

Hindsight bias: How knowledge and heuristics affect our reconstruction of the past

Ralph Hertwig

Columbia University, USA, and Max Planck Institute for Human Development, Berlin, Germany

Carola Faselow and Ulrich Hoffrage

Max Planck Institute for Human Development, Berlin, Germany

Once people know the outcome of an event, they tend to overestimate what could have been anticipated in foresight. Although typically considered to be a robust phenomenon, this *hindsight bias* is subject to moderating circumstances. In their meta-analysis, Christensen-Szalanski and Willham (1991) observed that the more experience people have with the task under consideration, the smaller is the resulting hindsight bias. This observation is one benchmark against which the explanatory power of process models of hindsight bias can be measured. Therefore, we used it to put the recently proposed RAFT model (Hoffrage, Hertwig, & Gigerenzer, 2000) to another test. Our findings were consistent with the “expertise effect.” Specifically, we observed—using computer simulations of the RAFT model—that the more comprehensive people’s knowledge is in foresight, the smaller is their hindsight bias. In addition, we made two counterintuitive observations: First, the relation between foresight knowledge and hindsight bias appears to be independent of how knowledge is processed. Second, even if foresight knowledge is false, it can reduce hindsight bias. We conclude with a discussion of the functional value of hindsight bias.

Recollection or re-evaluation of past events can be affected by what has happened since. While interesting in its own right—as it sheds light on the working of human memory—this phenomenon, *hindsight bias*, also matters because it affects how we evaluate the actions of others. Take historians as an example. Historians are hermeneuts of the past, trying to explain why things turned out the way they did. They must, for instance, evaluate the appropriateness of *ex ante* behaviour (e.g., Napoleon’s decision to invade Russia) that resulted in bad or good *ex post* outcomes. By necessity historians are cognisant of the outcome, and this knowledge can affect their evaluations. A behaviour, for instance, may be judged to be more neglectful

than it would have been judged without knowing its negative consequences.

As an example, consider the tragic failure of an adventure that has captured the public’s imagination: Robert F. Scott’s race to be first to reach the South Pole. In November 1911, Scott led a British team in an attempt to reach the Pole. After marching and skiing more than 900 miles, Scott and his four companions reached their goal in January 1912, only to find that Amundsen and his Norwegian colleagues had beaten them by almost a month. On their way back, Scott and his compatriots froze to death in a tent just a few miles short of a depot of food and heating oil. “When words of their deaths reached England, Scott was hailed as a hero, an exemplar of

Requests for reprints should be sent to Ralph Hertwig, Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzeallee 94, D-14195 Berlin, Germany. Email: hertwig@mpib-berlin.mpg.de

This work was completed while Ralph Hertwig was a visiting scholar at Columbia University. We thank the Deutsche Forschungsgemeinschaft for its financial support of Ralph Hertwig with Grant HE 2768/6-1 and of Ulrich Hoffrage with Grant Ho 1847/1. We are grateful to Oliver Hardt, Rüdiger Pohl, Lael Schooler, and Anders Winman for many helpful comments, and we thank Callia Piperides and Anita Todd for editing the manuscript.

English gentlemanly pluck in the face of dire adversity" (*The New York Times*, Science section, 28 August 2001). In recent decades, however, historians have turned to less flattering second-guessing of Scott's actions. For instance, the British historian Roland Huntford sought to revise the public's view of Scott. With the benefit of hindsight, he questioned many of Scott's decisions, such as why Scott and his men acted as their own pack animals, pulling a sled loaded with more than 200 pounds of equipment and supplies. He also asked how it was possible that Scott and his crew were not prepared for the gruelling temperatures. In his foreword to the new edition of Huntford's book (1999), the well-known travel writer Paul Theroux continued the tradition of denigrating Scott, describing him as "insecure, dark, panicky, humorless, an enigma to his men, unprepared, and a bungler, but in the spirit of a large-scale bungler, always self-dramatizing" (p. xiv).

To what extent is Huntford's, and for that matter Theroux's, view of Scott tainted by the knowledge of the expedition's tragic end? In a new book, Susan Solomon (2001), senior scientist at the US National Oceanic and Atmospheric Administration, has analysed meteorological data of the last 17 years from weather stations in Antarctica and compared them with weather information from the diaries and letters of the men on the Scott expedition. Based on these data she argued that an extremely rare spell of dramatic cold was the deciding factor in Scott's fatal expedition. Contradicting Huntford's judgement, she concluded that Scott and his crew "planned meticulously, and they were undone by an act of nature . . . It would have been a perfectly workable plan in a normal year" (*The New York Times*, Science section, 28 August 2001).

What was Scott—a neglectful bungler or a meticulous planner? Although we may never learn the truth, psychological research can at the least elucidate how such drastically different views of Scott's personality could have emerged. Clearly, book authors writing on the same events have strategic incentives to overemphasise their differences in opinion. Hindsight bias research, however, suggests another key factor that may have contributed to Huntford and Solomon's diverging views. For the sake of argument, let us assume that Huntford's judgement is tainted more than Solomon's by the benefit of knowing the fatal outcome of the expedition. How could that be? One possibility is that Solomon's edge in know-

ledge, that is, her knowing about the exceptionally icy temperatures, was instrumental in protecting her from exaggerating what Scott could have anticipated in foresight. In other words, being more knowledgeable may in fact have guarded her from concluding that the expedition's fatal outcome was inevitable and foreseeable. Is there reason to believe that a person's amount of knowledge is related to the hindsight bias? The brief answer is yes. As we describe in the next section, an extensive meta-analysis concluded that the more experienced a person is with the task under consideration, the smaller is the effect of the hindsight bias.

MODERATOR VARIABLES OF HINDSIGHT BIAS

Hindsight bias is one of the most frequently cited cognitive biases (Christensen-Szalanski & Beach, 1984). Not surprisingly, it is also one of the most researched. In a meta-analysis of hindsight bias research, Christensen-Szalanski and Willham (1991) analysed a total of 128 studies to identify important moderator variables of the phenomenon. They focused on two variables—one is the question of whether or not hindsight bias is more pronounced when people are told that an event occurred versus did not occur. The second variable is the effect of what the authors referred to as people's "familiarity", "expertise", and "experience" with the task. In what follows, we focus on this latter moderator variable (see Hertwig, Gigerenzer, & Hoffrage, 1997, for an account of the first variable).

To examine the impact of people's experience, Christensen-Szalanski and Willham (1991) coded each study within a large set of studies on hindsight bias as either "familiar" or "unfamiliar". For example, Arkes, Wortmann, Saville, and Harkness's (1981) study was coded as familiar because their participants were experts in the field from which questions were sampled. Specifically, Arkes et al. asked physicians to assign to each of four possible diagnoses entertained in a medical case history, the probability they thought they would have assigned. One group of physicians made their estimates without knowing the outcome (i.e., the actual disease), whereas all others arrived at their estimates after having been told which of the four possible diagnoses was correct. In their pool of studies, Christensen-Szalanski and Willham (1991) coded about half of the studies as

“familiar” and “unfamiliar”, respectively. Did the size of hindsight bias differ between these two sets? It did. Christensen-Szalanski and Willham (1991) reported that “the more familiar the subject is with the task, the smaller the effect of the hindsight bias” (p. 155). This effect is of medium to large size ($r = 0.42$ when corrected for sampling error). Thus, Christensen-Szalanski and Willham’s meta-analysis established experience to be a key moderator of hindsight bias. By identifying this effect, henceforth the *expertise effect*, their meta-analysis has provided hindsight bias research with a key empirical benchmark against which the explanatory power of models of hindsight bias can be evaluated.

EXPERT PERFORMANCE

How do experts and novices differ, and can these differences help to explain why novices are more disposed to the hindsight bias? The prolific research on expert performance (e.g., Chase & Simon, 1973; Chi, Feltovitch, & Glaser, 1981; de Groot, 1946/1965; Larkin, McDermott, Simon, & Simon, 1980) has revealed a number of differences in the way experts and novices go about solving problems; some of those are immediately relevant to hindsight bias. First and foremost, knowledge emerges as an “essential prerequisite to expert skill”, and “the extent of the knowledge an expert must be able to call upon is demonstrably large” (Larkin et al., 1980, p. 1342). How large an expert’s repertoire of knowledge can be is illustrated in Simon and Barenfeld’s (1969) classic study of master chess players. According to their estimates, a player (after at least several years of serious occupation with the game) is expected to have acquired a “vocabulary” of familiar sub-patterns with a size of 10,000 to 100,000 patterns.

However, experts differ from novices in more than just their sheer amount of knowledge. The famous German physicist Werner Heisenberg (1971) suggested that an expert is someone who knows some of the worst mistakes that can be made in his subject, and how to avoid them. One interpretation of Heisenberg’s portrayal of expert performance is that expert knowledge may also be more veridical than the knowledge of novices.

It makes intuitive sense to call upon differences in the quantity and quality of knowledge to explain expert–novice differences. These presumed differences in knowledge, however, present what Ericsson and Staszewski (1989) called a

“thorny” problem: How do experts process an enormous amount of information, given that they are subject to the same or similar elementary information-processing limits as novices? Shanteau (1992), a prominent scholar of expert decision making, asserted that “experts should use *all* relevant information” (p. 253, emphasis added), defying those limits. By suggesting that experts retrieve all the information available (either from internal or external memories) and combine various aspects into a single judgement, he thus depicts expert decision making as akin to rational decision making. Specifically, he echoed two commandments that are often taken as characteristics of rational choices and judgements, namely, “*complete search*” and “*compensation*” (see Gigerenzer & Goldstein, 1999). The former prescribes, “thou shalt find all the information available”, while the latter says, “thou shalt combine all pieces of information” (i.e., not rely on just one piece).

To conclude, expert and novice performance has been demonstrated or suggested to differ on multiple dimensions, among them the amount, the accuracy, and the processing of knowledge. In principle, both the combination of these dimensions as well as each one individually may be able to account for the expertise effect observed by Christensen-Szalanski and Willham (1991). In what follows, we investigate how each of these dimensions affects hindsight bias. In our investigation, we employ the RAFT model (Reconstruction After Feedback with Take The Best). This recent process model of the cognitive processes underlying the hindsight bias (Hoffrage & Hertwig, 1999; Hoffrage et al., 2000) affords us the opportunity to map the three dimensions of expertise considered here—amount, accuracy, and processing of foresight knowledge—into a single theoretical framework, and then to analyse their impact on hindsight bias.

