

AN EARLY HISTORY OF HINDSIGHT RESEARCH

Baruch Fischhoff
Carnegie Mellon University

The first studies of hindsight bias reflected the confluence of two desires. One arose from being part of the heady early days of Amos Tversky and Danny Kahneman's heuristics-and-biases research program. As the paradigm evolved, the challenge for the participating graduate students was to find a heuristic to call one's own or to find a way to elaborate one of the three "classics" (availability, representativeness, anchoring, and adjustment). Maya Bar Hillel and Ruth Beyth-Marom chose the latter route. However, I was still struggling to reconcile the political motives that had brought me to Israel, planning to live in a kibbutz for the rest of my life, with academic life—and its more realistic view of the pace of change in human affairs.

For one meeting of the seminar, we read Paul Meehl's (1973) "Why I Do Not Attend Case Conferences." One of his many insights concerned clinicians' exaggerated feeling of having known all along how cases were going to turn out. To me, this sounded a lot like the exaggerated claims of understanding political processes that permeated the political discussions to which I had long subjected myself. Those discussions often left me wondering, "If we're so prescient, why aren't we running the world?" Psychological research provided an opportunity (and an obligation) to discipline such observations with systematically collected evidence, interpreted in the context of evolving theory. It occurred to me that I might have my bias, if I could provide the evidence and theory.

The evidence came first. The research group was very sensitive to the need for normative analysis, establishing the standard for evaluating performance. The task that provided the clearest demonstration of hindsight bias came in the initial study. President Nixon was about to leave for his historic trips to China and the Soviet Union. Ruth and I asked people to assess the probabilities of various possible outcomes (e.g., Pres. Nixon will meet Chairman Mao; Pres. Nixon will announce that the trip was a success). After the trips were over, we asked subjects to recall their predictions. Fortunately, we had the foresight to ask them what they thought had happened, so that we

Address correspondence to Baruch Fischhoff, Department of Social and Decision Sciences, Carnegie Mellon University, 208 Porter Hall, Pittsburgh, PA 15213. E-mail: baruch@cmu.edu.

could evaluate any bias in their memory, relative to their own beliefs (which often differed from press accounts of what had happened).

Assuming that subjects accepted our claim of testing their memory, rather than their knowledge of international relations, identifying bias was relatively straightforward. Did they remember having given higher probabilities than they actually had for events that they thought had happened (and lower probabilities for ones that had not)? Theoretically, though, the bias seemed multiply determined, fitting any account of revising beliefs that left prior representations less available, if not lost entirely. In the spirit of the times, we thought of hindsight bias as arising from the use of a sense-making heuristic, whereby people integrate all they know about a topic into a coherent mental model. Like other heuristics, it was good for some things (looking forward with a full set of beliefs), but bad for others (reconstructing previous perspectives).

However, the recall design has its limits. Having to find participants twice makes it hard to study hindsight for temporally distant events. (We recently did it over a year's time [Fischhoff, Gonzalez, Lerner, & Small, 2005].) Moreover, the design cannot be used with events that people did not judge in foresight. A task that can be used with any event is to ask people what they would have predicted in a situation, had they been asked, after being told what had happened. Another friend, Aron Hirt-Mannheimer, and I searched history books for the right kind of event. Aron found the British-Gurkha story. Once we had prepared it as a stimulus, producing others was straightforward. (About 20 years later, someone asked for the other stimuli. I was stunned to discover how inappropriate one had become. Taken from Ellis (1966), it described a "troubled" homosexual, in terms that fit now-antiquated mores and theories. Although I sent it, I argued that it should not be used. In addition to being offensive, it had changed its meaning so much that using it would not represent a replication.)

Fischhoff (1975) describes other attempts to deal with possible demand characteristics with "what would you have said?" instructions (as do Slovic & Fischhoff, 1977). A nagging normative question was how much people had actually learned from observing an event. At one seminar meeting, Danny Kahneman suggested an analogy with drawing colored balls from an urn (a popular research task in those days; Slovic & Lichtenstein, 1971). In historical terms, the parallel involves a class of "four British-Gurkha-type struggles, each indistinguishable from the one used [in the study], two of which were won by the Gurkhas. Upon learning of another Gurkha victory, [one might] properly update that outcome's predictive (Before) probability of 50% to a higher postdictive (After) probability" (p. 293).

