Journal of Applied Psychology
1988, Vol. 73, No. 2, 305-307

0021-9010/88/$00.75

Eliminating the Hindsight Bias

Hal R. Arkes
Ohio University

David Faust
Brown University

Thomas J. Guilmette
Vanderbilt Rehabilitation Center
Newport, Rhode Island

Kathleen Hart
Brown University

Those who consider the likelihood of an event after it has occurred exaggerate their likelihood of having been able to predict that event in advance. We attempted to eliminate this hindsight bias among 194 neuropsychologists. Foresight subjects read a case history and were asked to estimate the probability of three different diagnoses. Subjects in each of the three hindsight groups were told that one of the three diagnoses was correct and were asked to state what probability they would have assigned to each diagnosis if they were making the original diagnosis. Foresight-reasons and hindsight-reasons subjects performed the same task as their foresight and hindsight counterparts, except they had to list one reason why each of the possible diagnoses might be correct. The frequency of subjects succumbing to the hindsight bias was lower in the hindsight-reasons groups than in the hindsight groups not asked to list reasons, $\chi^2(1, N = 140) = 4.12, p < .05$.

People who consider an event after it has occurred often claim that they would easily have been able to predict the event in advance if they had been asked to do so. However, research has shown that those people who are actually asked to predict that event do not find the prediction as easy as the hindsight subjects claim it is (Fischhoff, 1975). Hindsight bias is defined as the tendency for people considering a past event to overestimate their likelihood of having predicted its occurrence.

Hindsight bias has been demonstrated in a variety of applied settings including politics (Leary, 1982), historical judgment (Fischhoff, 1980), psychotherapy case histories (Fischhoff, 1975), medical diagnoses (Arkes, Wortmann, Saville, & Harkness, 1981; Dawson, Arkes, Blinkhorn, Lakshmanan, & Petrelli, 1986), and employee evaluation (Mitchell & Kalb, 1981). For example, in the Arkes et al. experiment, 15 physicians in the foresight group were asked to consider an actual case history and estimate the likelihood of four possible diagnoses. Each of four hindsight groups was told that one of the diagnoses was correct and then was asked what probability subjects would have assigned to each diagnosis if they had not known the correct answer. Hindsight subjects overestimated the likelihood that they would have assigned the "correct" diagnosis as compared with the estimate actually given by the foresight group.

A few attempts have been made to eliminate the hindsight bias; however, in general, these debiasing efforts have failed. For example, Fischhoff (1977) and Wood (1978) told subjects in their groups what the hindsight bias was and exhorted them to try to prevent its occurrence. Such warnings proved ineffective.

We used a debiasing procedure similar to that of Koriat, Lichtenstein, and Fischhoff (1980) to reduce the overconfidence people had in their answers to general knowledge questions. We felt that this procedure might be redirected to the goal of countering the hindsight bias. Koriat et al. presented subjects with two-alternative questions such as "The Sabines were part of (a) Ancient India or (b) Ancient Rome." Subjects chose the answer they believed to be correct and then indicated their confidence in the chosen answer. Control group subjects exhibited levels of confidence far greater than their levels of accuracy. Experimental group subjects were asked to state reasons why each of the two options might be correct and why each of the two might not be correct before they indicated their confidence in the chosen answer. These subjects manifested significantly less overconfidence than did control group subjects. We applied a similar procedure in the present study of differential diagnosis among professional neuropsychologists. Before hindsight subjects indicated the probability they would have assigned to the "correct" diagnosis, we asked them to provide a reason why each of the other diagnoses might have been correct. If the Koriat et al. results generalized, we would have expected the bias to be less in the "reasons" subjects compared with those who were not asked to list such reasons.

A procedure used by Lord, Lepper, and Preston (1984) that is closely related to the Koriat et al. (1980) procedure has proven effective. Lord et al. (Experiment 1) asked subjects who either did or did not favor capital punishment to consider two bogus pieces of research, one of which did and one of which did not support the utility of capital punishment in deterring murder. Control group subjects who examined this research information manifested a polarization of attitude; proponents were more in favor of the death penalty after reading the information, and opponents were more against it. "Consider the opposite" subjects were asked during information presentation whether they would have evaluated each piece of research the same had its results been exactly the opposite. These subjects showed less attitude polarization than did control group subjects. Koriat et al. found that considering the opposite was

Correspondence concerning this article should be addressed to Hal R. Arkes, Department of Psychology, Ohio University, Athens, Ohio 45701.

effective in reducing overconfidence. However, Fischhoff and Macgregor (1982) simply could not replicate the Koriat et al. results in a situation similar to that used in the original study. Thus, another goal of our study was to help determine the replicability of the Koriat et al. findings.