THE RAFT MODEL

Before introducing the RAFT model, we first distinguish two different types of research designs employed in hindsight bias research. The *hypothetical design* approximates the situation of historians (such as Huntford) who typically evaluate an event (e.g., Scott’s fatal expedition) in hindsight without having given an assessment prior to its occurrence. Specifically, this design

compares two groups of participants: One group has no outcome information, and the other has such information but is asked to ignore it (e.g., the physicians in Arkes et al.'s, 1981, study). Finally, a comparison is made between the judgements of both groups. In contrast, the *memory design* approximates everyday situations in which individuals (e.g., weather forecasters, political pollsters) predict an event, learn about the actual outcome, and then eventually remember their previous judgement. Because the RAFT model was primarily designed to model hindsight bias judgements in the context of the memory design, we henceforth focus on this design.

The RAFT model is based on the theory of probabilistic mental models (Gigerenzer, Hoffrage, & Kleinbölting, 1991). This theory models the cognitive processes in two-alternative-choice tasks, in which people are required to make inferences about which of two objects, *a* or *b*, has a higher value on some quantitative dimension (henceforth, *original choice*). The RAFT model applies this theoretical framework to a repeated measurement context in which a previous choice (made at Time 1) needs to be recalled (at Time 3) after receiving feedback (at Time 2) on the correct choice (Hoffrage et al., 2000).

The RAFT model makes three assumptions about this recollection process (at Time 3): First, if the original choice (made at Time 1) cannot be retrieved from memory, it will be reconstructed by rejudging the problem. Second, the reconstruction involves the attempt to recall the knowledge on which the original choice was based. Third, the outcome information received (at Time 2) is used to update old knowledge, in particular knowledge that was elusive and missing at Time 1. In conjunction, these assumptions suffice to explain the occurrence of hindsight bias. Thus, the RAFT model suggests that outcome information does not directly affect the memory trace for the original choice but exerts its impact indirectly by updating knowledge that is used to reconstruct the original choice (in the context of hindsight bias, the notion of reconstruction has been proposed by, for instance, Hawkins & Hastie, 1990; Stahlberg & Maass, 1998; see also Schwarz & Stahlberg, 2003-this issue).

We now specify in detail the cognitive processes underlying the original choice at Time 1 and the recalled choice at Time 3.

Time 1: Original choice

An anecdote helps to illustrate the proposed processes: A couple of months before the 2000 US presidential election, two German colleagues of ours, Peter and Michael, had bet on its outcome. While Peter deemed Al Gore to be the likely winner, Michael was convinced that George W. Bush would win. As we all know, the election was not settled until five weeks after election day when the US Supreme Court's intervention finally brought the contest to an end. This unusually long delay may have contributed to the fact that, very much to Michael's chagrin, Peter plainly forgot about their wager and when reminded, recalled having picked Bush rather than Gore as the likely winner (for a study of hindsight bias in the context of political events, see, for example, Blank, Fischer, & Erdfelder, 2003-this issue; Synodinos, 1986; Wendt, 1993).

How would RAFT account for Peter's retrospective belief that he had picked Bush rather than Gore after he learned that Bush won the election? The first step in the model is to account for the original choice: Not knowing who would win, Peter initially tried to infer the more likely winner from what he knew about the two candidates. According to the RAFT model, Peter constructed a probabilistic mental model to make the inference. Such a model connects the specific structure of the task with a probability structure of a corresponding natural environment (stored in long-term memory) and consists of *knowledge* in terms of a reference class, probability cues, and the cue values of the objects on the cues. Before we describe this knowledge in more detail, let us stress that henceforth we use the terms "knowledge" and "know" in a rather narrow sense, namely, to refer to the cue values a person has stored in long-term memory (regardless of whether or not those cue values are accurate).

Knowledge of cues and cue values. In Peter's case, the reference class might be some set of previous presidential elections (e.g., all elections since 1948) with the competing candidates in those races as objects that compose the reference class. Each candidate can be described on a number of cues related to the criterion "outcome of the election". Cues are variables that covary with the criterion, thus allowing a person to use them as predictors for the criterion variable. Cue values are the values of the objects on the cues. In the case of dichotomous cues, the cue values are "positive" and "negative".

Which cues may come to mind when one attempts to forecast the outcome of political elections? According to the common wisdom of political forecasters (e.g., Lichtman, 1996), the outcome of the presidential election can be inferred on the basis of predictors such as the *re-election*, the *incumbent party*, and the *economic prosperity* cues: The first cue refers to the observation that if the President is running for re-election, he (or she in the future) may have a head start (e.g., because of being well known to voters), and so may the candidate of the incumbent party. The third cue refers to the observation that the party that promises to better maintain prosperity in the future has an advantage (Campbell & Mann, 1996). Moreover, it appears that personal features of the candidates can also be predictive of their success or lack thereof. For instance, the candidate who is charismatic (e.g., Kennedy), a national hero (e.g., Eisenhower), or plainly taller (than the opponent) has been suggested to have better chances of winning.

Clearly, some of these cues are better predictors than others. A cue's predictive ability is captured in the notion of *ecological validity*, which is defined as the relative frequency with which the cue correctly predicts which object (here candidate) scores higher on the criterion (in a defined reference class). The validity of the re-election cue, for instance, has an ecological validity of 75% (assuming a reference class that consists of the presidential races between 1948 and 1996). In contrast, the incumbent party cue only has a modest validity of 54% (again considering the races between 1948 and 1996).

Inference mechanism. Let us assume that Peter's probabilistic mental model includes four cues, with the economic prosperity cue ranked highest, followed by the charisma, re-election, and incumbent party cues. The RAFT model accounts for Peter's inference with a processing strategy called the "Take The Best" heuristic. This lexicographic strategy assumes a subjective rank order of cues according to their validities and makes the inference on the basis of the best (i.e., most valid) cue that discriminates. The three building blocks of the heuristic (excluding the recognition principle, which is not relevant here; see Gigerenzer & Goldstein, 1996) are:

- *Search:* Choose the cue with the highest validity and retrieve the object's cue values from memory.

- *Stop:* If the best cue discriminates, stop searching. The cue is said to discriminate between two objects if one has a positive cue value and the other does not. If the best cue does not discriminate, continue with the next best cue until a cue that discriminates is found.
- *Decide:* Choose the object to which the cue points, that is, the object with the positive cue value (if criterion and cues are negatively correlated, then choose the object with the negative cue value). If no cue discriminates, then choose randomly. Note that for the purpose of illustration, we treat all cues in the Gore–Bush example as binary cues although some, such as the charisma cue, may be continuous (for the treatment of continuous cues within the RAFT framework, see Hoffrage et al., 2000).

To illustrate the heuristic's policy, Table 1 shows the Take The Best heuristic applied to Peter's knowledge about the candidates. At Time 1, Peter does not know the values for the highest-ranked cue, the economic prosperity cue, and thus Take The Best cannot use it. In addition, the second-ranked cue, charisma, does not discriminate between the candidates either: Although Peter knows that Gore is widely considered to lack charisma ("Gore the Bore"), he does not know how Bush scores on this cue dimension. Since neither of the candidates is the sitting president seeking re-election, Peter's third-ranked cue does

TABLE 1
Hindsight bias in the presidential election inference

Cues	Time 1		Feedback	Time 3	
	Gore	Bush		Gore	Bush
Economic prosperity	?	?		?	?
Charisma	–	?	→	–	+
Re-election	–	–		–	–
Incumbent party	+	–		+	–
Response	"Gore"			"Bush"	

The probabilistic mental model contains four cues ranked according to their (assumed) validity. Cue values are positive (+) or negative (–); missing knowledge is indicated by question marks (?). To predict the winner of the election, the Take The Best heuristic looks up only the cue values in the shaded areas. The final decision is determined solely on the basis of the cue values in the lowest shaded cell. At Time 3, the cue value of the charisma cue for Bush shifts towards feedback, that is, from "?" to "+". As a consequence, this cue now discriminates and points to Bush—hindsight bias occurs.

not discriminate either. Therefore, Take The Best determines the choice on the basis of the fourth-valid cue, the only one that discriminates between the candidates. Because Gore is the candidate of the incumbent party, the heuristic picks him as the candidate with the better chance of winning.

Time 3: Reconstruction

Why does Peter misremember his original choice? One necessary reason for the occurrence of Peter's hindsight bias is that he is not able to retrieve his original choice from memory directly. According to the RAFT model, if the original response cannot be retrieved, it will be reconstructed by reiterating the steps taken at Time 1.

The reconstruction process begins by retrieving the knowledge on which the choice at Time 1 was based, that is, by retrieving the original cues and their values. In some cases, veridical retrieval may be possible; in others, memory of the cue values may be elusive or missing—either because the knowledge retrieval from long-term memory is not completely reliable or because knowledge was elusive or missing at Time 1. The RAFT model's critical assumption is that outcome knowledge (e.g., Bush won the presidential race) transforms some of the elusive and missing cue values into positive or negative values, thus possibly turning non-discriminatory cues into discriminatory ones. This is due to the reversibility of the cue-criterion relationship: Because it is possible to draw inferences from a cue (e.g., height) to the criterion, the reverse is also possible—to draw inferences from the criterion to the cues. Thus, what used to be the distal variable (i.e., outcome of the election) at Time 1 now turns into a proximal cue and is used to infer what used to be a proximal cue at Time 1 (e.g., charisma) and what turns into a distal variable at Time 3. Such a reversal between proximal cues and distal variable is possible because cues and criterion are correlated with each other.

