

CLASSIC PAPER

Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty*

B Fischhoff

Qual Saf Health Care 2003;12:304–312

One major difference between historical and nonhistorical judgment is that the historical judge typically knows how things turned out. In Experiment 1, receipt of such outcome knowledge was found to increase the postdicted likelihood of reported events and change the perceived relevance of event descriptive data, regardless of the likelihood of the outcome and the truth of the report. Judges were, however, largely unaware of the effect that outcome knowledge had on their perceptions. As a result, they overestimated what they would have known without outcome knowledge (Experiment 2), as well as what others (Experiment 3) actually did know without outcome knowledge. It is argued that this lack of awareness can seriously restrict one's ability to judge or learn from the past.

historian Georges Florovsky¹ notes: "The tendency toward determinism is somehow implied in the method of retrospection itself. In retrospect, we seem to perceive the logic of the events which unfold themselves in a regular or linear fashion according to a recognizable pattern with an alleged inner necessity. So that we get the impression that it really could not have happened otherwise" (page 369).

An apt name for this hypothesized tendency to perceive reported outcomes as having been relatively inevitable might be "creeping determinism"—in contrast with philosophical determinism, which is the conscious belief that whatever happens has to happen.

Phenomena resembling creeping determinism have been noted by psychologists as well as historians. One example is Tversky and Kahneman's "law of small numbers," the belief that data which were observed more or less had to be observed.² A second example is the tendency to rework or reconstruct the biographies of deviants to show that their present diagnoses (labels) are inevitable products of their life histories.^{3–5} A third is the defensive attribution of responsibility for accidents, a process in which people carefully scrutinize the data describing accidents in order to uncover or impose a pattern that will increase their perceived predictability and avoidability.⁶

All of this evidence for creeping determinism is, however, either indirect, imprecise, unsystematic (anecdotal), or confounded by motivational and emotional issues. Experiment 1 directly tested the validity of the creeping determinism hypothesis and explored some of the concomitant effects of outcome knowledge on judgment.

EXPERIMENT 1

Method

Design

The six subexperiments described in this section are identical except for the stimuli used. In each, subjects were randomly assigned to one of five experimental groups, one Before group and four After groups. In each subexperiment, the Before group read a brief (150 word) description of a historical or clinical event for which four possible outcomes were provided. The After groups read identical passages to which a final sentence presenting one of the possible outcomes as the "true" outcome had been added. As the possible outcomes were mutually exclusive, three of the four After groups received "true" outcomes that actually had not happened. Subjects in all groups were asked to (a) estimate the likelihood of occurrence of each of the four possible outcomes, and (b) evaluate the relevance of each datum in

Hindsight and foresight differ formally in the information available to the observer. The hindsightful judge possesses outcome knowledge, that is, he knows how things turned out. The foresightful judge does not. Although foresight usually implies looking at the future, in the absence of outcome knowledge, past and future events can be equally inscrutable.

The studies presented here ask two questions about the judgmental differences between hindsight and foresight: (a) How does receipt of outcome knowledge affect judgment? (b) How aware are people of the effects that outcome knowledge has on their perceptions? Answers to these questions will shed light on how people do learn and might better learn from history.

The two hypotheses explored are: (a) Reporting an outcome's occurrence increases its perceived probability of occurrence; and (b) people who have received outcome knowledge are largely unaware of its having changed their perceptions in the manner described in the first hypothesis. In combination, these two hypotheses indicate that reporting an outcome produces an unjustified increase in its perceived predictability, for it seems to have appeared more likely than it actually was.

Indirect support for the first hypothesis may be found in a variety of sources. For example, the

*This is a reprint of a paper that appeared in *Journal of Experimental Psychology: Human Perception and Performance* 1975, Volume 1, pages 288–299. Copyright © 1975, Psychological Association. Reprinted with permission.

Correspondence to:
B Fischhoff, Hebrew
University of Jerusalem,
Israel

the event description. In two of the subexperiments subjects were also asked to indicate the relative extent to which they relied on the passage and on outside information.

Instructions

The cover sheet of each questionnaire read: "In this questionnaire we are interested in knowing how people judge the likelihood of possible outcomes of social events. A passage describing an unfamiliar historical event appears below. We will ask you to evaluate the probability of occurrence of each of the four possible outcomes of the event (including that which actually happened—for After subjects) in the light of the information appearing in the passage."