Method

Subjects

As part of a separate study, a list was compiled of all of the psychologists in the United States who indicated neuropsychology as a primary field or area of specialization in either or both the preprint of the *Directory of the American Psychological Association* (American Psychological Association, 1984) and the 1983 edition (plus the 1984–1985 supplement) of the *National Register of Health Service Providers in Psychology.* Of these 3,472 persons, the great majority were engaged in clinical practice. The majority were men; 94% had a PhD degree, and 3% were diplomates in neuropsychology. Number of years elapsed since obtaining the last degree ranged from less than 1 to 69, with a median of 11 years. Almost 90% had held their degree for at least 4 years and almost 80% for at least 6 years.

A total of 480 psychologists were randomly selected from the sample and were divided into eight groups of 60 each.

Procedure

The 60 subjects in the foresight group were mailed the following questionnaire:

> We would greatly appreciate your taking a few minutes to read the case below, which is part of an adjunct study on neuropsychological diagnosis. We would like you to decide what probability you would have assigned to each of three possible conditions had you been making the diagnosis.
>
> A fifty-five-year-old man presents to an outpatient psychiatric clinic in a mildly confused and agitated state. He is oriented to person and place but is uncertain about the date, misstating it by about a week. He says he is having problems with his "nerves," and he exhibits a mild bilateral intention tremor.
>
> Review of background information indicates an unsteady work history with frequent job changes over the last three years. He has not even held undemanding jobs over the last two years due to poor work performance. The man's social network is limited to "a few drinking buddies." His wife reports that he has been drinking steadily for at least five years, averaging a six-pack or more a day, but that he stopped drinking one week ago when a close friend was killed while driving home from a bar. She also says that "he can't remember like he used to, his memory is getting worse and worse," and that "both of his parents got the same problem and went right downhill," when they were in their fifties. When questioned, she is not certain what condition the parents had but indicates they died in their early sixties and fears her husband will die similarly.
>
> Neurological exam is normal except for bilateral tremor. A CT-Scan is "equivocal but probably normal," although mild cortical atrophy cannot be ruled out. Neuropsychological testing indicates mild to moderate dysfunction across most areas with the greatest difficulties shown in attention, concentration, and higher level reasoning. For example, Digit Span is 4 (forwards) and 3 (backwards), and performance is poor on Trails (Part B) and Wisconsin Card Sorting Test.
>
> Based on your consideration of the case, what is the probability you would assign to each of the diagnoses as the *primary* one?
>
> Assign a probability to each, making sure the probabilities add to 100%.
>
> Alcohol withdrawal _____%
> Alzheimer's disease _____%
> Brain damage secondary to alcohol abuse _____%

Table 1
Mean Probability Assigned to Each Diagnosis

Group	*N*	Outcome provided	Outcome evaluated: Alcohol withdraw (AW)	Alzheimer's disease (AD)	Brain damage/ alcohol abuse (BD)
No reasons					
Foresight	30	None	37	26	37
Hindsight	22	AW	44 (12)	24	33[a]
Hindsight	22	AD	27	34 (11)	38[a]
Hindsight	28	BD	22	28	50 (19)
Reasons					
Foresight	24	None	33	32	34
Hindsight	22	AW	29 (5)	34	36[a]
Hindsight	23	AD	22	39 (10)	39
Hindsight	23	BD	22	38	39[a] (13)

Note. All probabilities are multiplied by 100. The numbers in parentheses indicate the number of neuropsychologists whose probability for that particular diagnosis exceeded the corresponding foresight estimate.
[a] Row does not sum to 100.0 because of rounding error.

Each of the three hindsight groups had an extra sentence inserted between the two sentences that were in the first paragraph of instructions to the foresight group. The three sentences (one for each of the hindsight groups) were as follows: "This is the case of a man with a primary diagnosis of alcohol withdrawal." "This is the case of a man with a primary diagnosis of Alzheimer's disease." "This is the case of a man with a primary diagnosis of brain damage secondary to alcohol abuse."

Foresight-reasons-group subjects received a questionnaire like the one received by foresight subjects, except the following was inserted before the final paragraph: "After *each* of the following diagnoses, please jot down one piece of evidence from the case history that would support that particular diagnosis as the primary one." The three possible diagnoses were then listed.

The three hindsight–reasons questionnaires differed from the hindsight ones in an analogous way.

Results

Completed questionnaires were returned by 194 (40%) of the psychologists. Each of the eight cells ranged in size from 22 to 30 subjects. The mean probability assigned to each diagnosis by each group is presented in Table 1. Also included in the table within parentheses are the number of neuropsychologists whose probability for each diagnosis exceeded the corresponding foresight estimate.

