Selective activation as an explanation for hindsight bias

Markus Eisenhauer (markus.eisenhauer@psychol.uni-giessen.de)
FB 06 – Psychology, Justus-Liebig-University Gießen;
Otto-Behaghel-Str. 10, D-35394 Gießen, Germany

Rüdiger F. Pohl (ruediger.pohl@psychol.uni-giessen.de)
FB 06 – Psychology, Justus-Liebig-University Gießen;
Otto-Behaghel-Str. 10, D-35394 Gießen, Germany

Abstract

In hindsight, people often claim to have known more in foresight than they actually did. For example, the confidence for one of several possible outcomes is larger when it is known that this particular outcome occurred. A widespread explanation of hindsight bias assumes that the feedback serves as an anchor. How precisely this anchor takes effect and why it leads to a bias towards the anchor value has not been satisfactorily answered yet. One possible mechanism to explain hindsight bias assumes that the encoding of the feedback leads to a selective activation of the item-specific knowledge base. As a result, specific information units are strengthened and are thus more likely to be recalled when a person tries to reconstruct his or her original judgment. We tested the effect of selective activation in two hindsight experiments. The results showed a clear hindsight bias in that the recalled confidence ratings were distorted towards the feedback. Moreover, the consequences of selective activation were evident in that more information favoring the feedback was recalled

Introduction

Hindsight bias or the "Knew-it-all-along-effect" (Fischhoff, 1975) is a well-known systematic phenomenon that is of special interest for the insight it provides into the processes of judgment and recall. But how is our recall of previous knowledge states influenced by supplying new information (e.g., the outcome)? In the face of the outcome, we often seem to overestimate the quality of our previous knowledge, thus leading to a distortion towards the provided information. Suppose, for example, that a group of participants is being asked for the plausibility of absinthe being (a) a precious stone or (b) a liqueur? A second group of participants first receives the correct answer and is then being asked for the plausibility rating of the two alternatives with the instruction to ignore the solution. In comparison to the first group (without solution) the plausibility rating of the second group reveals a higher confidence in the correct alternative as suggested by the solution. That is, subjects of the second group seem to "guess better" (e.g., Hoch & Loewenstein, 1989).

The hindsight bias is even more intriguing if the same subjects are asked for the plausibility of absinthe being (a) a precious stone or (b) a liqueur and, then - usually after some time has elapsed - receive the correct answer, and finally are to remember their original plausibility rating. Now the remembered plausibility ratings are closer to the correct solution than the original ratings were (e.g., Fischhoff, 1977; Wood, 1978). The main difference between these two designs is the task that the subject has to perform, being a hypothetical judgment in the first case and a memory recollection in the second.

Of special interest` in the memory design is the stage at which the memory distortion actually occurs. Some researchers (e.g., Fischhoff, 1977; Loftus & Loftus, 1980) favor early stages, that is, they believe in destructive updating of the original information at the time of encoding the outcome information. Fischhoff (1975) used the term "creeping determinism" to point out that it is completely natural to assimilate outcome knowledge with the original information to create a coherent whole out of all the relevant knowledge. This process depicts learning from the outcome. Other experiments, though, suggest that the distortion takes place at a later stage. This can be inferred from studies showing post-outcome manipulations to be effective. Davies (1987, Experiment 1) found that supplying subjects with notes they had written in the first judgment session considerably reduced the hindsight bias. Equally effective was the post-feedback generation of reasons for all possible outcomes (Davies, 1987, Experiment 3). Hasher, Attig, and Alba (1981, Experiment 2) provided one of the rare examples in which subjects' recollections showed no hindsight bias. The critical debiasing manipulation was to warn subjects that they accidentally received false outcome information.

Conversely, Fischhoff (1977) found that informing the subjects about the bias did not reduce hindsight bias. However, his result was observed in a hypothetical design, in which the correct information was given before the first attempt to respond. Pohl and Hell (1996) found no effect of reducing hindsight bias in a memory design. Neither informing subjects in advance nor individual feedback about their recall performance reduced hindsight bias. The results showed that knowledge about the bias phenomenon did not help subjects to avoid the bias. Findings like these support automatic processes as an explanation for the observed bias and dismiss motivational accounts.

Explanations favoring the final rejudgment process as the point where biasing occurs might be labeled "cognitive-reconstruction" theories (Hawkins & Hastie, 1990). According to these, the hindsight bias is a necessary and unavoidable by-product of collecting evidence in the judgment

process. Hindsight bias is an automatic memory distortion that arises whenever the original response (that is being looked for) has been forgotten or - as in the case of hypothetical designs has never been encoded. The systematic memory distortion occurs because subjects are apparently unable to ignore outcome knowledge during the rejudgment process (cf. Kahneman & Tversky, 1974).