A typical passage, as taken from Woodward's *The Age of Reform*,⁷ was:

"(1) For some years after the arrival of Hastings as governor-general of India, the consolidation of British power involved serious war. (2) The first of these wars took place on the northern frontier of Bengal where the British were faced by the plundering raids of the Gurkas of Nepal. (3) Attempts had been made to stop the raids by an exchange of lands, but the Gurkas would not give up their claims to country under British control, (4) and Hastings decided to deal with them once and for all. (5) The campaign began in November 1814. It was not glorious. (6) The Gurkas were only some 12 000 strong; (7) but they were brave fighters, fighting in territory well-suited to their raiding tactics. (8) The older British commanders were used to war in the plains where the enemy ran away from a resolute attack. (9) In the mountains of Nepal it was not easy even to find the enemy. (10) The troops and transport animals suffered from the extremes of heat and cold, (11) and the officers learned caution only after sharp reverses. (12) Major-General Sir D Ochterlony was the one commander to escape from these minor defeats" (pp 383–384)

The possible outcomes offered were: (a) British victory, (b) Gurka victory, (c) military stalemate with no peace settlement, and (d) military stalemate with a peace settlement. For After subjects, the appropriate outcome was appended to the passage in the form of an additional sentence such as, "The two sides reached a military stalemate but were unable to come to a peace settlement".

Following the passage, subjects were asked, "In the light of the information appearing in the passage, what was the probability of occurrence of each of the four possible outcomes listed below? (The probabilities should sum to 100%)." On the following page, each datum appeared on a separate line followed by a seven-point scale on which subjects were asked to indicate "how relevant or important each datum in the event description was in determining the event's outcome". The numbers in the passage above indicate the division into datum units. They did not appear in the passage presented to subjects.

Stimulus selection

Four different events were used to achieve greater generality for the results obtained: Event A, the British–Gurka struggle cited above; Event B, the near-riot in Atlanta, Georgia in July 1967, as described in the Kerner Commission Report on Civil Disorders⁸; and Events C and D, clinical cases reported by Albert Ellis.^{9*} For Events C and D, the word "social" in the instructions was replaced by "individual" and the word "historical" was deleted.

Several methodological considerations guided the event selection process: (a) The event should be sufficiently familiar

to permit intelligent responses and sufficiently unfamiliar to rule out the possibility of subjects knowing what really happened—especially those receiving false outcome reports. (b) Past events were used to allow provision of "true" outcomes to the After groups.† (c) The space of possible outcomes had to be readily partitionable. For Events B, C, and D, the set of outcomes was constructed to be mutually exclusive and exhaustive. Although this is not the case for Event A, pretests indicated that these four outcomes constituted an effective partition.

Subjects

Approximately equal numbers of subjects participated in each group in each subexperiment. Event A (Gurkas) was administered twice, once to a group of 100 English-speaking students recruited individually at The Hebrew University campus in Jerusalem and once to a class of 80 Hebrew-speaking subjects at the University of the Negev in Beer Sheva. Event B (riot) was administered to two separate classes at The Hebrew University, one containing 87 Hebrew-speaking psychology majors with at least one year's study of statistics and one of 100 Hebrew-speaking students with no knowledge of statistics. Event C (Mrs. Dewar) was administered to the 80 University of the Negev students; Event D (George) to the 100 Hebrew University students without statistics training.

Procedure

Questionnaires for the various experimental groups were distributed randomly. Subjects devoted 20–30 min to the completion of each questionnaire.

Results

Probability estimates

Table 1 presents the mean probability assigned to each outcome by subjects in each experimental group for each subexperiment.

Similar patterns of data emerged in the two subexperiments using Event A (differing in subjects' language) and in the two using Event B (differing in subjects' knowledge of statistics). For the sake of tabular brevity, only one subexperiment in each pair is presented.

The creeping determinism hypothesis predicts that After subjects told that a particular outcome has happened will assign it a higher probability than will Before subjects. Four outcomes reported to different groups in each of six subexperiments afford 24 opportunities to test the hypothesis. The critical comparisons are between the outlined diagonal cells (those indicating the mean probability assigned to an outcome by subjects for whom that outcome was reported to have happened) and the Before cell in the top row above them.

In each of the 24 cases, reporting an outcome increased its perceived likelihood of occurrence ($p < 0.001$; sign test). Twenty-two of these differences were individually significant ($p < 0.025$; median test). Thus the creeping determinism effect was obtained over all variations of subject population, event description, outcome reported, and truth of outcome reported. The differences between mean Before and After probabilities for reported outcomes ranged from 3.6% to 23.4%, with a mean of 10.8%. Slightly over 70% of After subjects assigned the reported outcome a higher probability than the mean assignment by the corresponding Before subjects.