To illustrate this, let us consider Peter's updated probabilistic mental model at Time 3. After the 8 December ruling of the Supreme Court, Peter attempts to reconstruct his original choice. RAFT assumes that the new information concerning the de facto winner affords the mind inferences about some of the cue values that were unknown at Time 1. That is, not all initial cue values may be veridically remembered but some will have taken on values consistent with the newly acquired outcome information. As Table 1 shows, in Peter's *updated* mental model, Bush's value for charisma is now

seen as being "positive". Consequently, this cue now discriminates and points to Bush as the likely winner. If the same heuristic (here, Take The Best) is applied to this updated knowledge base, the reconstructed choice will be consistent with the outcome information. In other words, Peter remembers having originally deemed Bush to be the winner, thus exhibiting hindsight bias.¹

It is important to note that within the RAFT model, updating cue values is thought to be a probabilistic process in which some, but not all, missing cue values are updated. As a consequence, the RAFT model cannot predict which of the elusive and missing cue values will be updated. It can, however, use the updated cue values to predict which recollections of the original judgement will exhibit hindsight bias. In addition, RAFT assumes that knowledge retrieval from long-term memory is not perfectly reliable. In other words, it also posits random deviations between the cue values at Time 1 and Time 3. Because such alterations are independent of outcome knowledge, they can explain why for a given item hindsight bias (if they coincide with the direction of the actual outcome), reversed hindsight bias (if they are counter to the direction of the actual outcome), or no hindsight bias occurs.²

¹ While we use the Gore-Bush competition simply as an illustration, it does highlight an issue that, to the best of our knowledge, has hardly been addressed. Hindsight bias research typically assumes that the outcome of an event is unambiguous. To the chagrin of both presidential candidates, however, the outcome of the 2000 presidential election was ambiguous. In light of the fact that Gore won the popular vote (though did not reach the majority in the electoral college), some people (including a reviewer of this paper) argued that Gore was the "winner". Be this as it may, the outcome of the 2000 presidential election demonstrates that, in real life, outcomes can be ambiguous or at least may be perceived as such. If so, one may speculate that the benefit of knowing the outcome may lose some of its alluring impact.

² Explaining item-specific *reversed* hindsight bias through unsystematic changes in cue values does not exclude other accounts of the hindsight bias reversals, such as the surprise account proposed by Mazursky and Ofir (1990), and Ofir and Mazursky (1997) (for another account that treats feelings as information, see Werth & Strack, 2003-this issue). It is interesting to note that RAFT's core assumption of updating could also play a key role in Pezzo's (2003-this issue) sense-making model which builds on the very notion of surprise. His model assumes that in cases in which outcome information is incongruent with prior expectations, a sense-making process will be activated. Specifically, Pezzo predicts that if sense making succeeds, no "resultant surprise" is experienced, and hindsight bias will occur. On the assumption that sense making is more likely to succeed if updating has occurred, RAFT provides a cognitive rationale for why Pezzo's account predicts hindsight bias in this case.

The RAFT model can be summarised as follows: If the original choice cannot be retrieved from memory, an attempt is made to reconstruct the probabilistic mental model that led to it. An identical reconstruction requires the type of inference strategy for the original choice and its reconstruction to be the same, and the input into the strategy also to remain the same. Any violation of these requirements may lead to differences between the original and the reconstructed choice. The RAFT model assumes that feedback changes the input (i.e., the cue values) into the inferential strategy but does not exclude the possibility that other requirements may also be violated (and there are indeed such accounts of hindsight bias, e.g., Hawkins & Hastie, 1990).

Let us conclude this introduction of the RAFT model with a caveat. The model focuses on the context of choices (here between two alternatives, but it can, in principle, be applied to choices among multiple alternatives). Choices are in fact the paradigmatic context of economic theory and, according to Kahneman and Tversky (1984), “making decisions [i.e., risky and riskless choices] is like speaking prose—people do it all the time, knowingly or unknowingly” (p. 341). The context of choices, however, is not the only context in which hindsight bias has been observed. Others are estimation of quantities, and confidence judgements. For this reason, we need to caution that the scope of the RAFT model is limited to one task, albeit a ubiquitous one. Thus, the RAFT model does not exclude other accounts of hindsight bias such as the SARA model that has been designed to account for hindsight bias in numerical estimates (Pohl, Eisenhauer, & Hardt, 2003—this issue).

Before we describe how the RAFT model was implemented in the present computer simulation, we briefly review some of the available empirical evidence that can be marshalled in support of it.

THE RAFT MODEL: EMPIRICAL SUPPORT

How do people make choices between two objects based on a bundle of imperfect cues? The Take The Best heuristic embodies the bold possibility that only a single imperfect cue will be used to make such a choice, thus minimising both the information-searching costs (e.g., in terms of time) and the computational costs. This policy is what Gigerenzer, Todd, and the ABC Research Group (1999b) have called *one-reason decision making*.

One advantage of this decision-making policy is that it avoids conflicts between those cues that point in opposite directions. Avoiding such conflicts makes the Take The Best heuristic *non-compensatory*, which means that a cue supporting alternative *A* cannot be outweighed by any combination of less important cues, even if they all support alternative *B*.

Do people employ the Take The Best heuristic? There have been several independent investigations into the descriptive validity of the Take The Best heuristic and variants thereof (e.g., Bröder, 2000; Rieskamp & Hoffrage, 1999; Sleegers, Brake, & Doherty, 2000). Recently, Hertwig and Hoffrage (2001) reviewed this set of published studies. While acknowledging that conditions exist under which the majority of people could not be classified as using the Take The Best heuristic (see Bröder, 2000, Study 1), it seems fair to conclude that two key conditions of decision making—time pressure and the imposition of costs on information search and use—favour the use of non-compensatory strategies such as the Take The Best heuristic. For instance, Bröder (2000, Study 4) showed that when participants had to search for costly information, 65% of them were classified as using the Take The Best heuristic. In contrast, less than 10% could be classified as using a simple linear decision strategy (with unit weights).

Another core assumption of the RAFT model is that outcome knowledge (e.g., Bush won the presidential race) transforms some of the elusive and missing cue values into positive or negative values, thus possibly turning non-discriminatory cues into discriminatory ones. Does this updating actually occur? It does. As can be seen in Figure 1,

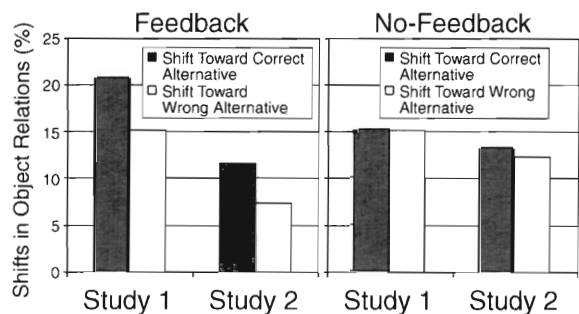


Figure 1. Percentage of shifts of object relations towards and away from the correct alternative in the feedback and no-feedback conditions in two studies (adapted from Hoffrage et al., 2000). The term *object relations* refers to the relation of objects with respect to a cue. This relation can be larger, smaller, equal, or unknown, and refers both to continuous cues (Study 1) and binary cues (Study 2).

Hoffrage et al. (2000) observed that after feedback on the correct alternative (at Time 2), more cue values (object relations; see figure legend) shifted towards the correct alternative than away from it (at Time 3). In contrast, cue values in the control condition (i.e., without feedback) shifted equally often towards and away from it. This finding was obtained in two independent studies—one that used binary cues (Study 1) and one that used continuous cues (Study 2).

The RAFT model accounts for the observed outcomes at Time 3 (hindsight bias, reversed hindsight bias, or veridical recall) on the basis of the cue values at Time 3. Do the cue values at Time 3 in fact determine the observed outcomes? Hoffrage et al. (2000, see their Figure 5) found that in 83.5% (Study 1) and 69.5% (Study 2) of the cases, the outcomes predicted by the RAFT model matched those actually observed (for various statistical tests of the performance of the RAFT model, see Hoffrage et al., 2000, pp. 572–577).

In sum, there are several empirical results that are consistent with the RAFT model and its building blocks such as the Take The Best heuristic and the assumption of knowledge updating. Can we add to this collection of results by demonstrating that the RAFT model can also account for Christensen-Szalanski and Willham's (1991) expertise effect? In what follows, we describe our investigation into this question in detail.

THE RAFT MODEL: IMPLEMENTATION

Using computer simulations, we investigated possible determinants of the expertise effect. Specifically, we examined the impact of the amount of foresight knowledge (i.e., how much does a person know at Time 1), knowledge accuracy (how accurate or inaccurate is a person's knowledge at Time 1), and knowledge processing (how is a person's knowledge processed at Times 1 and 3) on hindsight bias.

Environment

We conducted the simulations using a real-world environment, namely, German cities. In this environment, simulated "individuals" first (Time 1) answer real-world questions such as, "Which city has more inhabitants: (a) Essen or

(b) Bremen?". Then, (Time 2) they learn the correct answer (e.g., "Essen"). Finally, in the attempt to reconstruct the original answers, they rejudge the same questions at Time 3. The city environment consists of the set of German cities with more than 100,000 inhabitants (excluding Berlin, 82 cities), with population size as the criterion variable. The environment includes eight binary ecological cues (see Table 3) and the actual 8×82 positive and negative values of the objects (cities) on the cues. The cues include predictors such as the soccer-team cue ("Does the city have a team in the major league?") and the state-capital cue ("Is the city a state capital?"). The complete environment (e.g., cues and cue values) is shown in Gigerenzer and Goldstein (1996). Next, we describe the parameters that we systematically varied in the simulations (see Table 2).

Knowledge: Amount, accuracy, and updating

In the present simulation, the amount of knowledge was simply the percentage of cue values a "person" knows. Within the German city size environment, perfect knowledge means knowing all 656 cue values (i.e., the values of 82 cities on eight cues). Individuals with incomplete knowledge have only a portion of the total set of cue values at their disposal. To avoid selecting implausible knowledge parameters, we reanalysed a previous study in which we had asked 19 students (at the University of Munich) to recall their values on each of the eight cues for each of the 82 cities. On average, participants recalled 89% of all cue values ($SD = 10\%$). Amongst participants, the amount of knowledge ranged from 70 to 100% of all cue values. Informed by this analysis, we implemented different amounts of knowledge, ranging from 30 to 100% (see Table 2), thus extending the range beyond the (empirical) lower bound to examine how a small amount of knowledge affects hindsight bias.