Recently, Pohl (1998) found that when the data were separated according to whether participants considered the feedback value plausible or not, cases of unbiased recollections did emerge: feedback values that were labeled as estimates of another person and found to be implausible did not lead to hindsight bias. This finding argues against the view that hindsight bias is an automatic and unavoidable effect of feedback presentation. There are at least specific circumstances under which it is possible to avoid the influence.

In conclusion, most of the empirical evidence favors cognitive accounts, while motivational manipulations showed only minor effects. The same conclusion was drawn in a meta-analysis, covering 122 hindsight bias studies (Christensen-Szalanski & Willham, 1991). However the findings from Pohl (1998) point out that bias is not always as automatic and unavoidable as has been presumed

Previously proposed cognitive explanations of the hindsight bias are unfortunately not very satisfying. For example the anchoring and adjustment heuristic (Tversky and Kahneman, 1974) originally proposed to explain anchoring effects is also being discussed as an explanation for hindsight bias.

In a typical anchoring study (e.g., Tversky & Kahneman, 1974), participants are first asked whether the answer to a question is above or below a certain number. This number acts like an anchor because it distorts subsequent estimates towards it. Thus, mean estimates following a high anchor are higher than those following a low anchor are.

To explain such anchoring effects, Tversky and Kahneman (1974) proposed that participants start their estimation from the anchor and adjust the value in the direction they think plausible (i.e., higher or lower than the anchor). They stop at the first plausible value, thus leading to estimates that are biased towards the anchor (Jacowitz & Kahneman, 1995). Although plausible in its assumptions, it remains unclear how the anchor produces this restriction. Besides, it has been shown that highly implausible anchors lack any effects of anchoring and highly plausible anchors lead to anchoring (Pohl, 1988) albeit participants should respond with the anchor value in this case according to Jacowitz and Kahneman.

Pohl and Eisenhauer (1997) developed a detailed cognitive model that allows explaining anchoring and hindsight bias on a deeper level and that, moreover, can be used as a simulation model. As basic explanation for distorted judgment or recall, the model assumes a selective activation process of one's item-specific knowledge base. In order to reflect this focus, the model was termed SARA which stands for "Selective Activation, Reconstruction, and Anchoring" (Pohl & Eisenhauer, 1997). All processes (i.e., generating, encoding, forgetting, and reconstructing) change the associative pattern between the elements of one's knowledge base and possible retrieval cues, thus leading to a different probability of retrieval. SARA's general architecture is based on "SAM"--the *Search of Associative Memory* model (Raaijmakers & Shiffrin, 1980).

The subsequent part of this paper describes selective activation, the central assumption of SARA, in more detail and presents two experiments that support this explanation of hindsight bias.

In a typical hindsight experiment within the memory design, participants are asked to answer difficult almanac questions. Suppose for example, that you are asked for the height of the Eiffel tower? If you don't know the correct answer, there are two options that could lead to an answer: you could guess, or activate knowledge. In the second case, you are probably neither able nor willing to access all knowledge theoretically available to answer the question. The basic idea is that the representation of the information units in memory could be described as an associative network comprising all information dealing with the specific question: the knowledge base. Thus the task to give an estimate to an almanac question leads to the attempt to recall some of the information units of one's knowledge base, depending on their level of association. In other words, only strongly associated informations are likely of being activated. Depending on the time available and on the motivation, you would probably generate not more than two or three information units. This is a reasonable assumption especially if there are 50 or more almanac questions to be answered. For example, the mean height of buildings at the turn of the century, or that the Eiffel tower is a steel construction, could come to your mind. Those informations would be translated into numerical values and summarized in one value, the estimate for the question (e.g., "250 meters").

After some time has elapsed, the solution is provided (e.g., "300 meters"). Because it is the answer to the original question you will probably try to encode the solution. The process of encoding information shares many features with the generation of an estimate. It may be seen as a reverse retrieving process. The solution will be associated with information units in the knowledge base, systematically increasing the strength of association of information units close to the solution (e.g., "steel construction"; "build for a world's fare"). The central process of SARA is that the encoding of the solution leads to a selective activation of associated units of the knowledge base. The result is that information units strongly associated to the solution are increased in their associative-strength level. Finally, the solution is added to the knowledge base, thus completing its encoding.