No outcome was judged inevitable by any Before subject, whereas a small proportion (2.1%) of After subjects did assign

*Copies of these stimuli with the four offered outcomes will be supplied with all requested reprints. The permission granted by the Oxford University Press and Ronald Press to use these copyrighted materials is gratefully acknowledged.

†It might be wondered whether Before subjects might not behave like After subjects with reference to their predictions, which they know are actually postdictions. In a series of five experiments (Fischhoff, in press), we found that manipulating the temporal setting of possible outcomes has no effect on their perceived likelihood.

Table 1 Mean probabilities assigned to each outcome (Experiment 1)

Experimental group	n	Outcome provided	Outcome evaluated			
			1	2	3	4
Event A: British-Gurka struggle (English-speaking subjects)						
Before	20	None	33.8	21.3	32.3	12.3
After	20	1	57.2	14.3	15.3	13.4
	20	2	30.3	38.4	20.4	10.5
	20	3	25.7	17.0	48.0	9.9
	20	4	33.0	15.8	24.3	27.0
Event B: Near-riot in Atlanta (subjects with knowledge of statistics)						
Before	20	None	11.2	30.8	43.8	14.2
After	20	11	30.6	25.8	23.3	20.3
	20	2	5.5	51.8	24.3	18.5
	20	3	3.9	23.9	50.8	21.4
	20	4	16.7	31.9	23.4	27.9
Event C: Mrs Dewar in therapy						
Before	19	None	26.6	15.8	23.4	34.4
After	13	11	43.1	13.9	17.3	25.8
	17	2	26.5	23.2	13.4	36.9
	16	3	30.6	14.1	34.1	21.3
	17	4	21.2	10.2	22.6	46.1
Event D: George in therapy						
Before	17	None	27.4	26.9	39.4	6.3
After	18	11	33.6	20.8	37.8	8.0
	18	2	22.4	41.8	28.9	7.1
	20	3	20.5	22.3	50.0	7.3
	17	4	30.6	19.5	37.7	12.3

Note: The actual outcomes are numbers 1, 1, 4, and 2 for Events A, B, C, and D, respectively. Outlined cells are those with After probabilities of reported outcomes.

100% to reported outcomes. Evidently, most After subjects felt that in the light of the facts given in the description, other (unreported) outcomes were still possible (e.g. "The Gurkas had a 70% chance of winning, but the British still might have pulled it off"). Similarly, After subjects found a higher percentage of unreported outcomes to have been impossible (as indicated by a probability of 0%) than did Before subjects (11.5% versus 8.0%).

Another way to appraise the extent of creeping determinism is to translate mean Before probabilities into the form of a priori odds and the mean After probabilities for reported outcomes into a posteriori odds. The ratio of prior and posterior odds for outcome *I* provides a sort of average likelihood ratio for the impact of the datum "Outcome *I* did actually occur" (where the two hypotheses are "Outcome *I* occurs" and "Outcome *I* does not occur"). Over the 24 outcomes reported, these likelihood ratios varied from 1.2 to 3.5 ($M = 1.96$). Thus in the present sense, reporting an outcome's occurrence approximately doubles its perceived likelihood of occurrence.

Because the outcomes varied considerably in their mean Before probability (from 6.3% to 44.0%), reporting their occurrence may be seen as confirming (or disconfirming) subjects' expectations to varying degrees. There was a highly significant negative correlation ($\tau = 0.435$; $p < 0.001$) between the prior odds and likelihood ratios associated with reported outcomes (as computed in the preceding paragraph). Thus, the more unlikely an outcome report, the greater the impact it has.

Relevance judgments

Table 2 presents the mean relevance judgments for each datum in one subexperiment. Inspection reveals that the relevance attributed to any datum is highly dependent on which outcome, if any, subjects believe to be true. Some of these differences seem readily interpretable. For example, the fact that "the British officers learned caution only after sharp reverses" (datum no 11) was judged most relevant by subjects told of a British victory, and rather irrelevant by subjects told of a Gurka victory.

A less impressionistic analysis on the effects of outcome knowledge on relevance judgments proceeded in the following manner. For each subexperiment, a two-way (outcome reported \times datum evaluated) fixed model analysis of variance (ANOVA) was performed on subjects' judgments of data relevance. To accommodate the varying number of subjects in the experimental groups of subexperiments, the following procedure was adopted: The analysis was repeated three times to produce maximum, minimum, and middle solutions. For the maximum solution, subjects were randomly sampled from the smaller experimental groups and their responses duplicated, equating the size of all cells. For the minimum solution, subjects were randomly deleted from the larger groups until cell size was equated. For the middle solution, a combination of duplication and elimination was performed.