In addition to varying the sheer amount, we also varied knowledge accuracy. Knowledge accuracy is simply the percentage of cue values (among the known ones) that are correct. Perfectly accurate knowledge means that every cue value a "person" has stored is correct. Individuals with less than perfectly accurate knowledge need to rely on (some) cue values that are false. In the reanalysis of the Munich student sample, we also found that, on average, 86% of the known cue

TABLE 2
Parameters investigated in the simulations

<i>Parameter</i>	<i>Description</i>	<i>Implementation</i>	<i>Range</i>
Amount of knowledge	Percentage of cue values stored in long-term memory	Which of the cue values were designated to be unknown was randomly determined across all cue values	Complete knowledge: 100% (656 values) Incomplete knowledge: 30% (167 values), 40%, 50%, ..., 90%
Accuracy of knowledge	Percentage of false cues values among the known cue values	Positive and negative cue values were randomly replaced by the opposite cue values	Complete accurate knowledge: 0% false cue values Partly false knowledge: 5%, 10%, 15%, ..., 35%
Updating probability of cue values	Probability with which unknown cue values are updated in accordance with outcome information	Unknown cue values were randomly replaced with positive or negative cue values (in accordance with outcome information)	Updating probability: 0.05, 0.1 (default value), 0.15, 0.2

values were accurate ($SD = 6\%$) and that the proportion of false cue values ranged between 5 and 25%. Informed by these values, we implemented degrees of false knowledge (of cue values) ranging from 0 to 35% (see Table 2).

Finally, according to the RAFT model, elusive and missing cue values can be updated by feedback. For the sake of simplicity, we assumed that cue values (i.e., positive, negative, and unknown) are veridically retrieved from long-term memory (thus ignoring the fact that cue values at Time 1 and 3 may differ simply because memory retrieval is typically not completely reliable), and that updating would only occur (with some probability) if cue values were unknown in long-term memory. The updating probability for unknown cue values was set to range from 5 to 20% (Table 2). It is thus within close range of the rate that we observed in Hoffrage et al.'s (2000) Study 2.

Across all simulations, we examined 8 (amount of knowledge: 30%, 40%, ..., 90%, or 100% of the cue values were known) \times 8 (knowledge accuracy: 0%, 5%, ..., 30, or 35% of known cue values were false) combinations. Within each of the 64 combinations, we simulated 100 "individuals", who differed randomly from one another in the particular cue values that were false or missing. Similarly, for each individual (and for each of 100 runs of each individual) we randomly determined which of the missing cue values would be updated as well as the set of 41 city pairs (out of the 82 cities) to which an "individual" responded. The results were averaged across individuals and runs.

Let us also highlight that we did not pre-determine the same hierarchy of cues for all

simulated individuals. Rather, we calculated cue validities on the basis of the existing knowledge of cue values for each individual (at Time 1). Thus, the cue order of a person who knows, say, 50% of all cue values could be quite different from the cue order of another person with more or less knowledge and even different from the cue order of another person with the same amount of knowledge (depending on the distribution of known and unknown cue values).

In Simulation 1, we assumed knowledge to be completely accurate and we kept the updating probability constant (i.e., default value of 0.10; see Table 2). Here we examined how an increasing amount of cue-value knowledge affects hindsight bias. In Simulation 2, we replicated the knowledge simulation but explored the impact of an alternative inference mechanism on the hindsight bias. In Simulation 3, we analysed how an increasing proportion of false knowledge affects hindsight bias. The general policy we followed throughout the simulations was to vary the specific parameter under consideration, while keeping all others constant or averaging across them (i.e., *ceteris paribus* policy). Although this procedure restricted our ability to explore intricate interactions among parameters, it helped us to focus on the main effects, thus increasing the results' transparency and comprehensibility.

In empirical studies, hindsight bias occurs when the recalled choices are more accurate than the original choices. To control for other factors that may also affect the accuracy of the inferences at Time 3, the observed increase in accuracy is typically appraised against changes in accuracy in a control condition in which no outcome infor-

mation was provided, and in which typically no systematic differences occur. Similarly, in our simulations, we expected no systematic differences in the control conditions to occur, and, in fact, we found none. Therefore, we could simplify the hindsight bias measure: Specifically, we computed the difference between the average percentages of correct inferences at Time 3 (henceforth, *hindsight accuracy*) and Time 1 (henceforth, *foresight accuracy*). To control for different levels of foresight accuracy (thus avoiding the problem of a ceiling effect), hindsight bias was expressed as the ratio of this difference and the maximum difference between hindsight and foresight accuracy. Specifically, hindsight bias equals $100 * (\text{hindsight accuracy} - \text{foresight accuracy}) / (100 - \text{foresight accuracy})$.

SIMULATION 1: HOW DOES AMOUNT OF KNOWLEDGE AFFECT HINDSIGHT BIAS?

In their meta-analysis of hindsight bias studies, Christensen-Szalanski and Willham (1991) observed that more expertise in the domain from which the tasks were sampled yielded less hindsight bias. In the first simulation, we operationalised expertise in terms of the amount of knowledge about cue values at Time 1 (all of which are correct). Does access to more cue values guard against hindsight bias? Figure 2 depicts foresight accuracy, hindsight accuracy, and hindsight bias as a function of the amount of foresight knowledge. Clearly, more knowledge reduces hindsight bias. Specifically, the size of hindsight bias turns out to be a linear function of the amount of knowledge. That is, the more foresight know-

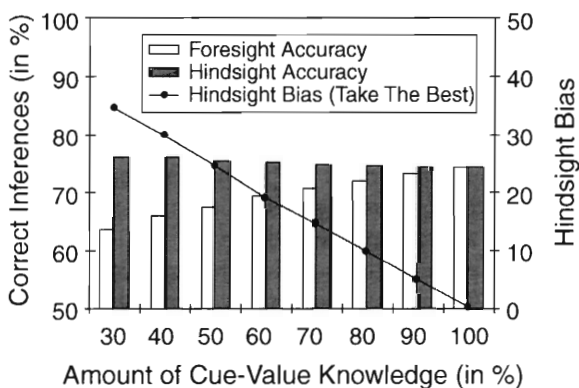


Figure 2. Foresight and hindsight accuracy (i.e., amount of correct inferences at Time 1 and Time 3) achieved by Take The Best, and hindsight bias as a function of amount of knowledge.

ledge a person has, the smaller the bias he or she tends to exhibit. For instance, a person who knows only 30% of all cue values displays a hindsight bias that is about seven times as large as a person who knows 90% of all cue values—34% versus 5%.

Which mechanism underlies the compelling relationship between cue-value knowledge and amount of hindsight bias? To explore this question, let us focus on two of the eight degrees of knowledge (at Time 1), namely, *scant* knowledge and *ample* knowledge (i.e., 30% versus 90% of cue values are known). In addition, let us introduce two new concepts, namely, *frugality* and *utilisation rate*. The former refers to the average number of cues a heuristic needs to look up before it can arrive at a decision (e.g., as to which of two objects scores higher on a criterion value); the latter refers to the percentage of choices in the total set that are determined by a given cue. Both concepts will help to explain why less foresight knowledge yields more hindsight bias.

According to the RAFT model, outcome knowledge can be used to infer missing cue values. Therefore, some of the cues that did not discriminate at Time 1 will discriminate at Time 3.³ Consequently, Take The Best becomes more *frugal* at Time 3, that is, it needs to look up fewer cues before it arrives at a decision. Of course, such savings in information search will be larger when the decision maker only has scant knowledge to begin with—with ample knowledge, Take The Best already tends to be frugal. Indeed, while the average number of cues looked up decreases from 5.7 (Time 1) to 4.6 (Time 3) for scant knowledge, it remains the same for ample knowledge (3.5 at Time 1 versus 3.5 at Time 3). This fact is illustrated in Figure 3, which shows the utilisation rate of all eight cues and the “guessing” cue at Time 1 and Time 3, respectively. For scant knowledge at Time 1, the three highest-ranked cues were used in about 25% of all inferences; at Time 3 they were used in more than 40% of all inferences. Finally, the guessing rate dropped from about 33% to 20%. In contrast, in the case of ample knowledge, the cues’ utilisation rates and the guessing rate remained almost unchanged.

The changes in the heuristic’s frugality and the cues’ utilisation rates point to a mechanism that

³ A cue is said to discriminate between two objects, *a* and *b*, if one object has a positive cue value and the other does not (i.e., it either has a negative value or is unknown) for this cue. Cue updating can turn a cue that does not discriminate between object *a* and *b* into one that does discriminate by updating one or both missing values of the objects for the cue.

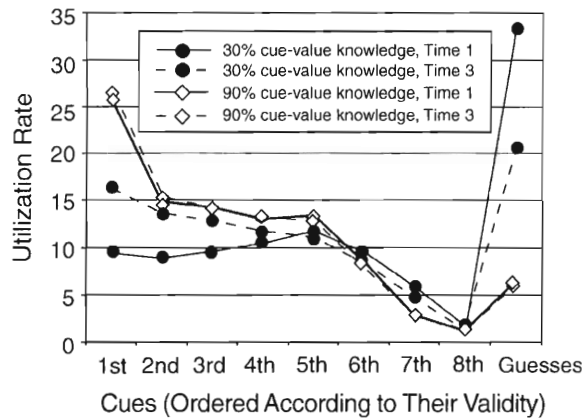


Figure 3. Utilisation rate at Time 1 and 3 for each of the eight cues, rank ordered according to their validity (see Table 3) as a function of amount of knowledge.

accounts for scant foresight knowledge yielding a larger hindsight bias: Scant knowledge leaves more room for updating to affect the process of reconstruction than ample knowledge does. Why? First and foremost, the fewer cue values a person knows at Time 1, the more cues the Take The Best heuristic has to look up to make an inference (e.g., with scant knowledge 5.7 cues). If the choice at Time 1 has been made, for instance, by the sixth-ranked cue, then each higher-ranked *and* updated cue now has a chance to be the one reason that the Take The Best heuristic uses to determine the choice at Time 3. In addition, less knowledge also means more missing cue values. Therefore, if a cue is retrieved at Time 3, the likelihood that this cue includes an updated value increases with less knowledge. Finally, if the Take The Best heuristic uses an updated cue to arrive at the choice, then this choice will necessarily be correct (in terms of the criterion value), because updating cue values is contingent on outcome knowledge. A correct choice at Time 3, in turn, can yield hindsight bias (if the choice at Time 1 was incorrect).