A less formal and direct form of normative analysis involved asking what experts in historical judgment advise. Fischhoff (1975) offers some connections to relevant sources in historiography; Fischhoff (1982) has more. The best guidance, though, came from a philosopher of science. Despite not believing in the social sciences, Imre Lakatos (1970) offered an account of scien-

tific thinking that made good cognitive sense. His “myth of the critical experiment” showed how scientists exaggerated how fully the implications of experiments (like Michelson–Morley) were understood when they first appeared (see also Motterlini, 1999).

The 1973 War interrupted the research (and many other things). The re-criminations over how surprised Israel had been seemed to have an element of hindsight bias, adding insult to the injury of the war. Despairing of getting something published in the general press, Ruth and I managed to place a book review (Fischhoff & Beyth, 1974) of Irving Janis’s (1972) *Victims of Groupthink*, which had become part of the popular debate. While conceding Janis’s insights into group processes, we wondered how much his selection and interpretation of good and bad cases (Cuban Missile Crisis, Bay of Pigs) had been influenced by knowing how they had turned out. I only had one opportunity to meet Prof. Janis, about ten years later. Our review was the first thing that he wanted to discuss.

It is challenging to reconstruct what were realistic expectations for hindsight research, at the time of these studies. Recently, Roediger (2004) wrote a presidential column for the *APS Observer* on most-cited dissertations. Mine had not come to his attention. Out of curiosity, I looked to see how it rated. To my amazement, it would have made the top ten. The two main articles from it, Fischhoff (1975) and Fischhoff and Beyth (1975) had received 533 and 207 citations, respectively. Given the base rate of citations for scientific publications, no evidence available at that time should have been diagnostic enough to predict anything like that. I suppose that the work deserves some credit. So does the creativity of those who have found it a useful platform for studying a variety of processes.

(The third article from the dissertation, Fischhoff (1976) considered whether being set in the past per se changed an event’s apparent probability. It had been cited 19 times. I was surprised that anyone had read it at all, given that it concluded by accepting the null hypothesis of no effect. Its two reviewers were divided: One thought the result was obvious; the other was sure that it wasn’t true. That contradiction and the fact that I was halfway around the world led the editor to accept it without changes.)

The clearest prediction at that time came from the woman running the office that processed dissertations. She took one look at my stack of six copies and said, “You won’t have any trouble.” I asked how she knew, given that she had not even seen the title. “The thin ones go right through.” Seeing my puzzlement, she took me into the next room and pointed to one that would have trouble. Each two-volume copy was thicker than my full set. We looked to see what the topic was. Something in French literature. I have often wondered how its author fared.

REFERENCES

- Ellis, A. (1966). Psychosexual and marital problems. In L. A. Berg & L.A. Pennington (Eds.), *An introduction to clinical psychology*. New York: Ronald.
- ▶ Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 104, 288–299.
- ▶ Fischhoff, B. (1976). The effect of temporal setting on likelihood estimates. *Organizational Behavior and Human Performance*, 15, 180–194.
- Fischhoff, B. (1982). For those condemned to study the past: Reflections on historical judgment. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*. New York: Cambridge University Press.
- Fischhoff, B., & Beyth, R. (1974). Failure has many fathers. Review of I. Janis, *Victims of Groupthink*. *Megamot*, 28, 325–328 (in Hebrew). Reprinted in *Policy Sciences*, 7, 388–393.
- ▶ 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., Gonzalez, R.M., Lerner, J.S., & Small, D.A. (2005). Evolving judgments of terrorism risks: Foresight, hindsight, and emotion. *Journal of Experimental Psychology: Applied*, 11, 124–139.
- Janis, I. (1972). *Victims of groupthink*. Boston: Houghton Mifflin.
- Lakatos, I. (1970). Falsification and the scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of scientific knowledge*. Cambridge, UK: Cambridge University Press.
- Meehl, P.E. (1973). Why I do not attend case conferences. In *Psychodiagnosis: Selected papers* (pp. 225–302). New York: Norton.
- Motterlini, M. (1999). *For and against method*. Chicago: University of Chicago Press.
- Roediger, H.L., III. (2004). Great dissertations: Mark I. *Observer*, 17(10).
- ▶ Slovic, P., & Fischhoff, B. (1977). On the psychology of experimental surprises. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 544–551.
- ▶ Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 6, 649–744.