The same ANOVA was performed on the three sets of data. The results discussed hold for all three solutions. (a) In each subexperiment there was a significant Outcome Reported \times Datum Evaluated interaction reflecting the differential effect

Table 2 Mean data relevance judgments for Event A, Experiment 1 (Hebrew-speaking subjects)

Outcome reported	Datum number											
	1	2	3	4	5	6	7	8	9	10	11	12
None	4.50	5.11	4.22	5.78	4.50	6.00	5.50	5.44	4.39	4.56	4.28	5.56
British victory	4.78	4.44	5.28	4.83	4.61	4.44	4.61	4.56	5.72	5.33	5.78	4.11
Gurka victory	3.66	4.83	3.55	4.44	5.89	5.11	4.11	4.61	3.72	5.22	4.11	4.78
Stalemate												
with peace treaty	4.50	4.72	4.55	5.89	5.50	4.17	4.22	5.00	4.22	5.22	4.89	4.94
without peace treaty	4.94	5.50	4.39	5.11	5.33	5.11	4.78	4.39	4.17	3.72	4.50	4.61

of outcome knowledge on relevance judgments discussed above and shown in table 2. (b) Over the six subexperiments, only two weak outcome-reported effects emerged. Thus, there is no indication of an entire set of data having greater mean relevance for one outcome than another. (c) Datum-evaluated effects appear in all but one subexperiment. They reflect data perceived to be either relevant or irrelevant whatever happens (e.g. Hastings' decision to deal with the Gurkas "once and for all" (datum no 4] was universally perceived as relevant).

Reliance

Subjects in two of the subexperiments were asked to indicate with a number from 0% to 100% the extent to which they had relied on the material presented in the passage compared with general (outside) knowledge. In both cases Before subjects indicated relying significantly ($p < 0.05$; median test) more on the passage than did After subjects.

Discussion

Reporting an outcome's occurrence consistently increases its perceived likelihood and alters the judged relevance of data describing the situation preceding the event. Hindsight-foresight differences in perceived data relevance have also been noted by historians observing the creeping determinism effect. Consider, for example, Tawney¹⁰: "Historians give an appearance of inevitability to an existing order by dragging into prominence the forces which have triumphed and thrusting into the background those which they have swallowed up" (page 177). Or, consider Wohlstetter¹¹: "It is much easier after the event to sort the relevant from the irrelevant signals. After the event, of course, a signal is always crystal clear. We can now see what disaster it was signaling since the disaster has occurred, but before the event it is obscure and pregnant with conflicting meanings" (page 387).

How justified are the judgmental changes affected by receipt of outcome knowledge? It is hard to say, simply because there is no unequivocal way to assign probabilities or relevance weights to unique events like the British-Gurka struggle. If, for example, someone claims that there was no chance (or a 7% chance or a 98.6% chance) of a thermonuclear war during the 1960s, who can prove him wrong? Indeed, the only wrong estimate is that it was 100% likely.

Were such events well defined and reoccurring, the wisdom of increasing the postdicted probability of some reported outcomes would be readily apparent. Consider a judge who has previously encountered four British-Gurka-type struggles, each indistinguishable from the one used here, two of which were won by the Gurkas. Upon learning of another Gurka victory, he may properly update that outcome's predictive (Before) probability of 50% to a higher postdictive (After) probability. Hearing of a Gurka victory may also justify some change in relevance judgments by showing, for example, the true importance of British suffering from climatic extremes. It may also teach him something about the nature of 19th century colonialism—and thus change the sort of "laws" or reasons he uses in drawing inferences from the event description.

Thus, the judgmental changes which we have called creeping determinism could conceivably reflect what judges learn from outcome reports. The skepticism expressed in the anecdotal observations presented above,^{1-5, 10} however, suggests that this is not the case. In the light of these comments it appears that what passes for the wisdom of hindsight often contains heady doses of sophistry—that the perceived inevitability of reported outcomes is imposed upon, rather than legitimately inferred from, the available evidence.

As described in these accounts, postdictive likelihood estimates are exaggerated through a largely unconscious process evoked by receipt of outcome knowledge. How aware people are of the effect that outcome knowledge has on their perceptions was examined in Experiment 2. Aside from helping to clarify the nature of creeping determinism, these results have considerable intrinsic interest. Awareness is clearly crucial to knowing what one has learned from the past (i.e. from outcome knowledge). It may be necessary for learning from the past at all.