In the last phase of a typical hindsight experiment, when you are asked to recall your original estimate the process will be exactly the same as in the generation phase. SARA assumes that the process should be based only on some of the information units of the knowledge base. The probability to access information units varies with their associative strength towards the currently present retrieval cues: again, only strongly associated information is likely of being activated. But the pattern of association strength, has changed because of the encoding of the solution with the result that certain information units are strengthened and more likely to

be activated when a person tries to reconstruct his original estimate. The probability to retrieve units closely associated to the solution in a following task should thus be increased. Consequently the recollection will most probably be systematically biased towards the solution (e.g., "250 meters"). In the remainder, we present two experiments that examine selective activation more closely.

Experiment 1

Method

Material It is next to impossible to lay out the specific information units that are potentially used to generate an estimate to a specific question. Therefore, we decided to supply the specific knowledge bases in our experiments.

We used a confidence task: Participants had to give ratings about how confident they were whether a certain quantity increased or decreased in value (e.g., increase or decrease in sales of a fictitious corporation). In order to judge how the quantity may have changed, participants received four arguments favoring an increase and four favoring a decrease of the quantity.

One example was following question:

Question:

"In 1988 88 % of the American adults believed in the right to beat their children. The percentage changed up to now."

Arguments favoring decrease:

- "authoritarian education being criticized"
- "TV-advertisements about serious consequences of violence in the family"
- "reporting in the media of abuse in the family"
- violence in the family leading theme of the universal day of the child"

Arguments favoring increase:

- "deficiency of antiauthoritarian education"
- "growth of authoritarian religious communities"
- "popularity of rigorous conservative colleges"
- "association of adolescent violence with missing limits in education"

Task:

"Please report how confident you are that the percentage increased"

Probability of increase (in percent): e.g. 40 %

Participants, design and procedure One hundred and eight students (80 female, 28 male; between 18 and 52 years old with a mean age of 23.6 years) of different faculties of the University of Trier took part in the experiment.

The experiment consisted of two sessions: In Session 1, participants had to fill out a questionnaire with 24 verification tasks and to indicate their confidence whether the quantity increased or decreased. In Session 2 (one week later) they had to recollect their confidence ratings of the first session. The questionnaire in the second session however was presented with some of the solutions, indicating in 8 cases that the quantity had increased and in other 8 cases that it had decreased. The remaining 8 cases contained no feedback and served as control cases. The selection of experimental items was counterbalanced across participants, so that all questions served equally often as experimental and as control items. Participants were asked to recall their own estimates given one week ago. The instructions stressed that there was no interest in the memory of the solutions but rather in the memory of the participants' first confidence ratings. The order of questions was identical in both sessions and there was no time limit. After the attempt to recollect the first confidence rating, participants had to remember in a free recall test as many arguments to each problem case as possible. At the end, participants were debriefed about hindsight bias and the goal of the experiment. The total experiment lasted about 60 minutes per person.

The dependent variable to measure hindsight bias (labeled "• %") was defined as the difference in the confidence rating for a decrease in the first session minus the confidence rating for a decrease in the second session.

•% = (confidence decrease (t1) - confidence decrease (t2))

A positive value of •% indicates a greater confidence for increase in Session 2 in comparison to Session 1, whereas a negative value indicates a greater confidence for increase. A shift towards the feedback indicates hindsight bias. The feedback that in fact "it increased" should lead to a positive value of •%. Conversely, the feedback "it decreased" should lead to a negative value of •%.

The number of arguments recalled favoring increase or decrease measured the assumed effects of selective activation. The level of significance was set to $\alpha = .05$ for all analyses.

Results

Confidence Ratings A repeated measures ANOVA for the feedback (increase – no feedback – decrease) revealed a distinct effect of the feedback. The feedback "increase" resulted in a positive value of • % (2.8), no feedback in a minor positive value (1.0) and the feedback "decrease" led to a negative value of • % (-3.3). The results showed a clear shift of the confidence ratings in the second session towards the feedback ($F_{(1,98)} = 30.7$).

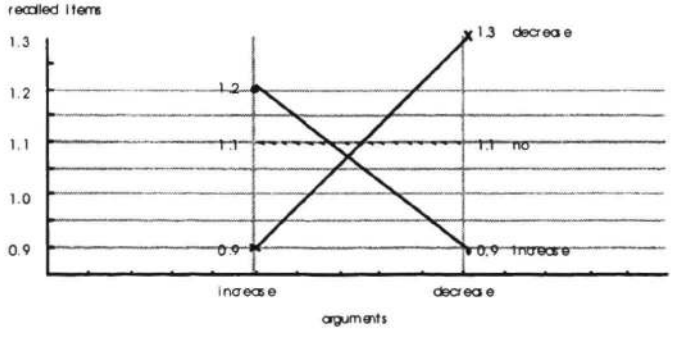

Figure 1: Mean number of recalled arguments in each experimental condition of Experiment 1.