To conclude, to the extent that the amount of knowledge of cue values reflects expertise, the RAFT model can account for the expertise effect reported in Christensen-Szalanski and Willham (1991). As Simulation 1 has shown, less foresight knowledge makes the veridical reconstruction of the original choice less likely. Within the RAFT model, one can delineate a mechanism that accounts for this effect. Less knowledge leaves more “room” for updating to affect the reconstruction process at Time 3. This explanation gives rise to an interesting question: Would this effect of foresight knowledge on hindsight bias also occur if

cues were processed in a completely different way? The next simulation examines the question of whether the effect of foresight knowledge arises specifically from the application of the Take The Best heuristic.

SIMULATION 2: IS THE EFFECT OF KNOWLEDGE ON HINDSIGHT BIAS ROBUST ACROSS DIFFERENT HEURISTICS?

How do experts make inferences about uncertain aspects of the world? As mentioned earlier, some researchers (e.g., Shanteau, 1992) have suggested that experts should bring all their relevant knowledge to bear. This idea reflects the widespread assumption in cognitive psychology that more information yields better performance (see Hertwig & Todd, in press). The research programme on fast and frugal heuristics (Gigerenzer et al., 1999b) has thoroughly challenged this ubiquitous assumption and some experts have been shown to rely on just one or a few pieces of information (e.g., Green & Mehr, 1997). Yet, it may be the case that those experts who attempt to look up all available information and to integrate it into one score exhibit a different relationship between foresight knowledge and hindsight bias than those who employ a one-reason decision-making strategy such as the Take The Best heuristic.

There is a plausible reason why this might be the case. If a choice has initially been made on the basis of a set of cues rather than on one single cue, it may prove to be more robust towards slight changes in the updated knowledge state. The Take The Best heuristic, in contrast, may amplify the effects of updating since a single updated cue value can lead to the opposite choice. What strategy integrates multiple pieces of information while still being psychologically plausible? Robyn Dawes (e.g., 1979) suggested a compensatory strategy that does not overtax the processing capabilities of the human mind. This strategy, which Gigerenzer et al. (1999b) called *Dawes' rule*, is a linear strategy with unit weights that has been advocated as a good approximation of weighted linear models (Dawes, 1979; Einhorn & Hogarth, 1975). It simply adds up the number of positive cue values and subtracts the number of negative cue values (ignoring missing values) and thus is very different from the Take The Best heuristic: While both are fast (i.e., they do not involve much computation), Dawes' rule is far

from being frugal—it bases its choice on all available pieces of information. Table 3 illustrates the policy of Dawes’ rule. It also demonstrates how updating does or does not give way to hindsight bias, depending on the inference mechanism employed.

Does more knowledge involve less hindsight bias if it is processed by Dawes’ rule? To answer this question, we conducted a simulation that exactly replicated the previous one with the exception that the Take The Best heuristic was replaced by Dawes’ rule. Before attending to hindsight bias, let us first consider the accuracy of the Dawes’ rule prior to updating. On average (i.e., across all degrees of knowledge of cue values), Dawes’ rule made 69.3% correct inferences at Time 1. The surprising observation is that Take The Best matched this performance: It scored 69.7% correct inferences although it used only about 44% of the information available to Dawes’ rule (for similar results see Gigerenzer, Czerlinski, & Martignon, 1999a).

Despite being much more frugal than Dawes’ rule, Take The Best does not pay a price in lost

accuracy. But does it pay a price in terms of a larger hindsight bias? The results (Figure 4) show that it does not: Reconstructing one’s original choice on the basis of Dawes’ rule does not attenuate hindsight bias. On the contrary, hindsight bias for Dawes’ rule is slightly but consistently larger than for the Take The Best heuristic. Across all degrees of knowledge, the average hindsight bias was 18.4% and 17%, respectively. Except for this slight difference in hindsight bias, the results in Figures 2 and 4 coincide: Again, the amount of hindsight bias is a linear function of the amount of knowledge. For instance, the hindsight bias of a person who knows only 30% of all cue values is more than seven times the size of that of a person who knows 90% of all cue values (38.2% versus 5.2%).

Thus, Dawes’ rule reproduces the same relation between foresight knowledge and hindsight bias as the Take The Best heuristic. The underlying mechanism, however, must be a different one as Dawes’ rule processes cues differently. To identify the mechanism, let us again focus on *scant* (i.e., 30% of cue values are known) versus *ample* (i.e., 90% of cue values are known) knowledge. To reiterate, the policy of Dawes’ rule is simply to add up the number of positive cue values and subtract the number of negative cue values (ignoring missing values) for each alternative. The decision rule is to choose the alternative with the higher score (the “winner”). To understand why less foresight knowledge yields less hindsight bias, it is instructive to consider the average difference between the “winner” and “loser” scores. With scant knowledge, the average difference is much narrower than with ample knowledge, namely, 2.0

TABLE 3
Choices at Time 1 and 3 as a function of knowledge states and inference mechanism (Dawes’ rule versus the Take The Best heuristic) for the German cities task

Cues	Time 1		Time 3	
	Essen	Bremen	Essen	Bremen
Exposition site	?	-	+	-
Soccer team	-	?	-	?
Intercity train	?	+	?	+
State capital	-	+	-	+
Licence plate	-	-	-	-
University	+	+	+	+
Industrial belt	?	?	?	?
East Germany	-	?	-	?
$\Sigma(+ \text{ values}) - \Sigma(- \text{ values})$	-3	+1	-2	+1
Choice (Dawes’ rule)	“Bremen”		“Bremen”	
Choice (Take The Best)	“Bremen”		“Essen”	

The probabilistic mental model contains eight cues ranked according to their ecological validity. Cue values are positive (+) or negative (-); missing knowledge is indicated by question marks. To infer whether Essen is larger than Bremen, Dawes’ rule adds up the number of positive cue values and subtracts the negative cue values. The Take The Best heuristic, in contrast, looks up only the cue values in the shaded areas. At Time 3, the cue value of the exposition site cue for Essen shifts towards feedback (i.e., Essen is larger than Bremen). As a consequence, the Take The Best heuristic predicts hindsight bias on the level of choice, while Dawes’ rule predicts veridical recollection.

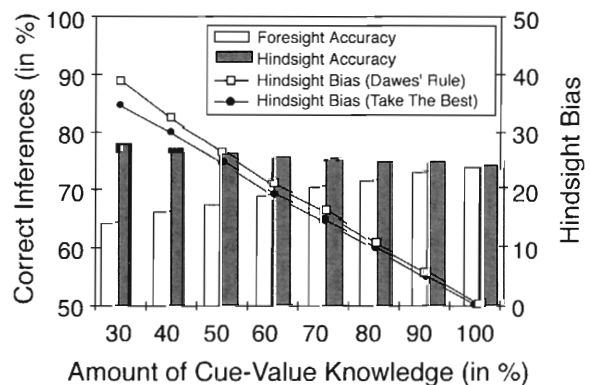


Figure 4. Foresight and hindsight accuracy (i.e., amount of correct inferences at Time 1 and Time 3) achieved by Dawes’ rule (and Take The Best), and hindsight bias as a function of amount of knowledge.

versus 3.7. This has an important implication: Updating one or only a few missing cue values can overturn the original choice much more easily for scant knowledge than for ample knowledge. As a consequence, hindsight bias is more likely to occur with scant knowledge.

To conclude, Dawes' rule, a linear strategy with unit weights, produces the same finding we observed earlier: More foresight knowledge results in less hindsight bias. In fact, both the non-compensatory Take The Best heuristic and the compensatory Dawes' rule yielded mostly identical results, with the latter exhibiting a slightly larger hindsight bias than the former. Thus, Christensen-Szalanski and Willham's (1991) expertise effect can be accounted for by two completely different processing policies. Although this finding may seem surprising, it is consistent with the fact that these authors observed the expertise effect across a wide range of studies, and thus very likely across different inference strategies.

SIMULATION 3: HOW DOES ACCURACY OF KNOWLEDGE AFFECT HINDSIGHT BIAS?

Up to this point in our investigation, we assumed knowledge of cue values to be completely accurate. Knowledge, however, may not always be accurate and, possibly, experts' knowledge may be more exact than that of novices. In the final simulation, we turned to the relationship between knowledge accuracy and hindsight bias. Specifically, we replicated Simulation 1 (using the Take The Best heuristic and assuming an updating probability of .10) and introduced one additional variable, namely, the accuracy of knowledge. Specifically, we implemented eight different degrees of false knowledge, ranging from 0 to 35% incorrect cue values (see Table 2). Figure 5 depicts the Take The Best heuristic's foresight accuracy, hindsight accuracy, and hindsight bias as a function of knowledge accuracy (the results are averaged across eight different amounts of knowledge, ranging from 100% knowledge to 30% knowledge).

Before attending to hindsight bias, let us first consider the effect of false knowledge on judgment accuracy at Time 1. Not surprisingly, it reduces the percentage of correct inferences in foresight. For illustration, compare the percentage of correct inferences when 5% and 35% of all cue

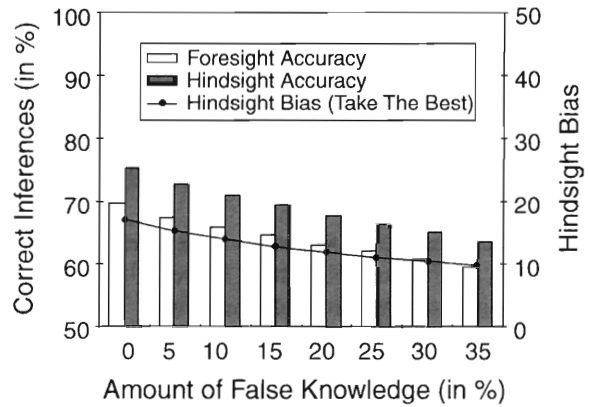


Figure 5. Foresight and hindsight accuracy (i.e., amount of correct inferences at Time 1 and Time 3) achieved by Take The Best, and hindsight bias as a function of amount of false knowledge.

values are incorrect. Henceforth, we refer to these two states as (relatively) *veridical* and *flawed* knowledge, respectively. While veridical knowledge yields about 67% correct inferences, drawing inferences from flawed knowledge brings inference accuracy down to about 60%. Does flawed knowledge also result in a larger hindsight bias? The surprising answer is no. As Figure 5 shows, the more flawed foresight knowledge is, the smaller the size of hindsight bias. For instance, a person whose knowledge is veridical displays a hindsight bias that is almost one and a half times larger than the bias of a person whose knowledge is flawed—15.2 versus 9.7%.