EXPERIMENT 2

Method

Design

Subjects were presented stimulus materials identical to those used in the After groups of Experiment 1, with each event description accompanied by a "true" outcome. They were asked to respond "as they would have had they not known the outcome". For each of the four events there were four After (ignore) groups, one receiving each possible outcome as true. If subjects are aware of the effect of outcome knowledge on their judgments, the responses of all of the After (ignore) groups should resemble those of that Before group in Experiment 1 which dealt with the same event. If After (ignore) subjects are completely unable to ignore the effect of outcome knowledge, their responses should resemble those of the After group in Experiment 1 which received the same outcome as "true".

Instructions

The cover of each test booklet read: "A number of short descriptions of real social and personal events appear below, each with a number of possible outcomes. On the basis of these data, we ask you to evaluate the likelihood of the outcomes listed. We thank you for your participation."

Each remaining page of the test booklet was identical to the corresponding page of the Experiment 1 booklet, except that each response section was preceded with the instruction to "answer as you would have had you not known what happened".

Subjects

Eighty members of an introductory statistics class at the University of the Negev participated.

Procedure

Questionnaires were randomly distributed to a single group of subjects. Each subject received one version of each of the four

Table 3 Mean probabilities assigned by subjects responding “as if you did not know what happened” (Experiment 2)

Experimental group	N	Outcome provided	Outcome evaluated			
			1	2	3	4
Event A: British-Gurka struggle						
Before	17	None	29.4	23.5	34.7	12.4
After (ignore)	20	11	29.8	27.4	24.9	18.4
	15	2	38.0	21.7	19.7	20.7
	18	3	22.1	31.8	31.9	14.3
	18	4	18.1	32.9	28.9	21.2
Event B: Near-riot in Atlanta						
Before	39	None	11.3	29.0	43.9	16.3
After (ignore)	17	11	24.6	27.0	28.3	19.8
	21	2	9.0	41.5	36.4	13.1
	20	3	6.3	24.5	43.5	25.8
	20	4	13.3	20.3	36.5	24.0
Event C: Mrs Dewar in therapy						
Before	19	None	26.6	15.8	23.4	34.4
After (ignore)	19	11	36.4	10.2	16.1	37.4
	19	2	24.7	28.8	15.5	31.9
	15	3	25.1	13.7	34.9	26.4
	20	4	18.3	12.3	21.8	52.8
Event D: George in therapy						
Before	17	None	26.4	26.9	39.4	6.3
After (ignore)	17	11	41.8	16.5	35.3	6.5
	18	2	24.6	35.9	32.4	7.0
	20	3	18.3	20.4	57.3	4.0
	18	4	21.0	21.1	38.4	19.6

Note: In each case the Before results are taken from the corresponding before (no outcome) group in Experiment 1 (subjects who actually responded not knowing what happened). Outlined cells are those with After (ignore) probabilities of reported outcomes.

different events. In a test booklet, Events A, B, and C alternated systematically as the first three events, with Event D (the least interesting) always appearing last. Order was varied to reduce the chances that subjects sitting in adjoining seats either copied from one another or discovered the experimental deception. All materials were in Hebrew. Questionnaires were anonymous.

Results

Probability estimates

Table 3 presents mean probability assignments by subjects in each of the After (ignore) groups along with the responses of the corresponding Before groups from Experiment 1. (The Hebrew-speaking group is used for Event A, the pooled responses of both relevant subexperiments for Event B.) The entries in each row will be called a profile. They indicate the probabilities subjects believed they would have assigned to the outcomes had they not known “what really happened”.

These reconstructed probabilities indicate no more than marginal awareness of the effects of outcome knowledge. In 13 of 16 cases the mean After (ignore) probability of the reported outcome was higher than the mean Before probability for the same event. For reported outcomes the mean Before–After(ignore) difference of 9.2% was slightly but not significantly less than the 10.8% mean Before–After difference in Experiment 1 ($p > 0.10$; Mann-Whitney U test).

The After (ignore) profiles closely resembled the corresponding After profiles. For 14 of 16 profiles, the mean absolute difference between corresponding cells was smaller for the After (ignore)–After comparison than for the relevant After (ignore)–Before comparison ($p < 0.002$; sign test). The median absolute difference between corresponding cells was 3.7% for After (ignore)–After, and 6.4% for After (ignore)–Before ($p < 0.001$; Mann-Whitney U test). There is no apparent reason, other than sampling error, for the weaker results obtained with Event A.