Recalled arguments The number of correctly recalled arguments was analyzed in a 3 x 2 MANOVA with the factors feedback (increase – no feedback – decrease) and type of argument (favoring increase or decrease). The interaction showed a clear and distinct effect of the feedback ($F_{(2,208)} = 33.5$; see Fig. 1). The feedback that "the quantity increased" led to a better recall of arguments favoring increase (1.2 vs. 0.9) whereas the feedback that "the quantity decreased" led to a better recall of arguments favoring decrease (1.3 vs. 0.9). In the condition without feedback no difference between arguments favoring increase or decrease could be observed.

Discussion

The significant shift • % in confidence ratings depending on the feedback denotes hindsight bias. The feedback that a quantity increased augmented the confidence for increase (or lowered the confidence for decrease) in the second session as compared to the first. Whereas the feedback that a quantity decreased augmented the confidence for decrease (or lowered the confidence for increase) in the second session as compared to the first. Without feedback confidence for increase augmented slightly indicating a minor positivity bias.

The analysis of the number of recollected arguments in the free recall showed a significant interaction between type of feedback and type of argument. Whenever the feedback indicated that the fact increased, significantly more arguments were recollected favoring increase. Accordingly, a feedback of decrease led to more recollected arguments favoring decrease. Thus, significantly more arguments favoring the feedback were recollected implying a selective activation of arguments favoring the feedback. The result of this experiment can be taken as a first confirmation of selective activation as a promising explanation for hindsight bias. In a second experiment, we tried to find more evidence for selective activation. Unlike Experiment 1, we used a hypothetical design with only one session.

Experiment 2

Method

Material and design The material was the same as in the questionnaire of Experiment 1 with the only difference that the experiment took part on a computer. Participants had to answer 24 verification tasks and to indicate their confidence weather the quantity increased or decreased. Eight cases were presented with solutions indicating that the fact had increased and eight cases that it had decreased. The remaining eight cases contained no feedback and served as control cases. The design of Experiment 2 corresponded with that of Experiment 1 with the exception that the experiment took place in one session. The dependent variable was the confidence in increase dependent upon the feedback. The number of arguments in free recall favoring increase or decrease was taken to indicate selective activation.

Participants and procedure One hundred and four students (69 female, 34 male; between 18 and 40 years old with a mean age of 22.9 years) of different faculties of the University of Trier took part in the experiment. The experiment consisted of one session. Similar to the task in Experiment 1, participants received a fact on the computer screen, but this time the solution was provided immediately (except for eight control-cases without solution). The arguments favoring increase and decrease of the quantity followed subsequently. Participants were then asked to report their own confidence independent from the feedback received. There was no time limit. At the end, participants were debriefed about hindsight bias and the goal of the experiment. The experiment lasted about 30 minutes.

Results

Confidence Ratings A repeated measures ANOVA for the feedback (increase – no feedback – decrease) revealed a distinct effect of the feedback. The feedback "increase" resulted in a greater confidence for increase (64.2 %), no feedback in a medium confidence rating (55.1 %) and the feedback "decrease" led to a lesser confidence for increase (39.8 %). The results showed a clear dependency of the confidence to the feedback ($F_{(2,204)} = 58.8$).

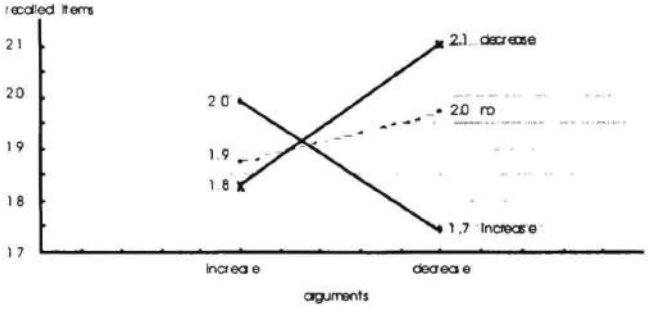

Figure 2: Mean number of recalled arguments in each experimental condition of Experiment 2.