Which mechanism might underlie this counterintuitive effect of false knowledge on hindsight bias? One candidate explanation concerns the impact of false knowledge on the heuristic's frugality. Although the insertion of incorrect cue values reduces foresight accuracy, it also reduces the number of cues that the Take The Best heuristic needs to look up before it can reach a decision. Specifically, with veridical knowledge the heuristic, on average, looks up 3.9 cues (of 9 cues including the guessing cue). In contrast, with flawed knowledge it only needs to look up 2.9 cues. Adopting the same logic as before (in the case of scant knowledge, see Simulation 1), this difference in the heuristic's frugality can account for why less accurate knowledge yields a smaller hindsight bias. At Time 3 hindsight bias can only occur if the Take The Best heuristic encounters a cue that has been updated and now discriminates between the two objects *before* it reaches the cue that initially discriminated at Time 1 (see Table 1 for an example). With flawed knowledge, how-

ever, the chances of coming across such a cue are smaller than with veridical knowledge. The reason is that prior to the cue that discriminated at Time 1 there are only, on average, 2.9 (as opposed to 3.9) cues for the effect of updating to occur. As a consequence, flawed knowledge admits less “room” for hindsight bias.

This candidate explanation raises the question of how is it possible that flawed knowledge increases the frugality of the Take The Best heuristic. To answer this question, one needs to analyse the informational structure of the knowledge environment. In environments in which the distribution of cue values is skewed such that “negative” cue values outnumber “positive” ones (e.g., in the German city environment, 71% of the cue values are negative), *random* insertion of false knowledge will reduce the asymmetry in the number of positive and negative cue values. Consequently, the cues’ discrimination rate increases, thus reducing the number of cues that need to be looked up. This fact has an interesting implication: If incorrect cue values were systematically rather than randomly distributed (e.g., if positive cue values were falsely considered to be negative but not vice versa) or if the frequency of positive and negative cue values was not as skewed, flawed knowledge might affect the reconstruction process at Time 3 in a rather different way.

Finally, it is noteworthy that flawed knowledge not only increases the discrimination rate, but it can also decrease the validity of cues below 50%. As a consequence, those “invalid” cues will be eliminated from consideration, thus providing even fewer cues for the effect of updating to occur. To test the extent to which both a higher discrimination rate and a smaller set of cues might contribute to the “debiasing” effect of false knowledge, we ran a modified version of Simulation 3. Here false knowledge was inserted such that the cues’ discrimination rate remained the same, and “invalid” cues (validity < 50%) were not eliminated from consideration (i.e., the number of available cues remained constant). In this simulation, the “debiasing” effect of false knowledge was reduced by half, thus indicating that both factors can provide a partial explanation for the surprising effect of false knowledge.

DISCUSSION

Hindsight bias may occur because of the attempt to *reconstruct* one’s original judgement (Hawkins & Hastie, 1990). The RAFT model (Hoffrage &

Hertwig, 1999) was proposed to account for this process of reconstruction. The model’s core assumption is that outcome knowledge (e.g., Bush won the presidential election) can be used to update the probabilistic knowledge from which we draw inferences. Thus, hindsight bias is not so much viewed as a bias but as a consequence of learning by feedback. By being explicit about the processes, the RAFT model opens the door to detailed analyses of the make-up of hindsight bias.

The present simulations rendered three major results: Consistent with Christensen-Szalanski and Willham’s (1991) observation, we found that more foresight knowledge results in a smaller hindsight bias (Simulation 1; see Figure 2). This relation appears to be independent of the inference strategy used to process a person’s knowledge of cues and cue values: Both a compensatory and a non-compensatory inference strategy yield comparable results, with the former showing a slightly but consistently larger hindsight bias (Simulation 2; see Figure 4). Finally, we observed that more flawed foresight knowledge led to a smaller hindsight bias (Simulation 3; see Figure 5).

Our investigations confirm the utility of developing and testing precise process models of hindsight bias. In addition, they provide additional empirical support for the RAFT model in so far as the model can predict and account for the well-known expertise effect. Finally, our results are also of importance as an existence proof that different cognitive strategies can yield similar predictions, thus suggesting the possibility that a particular judgement or memory phenomenon may be robust across a variety of different processing strategies. Next, we explore ways of testing the predictions that emerged from our simulations, discuss alternative accounts of the expertise effect, and, more generally, evaluate the role of hindsight bias.

A first test of the expertise prediction and the issue of policy capturing

Consistent with Christensen-Szalanski and Willham’s (1991) finding, the RAFT model predicts that more knowledge (of cue values) leads to less hindsight bias. This prediction can be evaluated in studies that manipulate or keep track of people’s cue knowledge. To this end, we reanalysed a previous study (Hoffrage et al., 2000, Study 2) that recorded such knowledge. In this study, participants were asked to assume the role of a health-insurance company employee. Assuming this role,

they learned some facts about a dozen fictional individuals who had submitted applications to purchase health insurance. These facts referred to the applicants' health status and included information about the presence or absence of three risk factors (parents' hypertension, excess of weight, and smoking).

Participants were then instructed that the cost of a person's health insurance depends on the presence or absence of health-risk factors: Unfortunately, applicants had forgotten to indicate their values for a key risk factor, namely, (high) blood pressure. Therefore, the participants' task was, among other things, to decide for pairs of two applicants "Which of them has higher blood pressure?". To be able to make this choice, participants also learned that parents' hypertension, excess of weight, and smoking were cues for high blood pressure, and they were told the validities of these cues (80%, 70%, and 60%). Either before or after they gave their response (Time 1), they recalled the values on all three cues they had learned—these values represent their amount of foresight knowledge. In the second session, participants received the correct answer (in one of three conditions; for the other two see Hoffrage et al., 2000) and then were asked to recall their original choice.

Was hindsight bias in this study larger for people with less foresight knowledge, as predicted by the RAFT model? To answer this question, we computed the correlation between participants' amount of foresight knowledge (here focusing only on correct cue values) and their respective hindsight bias. Consistent with the RAFT model's prediction, we found that the amount of foresight knowledge and the magnitude of hindsight bias are negatively correlated in the feedback condition ($r = -.15$). The same correlation was positive in the no-feedback condition ($r = .11$), and the effect size of the difference between the correlations in the two conditions amounted to $q = 0.26$ (which corresponds to a medium effect size; see Cohen, 1988, p. 115). In short, the empirical results confirm the results obtained in Simulation 1.

Before we turn to alternative accounts of the expertise effect, let us briefly discuss a methodological implication of the surprising finding that compensatory and non-compensatory processing strategies of cues produced similar sizes of hindsight bias (Simulation 2). In our view, this finding should not discourage experimenters from trying to find out which inference strategy people use.

True, on an aggregate level, the Take The Best heuristic and Dawes' rule yielded almost identical amounts of hindsight bias. On the level of individual judgements, however, they produced diverging judgements. Experimenters can take advantage of this fact. Specifically, if a study's aim is to identify which inference policy people use, the experimenter can compose a set of items that amplifies the differences between the strategies. For instance we observed that, averaged across all amounts of foresight knowledge and all proportions of false knowledge, the Take The Best heuristic and Dawes' rule yielded identical outcomes (i.e., hindsight bias, reversed hindsight bias, veridical recollection) in about two-thirds of all items. Therefore, an experimenter who tries to capture people's judgemental policies could sample from those choice tasks that discriminate between the policies instead of drawing a representative sample of tasks (see Dhimi, Hertwig, & Hoffrage, 2003, for a discussion of this topic).

Alternative accounts

Next to the RAFT's model account of the expertise effect, we can think of two alternative explanations for it—one follows from the RAFT model; the other refers to a process whose existence the RAFT model acknowledges (see Figure 1 in Hoffrage et al., 2000) but does not aim to model, namely, direct recall of the original judgement. We begin with the latter.

Better episodic memory. Experts are not only more knowledgeable than novices, they also seem to have the ability to learn mechanisms that afford rapid storage of information in long-term memory (e.g., Ericsson & Chase, 1982; Ericsson & Kintsch, 1995). In the context of hindsight bias research, this ability may enable experts to reliably recall the specific decision episode in which they determined which of two objects (for instance, Munich or Hamburg) is larger. In other words, experts may have a better *episodic* memory (for a review of research on episodic memory, see Tulving, 2002) of the context in which they arrived at their choice, thus rendering reconstruction processes unnecessary. If so, experts' average amount of hindsight bias would be smaller than that of novices. This explanation can be tested by using the number of veridical recollections (i.e., judgement at Time 1 = recalled judgement at Time 3) as an estimate of cases of direct recall. We know of

one hindsight bias study that kept track of the frequency of veridical recollections as a function of expertise. Pohl (1992) asked second-year psychology students (“novices”) and researchers (“experts”) to estimate numerical figures, such as “When did J. J. Gibson publish the book *The ecological approach to visual perception?*”. He then provided them with the correct answer and asked them to recall their original estimate. Consistent with the speculation that experts may have better episodic memory, Pohl observed that the experts’ rate of veridical reproduction of previous estimates was almost twice as high as that for novices, namely, 33% versus 19%, respectively.

These findings relate to another process account of hindsight bias—the SARA model (Pohl et al., 2003-this issue). This model conceptualises knowledge in terms of “images” or information units that are used when people estimate numerical values such as Goethe’s age at death. An image is, for instance, the (subjective) quantitative knowledge of the average life expectancy of Goethe’s contemporaries. In their simulations, Pohl et al. (2003-this issue) observed that the availability of more images brings the proportion of veridical recollections down because, so runs one plausible suggestion, it is more difficult to find the original estimate among a larger (as opposed to a smaller) set of images. In addition, if one assumed that experts have more images at their disposal, then the SARA model would predict that experts’ episodic memory (for their original judgements) should be worse than that of novices. However, the authors of the SARA model question this prediction, and suggest instead that expertise may be more related to the quality of images rather than their quantity. Consistent with this assumption, Pohl (1992) found that the original judgements of experts as compared to those of novices were significantly closer to the solution. More generally, however, the SARA model and the RAFT model concur in predicting that more knowledge (in terms of either more images or cues values) will reduce the size of the hindsight bias (see Pohl et al., 2003-this issue).