Relevance judgments

If After (ignore) subjects are able to ignore outcome knowledge, the outcome report they received should have no effect on their reconstructed relevance judgments. Instead, however, these relevance judgments clearly reflected the outcomes that After (ignore) subjects believed to have happened (but were instructed to ignore). For example, in Experiment 1 After subjects told of a British victory assigned substantially greater importance to the fact that “British officers learned caution only after sharp reverses” (datum no 11) than did Before subjects; those told of a Gurka victory assigned it slightly less importance. After (ignore) subjects in Experiment 2 who were asked to ignore a report of British victory believed that even without the report they would have perceived the relevance of datum no 11; those told to ignore a report of

Table 4 Mean probabilities assigned by subjects responding “as did other students who did not know what happened” (Experiment 3)

Experimental group	n	Outcome provided	Outcome evaluated			
			1	2	3	4
Event A: British-Gurka struggle						
Before	21	None	26.4	24.5	29.5	19.5
After (ignore)	17	11	39.4	22.4	20.3	18.8
	17	2	18.8	42.6	20.3	20.0
	22	3	31.1	21.2	26.6	20.0
	17	4	28.2	21.9	23.7	26.2
Event B: Near-riot in Atlanta						
Before	20	None	11.0	24.0	41.8	23.2
After (ignore)	17	11	15.0	24.7	36.5	23.8
	18	2	13.2	36.0	35.2	14.6
	19	3	4.8	22.5	51.1	21.6
	16	4	12.3	26.4	38.4	22.8
Event C: Mrs Dewar in therapy						
Before	21	None	19.6	15.9	24.0	40.5
After (ignore)	18	11	20.3	20.0	28.3	31.4
	18	2	31.9	23.3	14.8	30.0
	16	3	30.6	12.5	26.9	30.1
	19	4	12.5	20.4	22.6	44.4
Event D: George in therapy						
Before	19	None	30.7	22.4	39.2	7.8
After (ignore)	15	11	46.0	15.3	30.0	8.7
	16	2	22.5	36.6	34.1	6.9
	17	3	19.8	14.8	57.7	7.8
	16	4	23.5	18.3	40.3	17.8

Note: Outlined cells are those with After (ignore) probabilities of reported outcomes.

Gurka victory believed that they in foresight would have seen its irrelevance. When the relevance judgment ANOVA of Experiment 1 is repeated on the present data, this dependence is reflected in highly significant ($p < 0.0005$) Outcome Reported \times Datum Evaluated interactions.

Interestingly, for 128 of the 184 individual datum units evaluated by subjects in the four outcome groups of the four events After and After (ignore) relevance judgments were either both higher or both lower than the corresponding Before judgments (as was the case in the example, datum no 11, given above) ($z = 5.23$; sign test). There was no tendency for After and After (ignore) relevance judgments to be consistently higher or lower than Before relevance judgments, which might in itself account for this result.

Discussion

Experiment 1 showed that receipt of outcome knowledge affects subjects' judgments in the direction predicted by the creeping determinism hypothesis. Experiment 2 has shown that subjects are either unaware of outcome knowledge having an effect on their perceptions or, if aware, they are unable to ignore or rescind that effect. Both the relevance and the probability judgments of After (ignore) subjects suggest that subjects fail to properly reconstruct foresightful (Before) judgments because they are “anchored” in the hindsightful state of mind created by receipt of outcome knowledge.

It might be asked whether this failure to empathize with ourselves in a more ignorant state is not paralleled by a failure

to empathize with outcome-ignorant others. How well people manage to reconstruct the perceptions that others had before the occurrence of some event is a crucial question for historians, and indeed for all human understanding. The assumption that we clearly perceive how others viewed situations before receipt of outcome knowledge underlies most second guessing of their decisions. Experiment 3 examined this question.

EXPERIMENT 3

Method

Design

Subjects were presented with stimulus materials identical to those used in Experiments 1 and 2. They were asked to respond as had other student judges who had not known the true outcome. Before (others) subjects were not provided with any outcome knowledge. After (others) subjects received versions of the stimulus events with one of the four possible outcomes presented as the true outcome (what had actually happened). After (others) subjects' task was essentially to ignore outcome knowledge in order to respond like Before (others) subjects.

Instructions

The cover of each test booklet read: “Short descriptions of a number of real social and personal events appear below, each with several possible outcomes. These descriptions were presented to students of social science in other universities in

Israel. (However, they were not told which of the possible outcomes actually happened.) We will ask you to guess the judgments of these students regarding the likelihood of possible outcomes. We thank you for your participation."