Because the probabilities assigned to the three diagnoses are ipsative for every subject—the estimates must add up to 100—normal parametric statistical procedures are problematic. Another way to examine the data is simply to compare the number of hindsight and hindsight–reasons subjects who manifested a bias. Of 72 hindsight subjects, 42 (58%) gave a higher probability estimate to the correct diagnosis than the corresponding esti-

mate obtained from the foresight group. Only 28 of 68 hindsight–reasons subjects (41%) gave a higher probability estimate to the correct diagnosis than the corresponding estimate obtained from the foresight–reasons group. A chi-square analysis comparing these two splits (42:30 vs. 28:40) confirms the hypothesis that the frequency of subjects manifesting the hindsight bias is significantly greater among hindsight subjects than among hindsight–reasons subjects, $\chi^2(1, N = 140) = 4.12$, $p < .05$.

It is interesting to note that the hindsight subjects' data are a mirror image of the hindsight–reasons subjects' data; in the former, 6 more subjects than would be expected by chance gave higher estimates than did the corresponding foresight subjects. In the latter, 6 fewer subjects than would be expected by chance gave higher estimates than did the corresponding foresight subjects.

Discussion

The debiasing procedure previously used by Koriat et al. (1980) to reduce overconfidence in answers to general knowledge questions significantly reduced the number of subjects manifesting the hindsight bias. This demonstration of an effective debiasing technique is noteworthy because the hindsight bias has been shown to occur in such a wide variety of judgment tasks.

The practical consequences of this debiasing procedure should prove beneficial and may extend beyond the hindsight bias alone. Physicians, psychologists, and other diagnosticians often inaccurately claim they "knew it all along" when told the true diagnosis. The hindsight bias thereby reduces what the diagnostician can learn from the outcome information, because he or she thinks the outcome is already so obvious.

In some hindsight studies—for example, Fischhoff and Beyth (1975)—subjects are asked after an actual event has occurred what probability they would have assigned to the event had they been asked to predict it. In such experiments, the occurrence (or nonoccurrence) of a prior event is an incontrovertible fact. On the other hand, in diagnosis studies such as the present one, the initial diagnosis may be one about which the hindsight subject has some measure of doubt. The reasons manipulation may make this doubt more salient, thereby reducing the magnitude of the hindsight bias.

The hindsight bias is one expression of the overconfidence that has been demonstrated across a broad range of judges and tasks. (See Fischhoff, 1982, for an overview.) In one of the few related studies using neuropsychologists, Wedding (1983) obtained nonsignificant but consistently negative correlations between subjective confidence and accuracy on a series of diagnostic judgments. Among 14 judges, the 2nd most confident was the least accurate, and 2 of the least confident judges were the most accurate. Overconfidence may lead to premature conclusions and insufficient consideration of alternative possibilities, thus decreasing judgment accuracy.

Having the diagnostician list or generate reasons why other outcomes might have been expected should heighten appreciation of the difficulty of the case, the plausability of other diagnoses, and the information value of the correct answer. Under these circumstances the correct answer will more likely be seen as truly informative, and alternative possibilities may be given their just due. This will improve both the accuracy of the postdiction ("I realize that I wouldn't have made that diagnosis after all") and any subsequent predictions ("Next time I won't be so positive that it has to be disease X"). It may also counter premature tendencies to terminate the search for additional information and encourage attempts to disconfirm hypotheses.

References

American Psychological Association. (1984). *Directory of the American Psychological Association.* Baltimore, MD: Port City Press.

Arkes, H. R., Wortmann, R. L., Saville, P., & Harkness, A. R. (1981). The hindsight bias among physicians weighing the likelihood of diagnoses. *Journal of Applied Psychology, 66,* 252–254.

Dawson, N. V., Arkes, H. R., Blinkhorn, R., Lakshmanan, M., & Petrelli, M. (1986, October). *Hindsight bias: An impediment to accurate probability estimation in clinicopathologic conferences.* Paper presented at the meeting of the Society for Medical Decision Making, Chicago.

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 Perception and Performance, 3,* 349–358.

Fischhoff, B. (1980). For those condemned to study the past: Reflections on historical judgment. In R. A. Shweder & D. W. Fiske (Eds.), *New directions for methodology of behavioral science: Fallible judgment in behavioral research* (pp. 79–93). San Francisco: Jossey-Bass.

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 422–444). New York: Cambridge University Press.

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

Fischhoff, B., & Macgregor, D. (1982). Subjective confidence in forecasts. *Journal of Forecasting, 1,* 155–172.

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

Leary, M. R. (1982). Hindsight distortion and the 1980 presidential election. *Personality and Social Psychology Bulletin, 8,* 257–263.

Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology, 47,* 1231–1243.

Mitchell, T. R., & Kalb, L. S. (1981). Effects of outcome knowledge and outcome valence on supervisors' evaluations. *Journal of Applied Psychology, 66,* 604–612.

National Register of Health Service Providers in Psychology. (1983). Washington, DC: Council for the National Register of Health Service Providers in Psychology.

Wedding, D. (1983). Clinical and statistical prediction in neuropsychology. *Clinical Neuropsychology, 2,* 49–55.

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

Received January 9, 1987
Revision received August 21, 1987
Accepted June 28, 1987 ■