoring the British (e.g., because of their better equipment). Finally, hindsight bias can also be located in Hawkins and Hastie's third subtask (integration of the evidence implications). The evidence may be weighted differently before and after outcome knowledge is available. For instance, knowing that the British won could lead one to give greater weight

to weapons and equipment and less weight to motivation and familiarity with the territory.

RAFT shares with these previous attempts the assumptions that people construct their recollections by drawing cue-based inferences and that hindsight bias can be explained by different processing of the cues. The RAFT model adds a new candidate to Hawkins and Hastie's list: hindsight bias through updating of the evidence itself. We do not want to imply that any of the other suggestions are wrong; they neither contradict nor exclude but rather can complement each other. In fact, the RAFT model can account for a substantial number of the cases where (reversed) hindsight bias occurred, but not for all of them.

The assumption that those previous judgments, which cannot be recalled, are reconstructed is also the central notion of the change-of-standard approach (Higgins & Liberman, 1994; Higgins & Stangor, 1988). The change-of-standard approach explains memory biases related to the hindsight bias. Here a bias arises if a piece of information (e.g., Judge Jones sentenced a murderer to 15 years in prison) is stored in terms of a categorical judgment (e.g., he made a harsh decision) and the meaning of the category changes prior to the recollection of the information (e.g., because of information about the behavior of other judges under similar circumstances). Higgins and Liberman concluded that "using a judgment's past contextualized meaning is *not* natural. . . . Instead, people use a judgment's current categorical meaning as a default" (p. 255). Similarly, RAFT assumes that people use their current knowledge as a default.

RAFT can be used as a starting point to construct mechanisms for various phenomena where repeated measurement of choices or confidences is involved. For instance, variants of RAFT can be applied to phenomena such as the reiteration effect (Hasher, Goldstein, & Toppino, 1977), the exposure effect (Bornstein, 1989), eyewitness testimony (Loftus, 1979; McCloskey & Zaragoza, 1985), cognitive dissonance (Festinger, 1957), social conformity (Asch, 1958), or distorted recollection of the past due to outcome information and idiosyncratic expectancies of change (Conway & Ross, 1984; Hirt, McDonald, & Markman, 1998). To illustrate, consider the reiteration effect, where confidence in the truth of a statement increases by mere repetition of the statement. If one replaces the effect of feedback (updating elusive cue values) in RAFT by the parallel effect of repetition (updating recognition values, Goldstein & Gigerenzer, 1999; or familiarity values, Jacoby, 1991), then RAFT can be generalized to the reiteration effect: Repetition of a statement increases its likelihood of being recognized, which, in turn, may increase confidence that this statement is true.

Hindsight Bias as a By-Product of an Adaptive Process

We used the term hindsight bias because it is established in the literature. However, we do not view hindsight bias as a bias in the first place but as a consequence of learning by feedback (for a similar view, see Hoch & Loewenstein,

1989). Winman et al. (1998) recently proposed a model of the hindsight bias that is also based on the (helpful) role of feedback. Their "accuracy-assessment model" is formulated for tasks in which a "salient cognitive process" (p. 418) can be activated only to a low degree, such as for sensory discrimination tasks. Nevertheless, they arrive at a conclusion similar to ours, namely that the hindsight bias "is not an idiosyncratic and inexplicable information-processing bias but the consequence, or side-effect, of a perfectly reasonable consideration by the participants" (p. 429).

Whereas in the accuracy-assessment model, feedback affects the type of inference mechanism used, in the RAFT model it affects the input to the inference mechanism, which is the same prior to and after feedback. New incoming information, such as the feedback provided in our experiments, is evaluated against preexisting information; if the new information is more reliable, the preexisting information may be changed to obtain a more accurate corpus of knowledge. Such an automatic process of updating knowledge is consistent with Bartlett's (1932/1995) findings that schemata are constantly changing and being updated.