More updating. The RAFT model, however, offers yet another candidate reason beyond more knowledge for why experts’ hindsight bias is smaller: the possibility that the rate of updating is higher for novices than for experts. Would a higher updating rate yield more hindsight bias? To address this question, we varied the size of the

updating probability, and found that more updating produces in fact a larger hindsight bias. Figure 6 shows the amount of hindsight bias for four different updating probabilities: .05, .10, .15, and .20. Clearly, if novices’ knowledge were updated at a higher rate than experts’ knowledge, then, all else being equal, the former would show a larger average hindsight bias. The results in Figure 6 also show that the impact of the updating rate on size of hindsight bias increases with increasing lack of knowledge. But is novices’ knowledge indeed more thoroughly updated than experts’ knowledge? As we discuss in the following section, this is just one of the novel and, as we believe, exciting questions that emerge from the RAFT model.

New questions

Hindsight bias is not a simple, uniform phenomenon. It is likely to be shaped by processes of recollection and reconstruction (Erdfelder & Buchner, 1998) as well as by judgement phenomena such as the reiteration effect (Hertwig et al., 1997). The results of our simulations support the view of hindsight bias as the sum of multiple determinants. Specifically, we find that people’s hindsight bias is, all else being equal, a function of their amount of foresight knowledge, the proportion of false knowledge, and the degree to which feedback updates their missing knowledge. Thus, hindsight bias is a nontrivial composite of effects that take different directions: While the effects of knowledge accuracy and updating go in the same direction (i.e., hindsight bias increases with

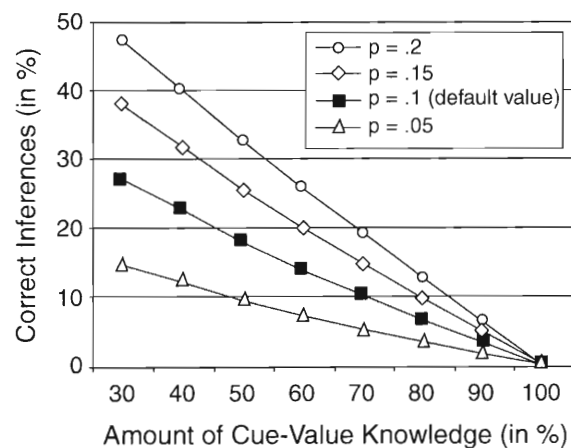


Figure 6. Hindsight bias as a function of amount of knowledge and the probability of updating ($p = .05$, $p = .10$, $p = .15$, $p = .20$).

increasing accuracy and updating probability), the effect of the amount of knowledge goes in the opposite direction (i.e., hindsight bias shrinks with increasing amount of knowledge)—at least in the present knowledge domain.

To the extent that hindsight bias reflects the combination of those three effects, their individual sizes will determine the size of hindsight bias. In the simulations, we observed that the “debiasing” effect of more knowledge was larger than the “biasing” effect of more accurate knowledge.⁴ In fact, a larger “debiasing” effect of knowledge amount could also explain why, according to Christensen-Szalanski and Willham’s (1991) meta-analysis, experts exhibit less hindsight bias than novices although the former may enjoy both ample and accurate knowledge.

However, to understand better how much influence each of the three effects exerts on hindsight bias, we need to know not only the range of parameter values but also how they are distributed. Consider updating probability as an example. In our simulations, we assumed updating to be constant across levels of knowledge. But this need not be so. Alternatively, the updating may linearly (or exponentially) increase or decrease as a function of knowledge, or updating may follow a (inverse) U-shaped function. Thus, by focusing on the updating process, the RAFT model raises new questions—for instance, do novices’ updating probabilities exceed those of experts, or vice versa—that, to the best of our knowledge, have not or only rarely been raised before. To the extent that knowledge updating represents a general learning mechanism that enables and supports people’s inductive inferences, the answers to these questions will have relevance beyond the limits of hindsight bias research.

⁴This difference in effect sizes is not due to the fact that we examined a larger parameter space for knowledge amount (30–100% known cue values) than for knowledge accuracy (0–35% false cue values). To test for this possibility we simulated the effect size for knowledge amount and knowledge accuracy assuming the same parameter space. The hindsight bias for 100% and 65% amount of knowledge (averaged across all states of knowledge accuracy) amounted to 0% and 12%, respectively. In contrast, the hindsight bias for 0% and 35% false cue knowledge (averaged across all states of known cue values) was 8.3% and 4.4% respectively. In addition, we simulated an empirically derived parameter space (derived from the aforementioned Munich experiment). Here we found that the “debiasing” effect of knowledge amount strongly exceeded the “biasing” effect of knowledge accuracy. Specifically, while more knowledge decreased by hindsight bias by 6.5 percentage points, more accurate knowledge increased the hindsight bias by 0.7 percentage points.

Another novel issue concerns the question of how robust the RAFT model’s predictions are across different environments. For illustration, consider the following property of the German city environment. In it, negative cue values outnumber positive cue values (71% vs 29%, respectively). Does this ecological property affect the amount of hindsight bias predicted by the RAFT model? To examine this question, we conducted another simulation in which, briefly sketched, we artificially created environments that mimicked the German city environment (i.e., number of objects, number of cues, and cue validities), except that positive cue values now outnumbered the negative ones (70% vs 30%, respectively); as in Simulation 1, amount of knowledge ranged from 30 to 100% and updating probability was .10. Compared to the results of Simulation 1, we now found the same linear increase of the hindsight bias as a function of more knowledge (see Figure 2); however, the amount of hindsight bias was smaller, and ranged from 0% to 22% (compared to 0% and 34% in Simulation 1).

Why is this? This difference in the amount of hindsight bias is due to one of Take The Best’s key building blocks, its stopping rule. The reader familiar with the Take The Best heuristic presented in Hoffrage et al. (2000) will have noticed that in our simulations we simplified the stopping rule. Here we employed a *lenient* stopping rule that terminates search if one object has a positive cue value and the other does not. Previously, we employed a *strict* stopping rule that terminated search only when one object had a positive (larger) value and the other a negative (smaller) one. This change in the heuristic’s architecture was made to achieve consistency with other applications of the Take The Best heuristic (e.g., Gigerenzer & Goldstein, 1999). Moreover, the lenient stopping rule with its “positive bias” is ecologically rational for those environments in which negative cue values outnumber positive ones, and thus unknown values are most likely negative ones (Gigerenzer & Goldstein, 1999). As the new simulation showed, the lenient stopping rule leads to a smaller hindsight bias once the environment exhibits a “negative” bias. To test whether this change is due to the nature of the stopping rule, we also implemented the strict stopping rule in the artificial environment (see above), and found that with this rule the amount of hindsight bias is independent of the ratio of positive and negative cue values.

To conclude, these novel results indicate that environmental structures may interact with the building blocks of heuristics. Given this interaction, we can now ask and examine the question of whether or not the empirically observed hindsight bias in reality changes as a function of environmental structures. Depending on the answer, we could then make inferences about the building blocks of people's heuristics, such as the nature of their stopping rules—strict or lenient.

Hindsight bias: Wagering on the future

Typically, hindsight bias has been interpreted as detrimental to people's future predictive abilities. Take Fischhoff's view, for instance. His early experimental studies carved out this new topic for memory researchers. He stressed that hindsight bias is not only robust and difficult to eliminate (Fischhoff, 1982a), but also has potentially harmful consequences:

When we attempt to understand past events, we implicitly test the hypotheses or rules we use both to interpret and to anticipate the world around us. If, in hindsight, we systematically underestimate the surprises that the past held and holds for us, we are subjecting those hypotheses to inordinately weak tests and, presumably, finding little reason to change them. *Thus, the very outcome knowledge which gives us the feeling that we understand what the past was all about may prevent us from learning anything from it.* (Fischhoff, 1982b, p. 343, emphasis added).

Similarly, Bukszar and Connolly (1988) concluded that, "those findings [of the hindsight bias] raise serious questions about the ability of humans to learn from experience. . . . [I]n retrospect, people see the world as unfolding inevitably toward the present. Outcomes fail to surprise" (p. 630).

We advocate an alternative view (Hoffrage & Hertwig, 1999; see also Roese & Olson, 1996). Although we do not deny that hindsight bias can have tangible consequences, we interpret it as a by-product of an efficient memory system. Updating is an important adaptive process because, among other reasons, "it is too expensive to maintain access to an unbounded number of items" (Anderson & Schooler, 1991, p. 396). A stockpile of memories (e.g., the memories of previous judgements, choices, decisions) would interfere with the only information that is relevant

right now. In this sense, forgetting in conjunction with updating information may be necessary for memory to maintain its function. It prevents us from using old and possibly outdated information (see Bjork & Bjork, 1988; Ginzburg, Janson, & Ferson, 1996).

Given this view, does hindsight bias prevent us from learning from the past as suggested by Fischhoff (1982b)? No, on the contrary, the very existence of bias suggests that we learn from the past. In the RAFT model, updating—that is, learning—occurs on the level of the imperfect cues from which we draw inferences. This learning is possible because of what Egon Brunswik (1952) called "vicarious functioning". If information on one cue is not available, another can replace this cue. Further, it is not only cues that are interchangeable: For many cases, the possibility of drawing inferences from a cue to a criterion can be reversed. For instance, not only can the candidates' charisma tell us about their chances in the election contest, but also the reverse is true. Therefore, new information about the criterion can be used to update related knowledge in semantic memory—similar to the updating of outdated information in episodic memory (see Bjork & Bjork, 1988).

Updating is adaptive: It increases the coherence of our knowledge and the accuracy of our inferences. Consider the results reported in Figure 2: On average (across all knowledge states), the proportion of correct inferences increases from 64.1 in foresight to 68.8 in hindsight. Accuracy improves because decision makers have new knowledge (about cue values). This new knowledge, however, can not only be used to rejudge the initial choices but also to make inferences about novel choices for which no outcome information was provided. In fact, in the German city environment the test set of 41 choices represents only about 1% of all possible choices ($n = 3,321$). Does the updated knowledge also prove beneficial for the 99% novel inferences? It does. As Figure 7 shows, the percentage of correct inferences for the complete set of old and new pair comparisons *after* updating is higher than *prior to* it. That is, updating knowledge benefits future, novel choices.