The section in parentheses only appeared in the instructions for After (others) subjects. Each page of the test booklets was identical to the corresponding page of the Experiment 1 test booklets, except for the addition of a reminder: "Answer as you think other students (who did not know what happened) answered" before each response section.

Subjects

Ninety four members of an intermediate statistics class at the University of the Negev participated.

Results

Probability estimates

Table 4 presents mean probability assignments by subjects in each group. After (others) subjects' inability to ignore the effects of creeping determinism is clearly evident. For 14 of the 16 reported outcomes ($p < 0.002$; sign test) they attributed higher probabilities to outcome-ignorant others than did Before (others) subjects. As in Experiment 2, being told to ignore outcome knowledge slightly but not significantly ($p > 0.10$; Mann-Whitney U test) reduced its impact. The mean Before (others)–After (others) difference was 8.7% compared with the mean Before–After difference of 10.8% in Experiment 1.

Relevance judgments

After (ignore) subjects who had received different outcome reports attributed markedly different relevance judgments to the outcome-ignorant others. The dependence of the relevance judgments that they attributed on the outcome knowledge they were to ignore produced significant ($p < 0.01$) Outcome Reported \times Datum Evaluated interactions for each of the four events. Thus, After (ignore) subjects expected other subjects to have seen in foresight patterns of data relevance that they themselves only saw in hindsight.

Projection

Comparing tables 1 and 4 and tables 3 and 4, it is apparent that the entries in corresponding Before and Before (others) cells are quite similar, as are corresponding After (others) and After (ignore) cells. The mean absolute difference between entries in corresponding cells is 3.5% for the first comparison, 5.1% for the latter. This suggests that, when asked to respond like similar others, subjects respond as they believe they themselves would have responded in similar circumstances (i.e. by projection). Both the probability and relevance judgments of After (others) subjects more closely resembled those of After (ignore) and After subjects than those of Before (others) subjects.

Reasons

Some 87% of the subjects provided reasons for their judgments. Although content analysis of these reasons proved intractable, one interesting finding is that After (others) subjects offered consistently more reasons than Before subjects ($p < 0.05$; median test). In Experiment 1, After subjects reported relying more on outside information (as compared with the text) than did Before subjects. Perhaps in both cases knowing what happened facilitates knowing where to look for and what to accept as reasons.

GENERAL DISCUSSION

Finding out that an outcome has occurred increases its perceived likelihood. Judges are, however, unaware of the effect that outcome knowledge has on their perceptions. Thus,

judges tend to believe that this relative inevitability was largely apparent in foresight, without the benefit of knowing what happened.

In a fourth study¹² subjects were asked on the eve of former President Nixon's trips to China and the USSR (in early 1972) to estimate the probability of various possible outcomes of the visits (e.g. Nixon's meeting Chairman Mao, visiting Lenin's tomb, or announcing that the trips were successful). From 2 weeks to 6 months after the trips' completion, these same subjects were asked to remember as best they could their own original predictions. They were also asked to indicate for each event whether or not they believed that it had actually happened.

Results showed that subjects remembered having given higher probabilities than they actually had to events believed to have occurred and lower probabilities to events that hadn't occurred. Their original predictions showed considerable overestimation of low probabilities—that is, too many events that they judged to be extremely unlikely or impossible did occur. The probability judgments that they remembered, however, consistently underestimated low probabilities. Indeed, almost no events to which they remembered assigning low probabilities were perceived to have occurred.

Thus, undiagnosed creeping determinism not only biases people's impressions of what they would have known without outcome knowledge, but also their impressions of what they themselves and others actually *did* know in foresight.

Explanations

The simplest hypothesis regarding the manner in which judges process outcome knowledge suffices to account for these results. Assume that upon receipt of outcome knowledge judges immediately assimilate it with what they already know about the event in question. In other words, the retrospective judge attempts to make sense, or a coherent whole, out of all that he knows about the event. The changes in relevance judgments could reflect such assimilative meaning adjustment.

Assimilation of this type would tend to induce creeping determinism for judges using any of the techniques for producing subjective probability estimates appearing in Tversky and Kahneman's compendium.¹³ Judges using the heuristic of "representativeness" perceive outcomes as likely when they match or represent the dominant features of the situation that produced them. Assimilation of outcome knowledge should certainly increase the perceived "fit" between the reported outcome and the situation that preceded it. A second heuristic leads judges to evaluate an outcome's likelihood by the relative "availability" of scenarios leading to its occurrence and non-occurrence. The judge who knows what happened, and has adjusted his perceptions in the light of that knowledge, may well find it difficult to imagine how things could have turned out otherwise.