The adaptive function of this knowledge updating in our semantic memory is that it enables us to improve our inferences over time. In the case of hindsight bias, however, inferences about what we previously said may be in error, thus making it difficult for us to learn from our past. Nevertheless, hindsight bias may not be much of an adaptive disadvantage. Remembering the real state of affairs (e.g., whether something is true or really happened) is generally more important than remembering what one thought about it before learning the truth. As Bartlett (1932/1995) put it: "In a world of constantly changing environment, literal recall is extraordinarily unimportant" (p. 204). Moreover, the ability to access our previous knowledge states would require significant storage space and would lead to memory overload; forgetting may be necessary for memory to maintain its function (Hoffrage & Hertwig, 1999). Another advantage of forgetting is that it prevents one from using old information that may be outdated because of changes in the environment (Bjork & Bjork, 1988; Ginzburg, Janson, & Ferson, 1996). Taken together, the disadvantage of hindsight bias is a relatively cheap price to pay for making better inferences and maintaining a well-functioning memory.

References

- Arkes, H. R., Wortmann, R. L., Saville, P. D., & Harkness, A. R. (1981). Hindsight bias among physicians weighting the likelihood of diagnoses. *Journal of Applied Psychology*, 66, 252-254.
- Asch, S. E. (1958). Effects of group pressure upon the modification and distortion of judgments. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology* (3rd ed., pp. 174-183). New York: Holt.
- Bartlett, F. C. (1995). *Remembering*. Cambridge, England: Cambridge University Press. (Original work published 1932)
- Bjork, E. L., & Bjork, R. A. (1988). On the adaptive aspects of retrieval failure in autobiographical memory. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol II-1, pp 283-286). Chichester, England: John Wiley.

- Bornstein, R. F. (1989). Exposure and affect: Overview and meta-analysis of research, 1968–1987. *Psychological Bulletin*, 106, 265–289.
- Bröder, A. (in press). Assessing the empirical validity of the “Take The Best” heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, 50, 255–272.
- Brunswik, E. (1952). The conceptual framework of psychology. In *International encyclopedia of unified science* (Vol. 1, No. 10, pp. 4–102). Chicago: University of Chicago Press.
- Bukszar, E., & Connolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. *Academy of Management Journal*, 31, 628–641.
- Christensen-Szalanski, J. J., & Fobian Willham, C. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48, 147–168.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Conway, M., & Ross, M. (1984). Getting what you want by revising what you had. *Journal of Personality and Social Psychology*, 47, 738–748.
- Cooksey, R. W. (1996). *Judgment analysis: Theory, methods, and applications*. San Diego, CA: Academic Press.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). In G. Gigerenzer, P. M. Todd, & the ABC Research Group. *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.
- Davies, M. F. (1987). Reduction of hindsight bias by restoration of foresight perspective: Effectiveness of foresight-encoding and hindsight-retrieval strategies. *Organizational Behavior and Human Decision Processes*, 40, 50–68.
- Dehn, D., & Erdfelder, E. (1998). What kind of bias is hindsight bias? *Psychological Research*, 61, 135–146.
- Dellarosa, D., & Bourne, L. E., Jr. (1984). Decisions and memory: Differential retrievability of consistent and contradictory evidence. *Journal of Verbal Learning and Verbal Behavior*, 23, 669–682.
- Doherty, M. E. (Ed.). (1996). Social judgment theory [Special Issue]. *Thinking & Reasoning*, 2, 105–248.
- Erdfelder, E., & Buchner, A. (1998). Decomposing the hindsight bias: A multinomial processing tree model for separating recollection and reconstruction in hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 387–414.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford, CA: Stanford University Press.
- Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.
- Fischhoff, B. (1982). Debiasing. In D. Kahnemann, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge, England: Cambridge University Press.
- Fischhoff, B., & Beyth, R. (1975). “I knew it would happen.” Remembered probabilities of once-future things. *Organizational Behavior and Human Performance*, 13, 1–16.
- Fishburn, P. C. (1974). Lexicographic orders, utilities and decision rules: A survey. *Management Science*, 20, 1442–1471.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506–528.
- Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Ginzburg, L. R., Janson, C., & Ferson, S. (1996). Judgment under uncertainty: Evolution may not favor a probabilistic calculus. *Behavioral and Brain Sciences*, 19, 24–25.
- Goldstein, D., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & the ABC Research Group. *Simple heuristics that make us smart* (pp. 59–72). New York: Oxford University Press.
- Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review*, 62, 255–262.
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16, 107–112.
- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgment of the past events after the outcomes are known. *Psychological Bulletin*, 107, 311–327.
- Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory & Cognition*, 16, 533–538.
- Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. *Psychological Review*, 104, 194–202.
- Hertwig, R., Hoffrage, U., & Martignon, L. (1999). Quick estimation: Letting the environment do the work. In G. Gigerenzer, P. M. Todd, & the ABC Research Group. *Simple heuristics that make us smart* (pp. 209–234). New York: Oxford University Press.
- Higgins, E. T., & Liberman, A. (1994). Memory errors from a change of standard: A lack of awareness or of understanding? *Cognitive Psychology*, 27, 227–258.
- Higgins, E. T., & Stangor, C. (1988). A “change-of-standard” perspective on the relations among context, judgment, and memory. *Journal of Personality and Social Psychology*, 54, 181–192.
- Hirt, E. R., McDonald, H. E., & Markman, K. D. (1998). Expectancy effects in reconstructive memory: When the past is just what we expected. In S. J. Lynn & K. M. McConkey (Eds.), *Truth in memory* (pp. 62–89). New York: Guilford Press.
- Hoch, S. J., & Loewenstein, G. F. (1989). Outcome feedback: Hindsight and information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 605–619.
- Hoffrage, U. (1995). *Zur Angemessenheit subjektiver Sicherheitsurteile. Eine Exploration der Theorie der probabilistischen mentalen Modelle* [The adequacy of subjective confidence judgments: Studies concerning the theory of probabilistic mental models]. Doctoral dissertation, University of Salzburg, Salzburg, Austria.
- Hoffrage, U., & Hertwig, R. (1999). Hindsight bias: A price worth paying for fast and frugal memory. In G. Gigerenzer, P. M. Todd, & the ABC Research Group. *Simple heuristics that make us smart* (pp. 191–208). New York: Oxford University Press.
- Hoffrage, U., Martignon, L., & Hertwig, R. (1997, August). Does “judgment policy capturing” really capture the policies? Poster presented at Subjective Probability, Utility, and Decision Making-16, Leeds, UK.
- Huber, O. (1979). Nontransitive multidimensional preferences: Theoretical analysis of a model. *Theory and Decision*, 10, 147–165.
- Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, 30, 513–541.

- Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
- Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica*, 24, 178–191.
- Martignon, L., & Laskey, K. B. (1999). Bayesian benchmarks for fast and frugal heuristics. In G. Gigerenzer, P. M. Todd, & the ABC Research Group. *Simple heuristics that make us smart* (pp. 169–188). New York: Oxford University Press.
- McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114, 1–6.
- Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, 9, 1–22.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 534–552.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. New York: Cambridge University Press.
- Pennington, D. C. (1981). The British fireman's strike of 1977/78: An investigation of judgments in foresight and hindsight. *British Journal of Social Psychology*, 20, 89–96.
- Pohl, R. F., & Eisenhauer, M. (1997). SARA: An associative model for anchoring and hindsight bias. In M. G. Shafto, & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (p. 1103). Mahwah, NJ: Erlbaum.
- Rieskamp, J., & Hoffrage, U. (1999). When do people use simple heuristics, and how can we tell? In G. Gigerenzer, P. M. Todd, & the ABC Research Group. *Simple heuristics that make us smart* (pp. 141–167). New York: Oxford University Press.
- Simon, H. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.
- Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprise. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 544–551.
- Stahlberg, D., & Maass, A. (1998). Hindsight bias: Impaired memory of biased reconstruction? In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 8, pp. 106–132). New York: Wiley.
- Tolstoy, L. (1982). *War and peace*. London: Penguin Classics. (Original work published 1869)
- Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281–299.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Winman, A., Juslin, P., & Björkman, M. (1998). The confidence-hindsight mirror effect in judgment: An accuracy-assessment model for the knew-it-all-along phenomenon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 415–431.
- Wohlsetter, R. (1962). *Pearl Harbor: Warning and decision*. Stanford, CA: Stanford University Press.

Received May 3, 1998

Revision received November 1, 1999

Accepted November 16, 1999 ■