Updating, however, is not a magical panacea. Although it increases the accuracy of our inferences, it may also account for mental models of the environment that are not completely correct. That is, updating can generate cue values that conform to the outcome of an event but never-

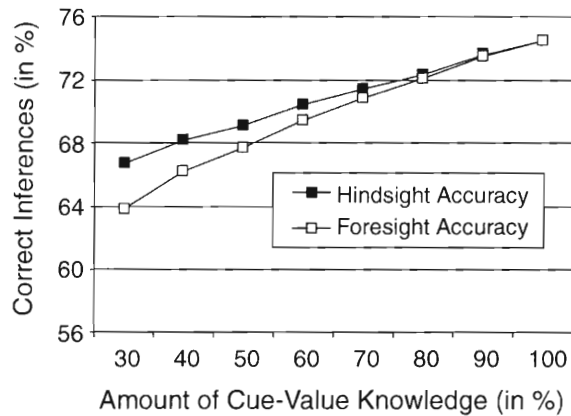


Figure 7. Increase of Take The Best's predictive accuracy after knowledge updating for the total choice set (assuming the same parameters as in Simulation 1: false knowledge = 0%, and updating probability = 0.1).

theless are false.⁵ Consider our previous example of US presidential elections. Among candidates' personal characteristics, their height has proved to be highly predictive: The taller candidate has won every presidential race since World War II except in the post-Watergate election of 1976 (Carter versus Ford), and in the notorious 2000 election. Imagine a voter who initially did not know that Gore is taller than Bush. After the election, the voter may have erroneously inferred that Bush is likely to be taller than Gore. Admittedly, updating by no means guarantees correct knowledge. However, we suggest that updating will typically result in correct knowledge. In fact, in a reanalysis of Hoffrage et al.'s (2000) Study 1 we found two-thirds of the updated information to be correct.

CONCLUSION

Winston Churchill was once asked what the desirable qualifications were for any young person who might wish to become a politician. He responded that "it is the ability to foretell what is going to happen tomorrow, next week, next month, and next year. And to have the ability afterwards to explain why it didn't happen" (cited in Buchanan, 2000, p. 185). The interesting insight from the present findings is that being able to

⁵Theoretically, false cue values may also cause subjective cue validities to diverge from ecological validities (based on environmental frequencies). How distorting the impact of a false cue value is, however, depends on factors such as the size of the sample of cue values from which the validities are calculated, and the degree to which false cue values cancel each other out.

explain in hindsight why something did or did not happen (in contrast to one's expectations) is the key to improving our ability to foretell what is going to happen in the future. In the same spirit, we shall point out that while less knowledgeable decision makers may be less apt to veridically reconstruct past judgements, they enjoy the benefit of learning more from new knowledge. In the trade-off between accurate memories of the past and accurate inferences in the future, they wager on the latter.

REFERENCES

- Anderson, J.R., & Schooler, L.J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396-408.
- Arkes, H.R., Wortmann, R.L., Saville, P.D., & Harkness, A.R. (1981). Hindsight bias among physicians: Weighing the likelihood of diagnoses. *Journal of Applied Psychology*, 66, 252-254.
- 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. I: Memory in everyday life* (pp. 283-288). New York: Wiley.
- Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory*, 11, 491-504.
- Bröder, A. (2000). 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*, 26, 1332-1346.
- Brunswik, E. (1952). The conceptual framework of psychology. *International Encyclopedia of Unified Science*. Chicago: University of Chicago Press.
- Buchanan, M. (2000). *Ubiquity*. London: Weidenfeld & Nicolson.
- Bukszar, E., & Connolly, T. (1988). Hindsight bias and strategic choice: Some problems in learning from experience. *Academy of Management Journal*, 31, 628-641.
- Campbell, J.E., & Mann, T.E. (1996). Forecasting the presidential election: What can we learn from the models? *The Brookings Review*, 14, 26-31.
- Chase, W.G., & Simon, H.A. (1973). Perception in chess. *Cognitive Psychology*, 4, 55-81.
- Chi, M.T.H., Feltovitch, P.J., & Glaser, R. (1981). Categorization of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Christensen-Szalanski, J.J.J., & Beach, L.R. (1984). The citation bias: Fad and fashion in the judgment and decision literature. *American Psychologist*, 39, 75-78.
- Christensen-Szalanski, J.J.J., & Willham, C.F. (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: Lawrence Erlbaum Associates Inc.

- Dawes, R.M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.
- de Groot, A.D. (1965). *Thought and choice in chess*. The Hague: Mouton (first edition in Dutch, 1946).
- Dhimi, M.K., Hertwig, R., & Hoffrage, U. (2003). *The role of representative design and sampling in an ecological approach to cognition*. Manuscript submitted for publication.
- Einhorn, H.J., & Hogarth, R.M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Decision Processes*, *13*, 171–192.
- 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.
- Ericsson, K.A., & Chase, W.G. (1982). Exceptional memory. *American Scientist*, *70*, 607–615.
- Ericsson, K.A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*, 211–245.
- Ericsson, K.A., & Staszewski, J.J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Hillsdale, NJ: Lawrence Erlbaum.
- Fischhoff, B. (1982a). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). Cambridge: Cambridge University Press.
- Fischhoff, B. (1982b). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 335–351). Cambridge: Cambridge University Press.
- Gigerenzer, G., Czerlinski, J., & Martignon, L. (1999a). How good are fast and frugal heuristics? In J. Shanteau, B.A. Mellers, & D.A. Schum (Eds.), *Decision science and technology: Reflections on the contributions of Ward Edwards* (pp. 81–103). Norwell, MA: Kluwer.
- Gigerenzer, G., & Goldstein, D.G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*, 650–669.
- Gigerenzer, G., & Goldstein, D.G. (1999). Betting on one good reason: The Take The Best heuristic. In G. Gigerenzer, P.M. Todd, & the ABC Research Group, *Simple heuristics that make us smart* (pp. 75–95). New York: Oxford University Press.
- Gigerenzer, G., Hoffrage, U., & Kleinböling, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*, 506–528.
- Gigerenzer, G., Todd, P.M., & the ABC Research Group. (1999b). *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.
- Green, L., & Mehr, D.R. (1997). What alters physicians' decisions to admit to the coronary care unit? *The Journal of Family Practice*, *45*, 219–226.
- Hawkins, S.A., & Hastie, R. (1990). Hindsight: Biased judgment of past events after the outcomes are known. *Psychological Bulletin*, *107*, 311–327.
- Heisenberg, W.K. (1971). *Physics and beyond: Encounters and conversations*. New York: Harper & Row.
- Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. *Psychological Review*, *104*, 194–202.
- Hertwig, R., & Hoffrage, U. (2001). Empirische Evidenz für einfache Heuristiken: Eine Antwort auf Bröder [Empirical evidence for simple heuristics: A response to Bröder]. *Psychologische Rundschau*, *52*, 162–165.
- Hertwig, R., & Todd, P.M. (in press). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment and decision making*. New York: Wiley.
- 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., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 566–581.
- Huntford, R. (1999). *The last place on earth*. New York: Modern Library.
- Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist*, *39*, 341–350.
- Larkin, J., McDermott, J., Simon, D.P., & Simon, H.A. (1980). Expert and novice performance in solving physics problems. *Science*, *208*, 1335–1342.
- Lichtman, A.J. (1996). *The keys to the White House, 1996*. Lanham, MD: Madison Books.
- Mazursky, D., & Ofir, C. (1990). "I could never have expected it to happen": The reversal of the hindsight bias. *Organizational Behavior and Human Decision Processes*, *46*, 20–33.
- Ofir, C., & Mazursky, D. (1997). Does a surprising outcome reinforce or reverse the hindsight bias? *Organizational Behavior and Human Decision Processes*, *69*, 51–57.
- Pezzo, M. (2003). Surprise, defence, or making sense: What removes hindsight bias? *Memory*, *11*, 421–441.
- Pohl, R.F. (1992). Der Rückschau-Fehler: Systematische Verfälschung der Erinnerung bei Experten und Novizen [Hindsight bias: Systematic distortions of the memory of experts and laymen]. *Kognitionswissenschaft*, *3*, 38–44.
- Pohl, R.F., Eisenhauer, M., & Hardt, O. (2003). SARA: A cognitive process model to simulate the anchoring effect and hindsight bias. *Memory*, *11*, 337–356.
- 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.
- Roese, N.J., & Olson, J.M. (1996). Counterfactuals, causal attributions, and the hindsight bias: A con-

- ceptual integration. *Journal of Experimental Social Psychology*, 32, 197–227.
- Schwarz, S., & Stahlberg, D. (2003). Strength of hindsight bias as a consequence of meta-cognitions. *Memory*, 11, 395–410.
- Shanteau, J. (1992). Competence in experts: The role of task characteristics. *Organizational Behavior & Human Decision Processes*, 53, 252–266.
- Simon, H.A., & Barenfeld, M. (1969). Information-processing analysis of perceptual processes in problem solving. *Psychological Review*, 76, 473–483.
- Slegers, D.W., Brake, G.L., & Doherty, M.E. (2000). Probabilistic mental models with continuous predictors. *Organizational Behavior and Human Decision Processes*, 81, 98–114.
- Solomon, S. (2001). *The coldest march: Scott's fatal Antarctic expedition*. New Haven, CT: Yale University Press.
- Stahlberg, D., & Maass, A. (1998). Hindsight bias: Impaired memory or biased reconstruction? In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 8, pp. 105–132). New York: Wiley.
- Synodinos, N.E. (1986). Hindsight distortion: "I knew-it-all along and I was sure about it." *Journal of Applied Social Psychology*, 16, 107–117.
- Tulving, E. (2002). Episodic memory: From mind to brain. *Annual Review of Psychology*, 53, 1–25.
- Wendt, D. (1993). No hindsight bias ("knew-it-all-along effect") during province parliament elections in Schleswig-Holstein in 1988 and 1992. *Zeitschrift für Sozialpsychologie*, 24, 273–279.
- Werth, L., & Strack, F. (2003). An inferential approach to the knew-it-all-along phenomenon. *Memory*, 11, 411–419.