Recalled arguments The number of correctly recalled arguments was analyzed in a 3 x 2 MANOVA with the factors feedback (increase – no feedback – decrease) and type of argument (favoring increase or decrease). The interaction showed an effect of the feedback ($F_{(2,204)} = 12.4$; see Fig. 2). The feedback that "the quantity increased" led to a better recall of arguments favoring increase (2.0 vs. 1.7) whereas the feedback that "the quantity decreased" led to a better recall of arguments favoring decrease (2.1 vs. 1.8). In the condition without feedback no noticeable difference between arguments favoring increase or decrease (1.9 vs. 2.0) could be observed.

Discussion

The difference in confidence ratings in the different feedback conditions revealed hindsight bias. The feedback that a quantity increased augmented the confidence for increase as compared to no feedback. The feedback that a fact decreased lowered the confidence for increase compared to no feedback.

The analysis of the number of recollected items in the free recall showed again a significant interaction between type of feedback and type of argument. Whenever the feedback indicated that the quantity increased significantly more arguments were recollected favoring increase. Correspondingly a feedback of decrease led to more recollected arguments favoring decrease. Significantly more arguments favoring the feedback were recollected implying again a selective activation of arguments favoring the feedback. The result of this experiment substantiates those from Experiment 1 and consolidates selective activation as a promising explanation for the hindsight bias.

Conclusion

Both experiments successfully demonstrated the existence of selective activation. Selective activation thus appears to be a promising explanation of hindsight bias. The solution or anchor proved to have a distinctive influence: Items favoring the anchor were recollected more often compared to those supporting the opposite. The central concept of SARA (Pohl & Eisenhauer, 1997), namely selective activation of the item specific knowledge base, was confirmed in the two reported experiments. SARA makes detailed assumptions about a person's pre-experimental knowledge base and how it is altered in the course of the experiment. All processes (i.e., generating, encoding, forgetting, and reconstructing) change the associative pattern between the elements of one's knowledge base and possible retrieval cues, thus leading to a different probability of retrieval. According to SARA, anchored reconstruction results from a *selective activation* of one's item-specific knowledge base (Hawkins & Hastie, 1990; Strack & Mussweiler, 1997). This activation is governed by the anchor value and is considered being selective, because information that is more similar to (or consistent with) the anchor will receive more activation than other information (Kahneman & Miller, 1986). After selective activation, the probability of retrieving a certain piece of information from one's knowledge base has changed. As a consequence, any attempt to generate or to reconstruct an "unbiased" estimate is bound to fail. Most probably, the resulting estimate will be biased towards the anchor value.

Selective activation is able to explain most of the findings in the field of anchoring. For example, Fischhoff's experiments (1975) on "creeping determinism" share many features with the reported experiments. Its participants received passages describing an unfamiliar historical event and had to evaluate the probability of four possible outcomes in the light of a solution. As in our experiments, a shift towards the solution was observed. Contrary to Fischhoff, however we don't assume an irreversible and immediate assimilation of the solution. In our opinion the shift towards the solution reflects a selective activation of the knowledge base promoting arguments in favor of the provided solution.

The results of the reported experiments support the basic idea to explain and to model distortions in judgment and memory through selective activation of one's item-specific knowledge base.

References

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

Davies, M. F. (1987). Reduction of hindsight bias by restoration of foresight perspective: Effectiveness of foresight-encoding and hindsight retrieval strategies. *Organizational Behavior and Human Decision Processes, 40*, 50–68.

Fischhoff, B. (1975). Hindsight = foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental Psychology: Human Performance and Perception, 3*, 349–358.

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

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

Hell, W., Gigerenzer, G., Gauggel, S., Mall, M., & Müller, M. (1988). Hindsight bias: An interaction of automatic and motivational factors? *Memory and Cognition, 16*, 533–538.

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

Jacowitz, K. E. & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin, 21*, 1161–1166.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93*, 136–153.

Loftus, E. F., & Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist, 35*, 409–420.

Pohl, R.-F. (1998). The effects of feedback source and plausibility of hindsight bias. *European-Journal-of-Cognitive-Psychology, 10(2)*, 191–212

Pohl, R. F., & Eisenhauer, M. (1997). SARA: An associative model for anchoring and hindsight bias. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society* (p. 1103). Mahwah, NJ: Erlbaum.

Pohl, R.-F., & Hell,-W. (1996). No reduction in hindsight bias after complete information and repeated testing. *Organizational-Behavior-and-Human-Decision-Processes, 67(1)*, 49–58.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14; pp. 207–262). San Diego, CA: Academic Press.

Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology, 73*, 437–446.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*, 1124–1131.

Wood, G. (1978). The "knew-it-all-along" effect. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 345–353.