An alternative mode of explanation focuses on structural differences between the tasks of hindsight and foresight. Judges possessing outcome knowledge may, for example, tend to reverse their temporal perspective and produce scenarios that proceed backward in time, from the outcome to the preceding situation. Such scenario retrodiction may effectively obscure the ways in which events might not have taken place, much as solving a maze backwards can obscure the ways in which one might have got lost entering from the beginning.

Receipt of outcome knowledge may also restructure the task of judges using the "anchoring and adjustment" heuristic.¹³ Judges may estimate the likelihood of a reported outcome by initially assigning it 100%, the most salient possible value, and then looking for reasons to adjust downward from there. Adjustment from initial values is typically inadequate and would produce creeping determinism with this task.

Of these explanations, those based on assimilation most readily account for the underestimation of creeping determinism found in Experiments 2 and 3 and in the study by Fischhoff and Beyth.¹² Making sense out of what one is told about the past seems so natural and effortless a response that one may be unaware that outcome knowledge has had any effect at all on him. Judges who are aware that outcome knowledge has affected their perceptions still face the unenviable task of reconstructing their foresightful state of mind. “Undiagnosed creeping determinism” would characterize the responses of subjects who, in reconstruction, were unable to adequately unanchor themselves from the perspective of hindsight.

Implications

In the short run, failure to ignore outcome knowledge holds substantial benefits. It is quite flattering to believe, or lead others to believe, that we would have known all along what we could only know with outcome knowledge—that is, that we possess hindsightful foresight. In the long run, however, unperceived creeping determinism can seriously impair our ability to judge the past or learn from it.

Consider a decision maker who has been caught unprepared by some turn of events and who tries to see where he went wrong by recreating his preoutcome knowledge state of mind. If, in retrospect, the event appears to have seemed relatively likely, he can do little more than berate himself for not taking the action which his knowledge seems to have dictated. He might be said to add the insult of regret to the injury inflicted by the event itself. When second guessed by a hindsightful observer, his misfortune appears to have been incompetence, folly, or worse.

In situations where information is limited and indeterminate, occasional surprises—and resulting failures—are inevitable. It is both unfair and self-defeating to castigate decision makers who have erred in fallible systems, without admitting to that fallibility and doing something to improve the system. According to historian Roberta Wohlstetter,¹¹ the lesson to be learned from American surprise at Pearl Harbor is that we must “accept the fact of uncertainty and learn to live with it. Since no magic will provide certainty, our plans must work without it” (page 401).

When we attempt to understand past events, we implicitly test the hypotheses or rules we use to both interpret and anticipate the world around us. If, in hindsight, we systematically underestimate the surprises which 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. Elaboration on this point as well as speculation on how hindsight can be improved may be found in Fischhoff¹⁴ and Fischhoff and Beyth.¹⁵

.....

Preparation of this report was supported by the Advanced Research Projects Agency of the Department of Defense (ARPA Order 2449) and was monitored by the Office of Naval Research under Contract No. N00014-73-C-0438 (NR 197026). The research reported constitutes part of a doctoral dissertation submitted to The Hebrew University of Jerusalem. I am deeply indebted to Amos Tversky, Daniel Kahneman, Paul Slovic, Ruth Beyth, and Sarah Lichtenstein for their contributions to this project. The detailed comments of two anonymous reviewers on a previous draft are gratefully acknowledged.

REFERENCES

- 1 **Florovsky G.** The study of the past. In: Nash RH, ed. *Ideas of history*. Vol 2. New York: Dutton, 1969.
- 2 **Tversky A, Kahneman D.** Belief in the law of small numbers. *Psychol Bull* 1971;**76**:105–10.

- 3 **Lofland R.** *Deviance and identity*. Englewood Cliffs, NJ: Prentice-Hall, 1969.
- 4 **Rosenhan D.** On being sane in insane places. *Science* 1973;**79**:250–2.
- 5 **Schur E.** *Labelling deviant behavior*. New York: Harper & Row, 1971.
- 6 **Walster E.** “Second guessing” important events. *Human Relations*, 1967;**20**:239–50.
- 7 **Woodward EL.** *Age of reform*. London: Oxford University Press, 1938.
- 8 **National Advisory Commission.** *Kerner Commission report on civil disorders*. New York: Bantam, 1968.
- 9 **Ellis A.** Psychosexual and marital problems. In: Berg LA, Pennington LA, eds. *An introduction to clinical psychology*. New York: Ronald, 1966.
- 10 **Tawney RH.** *The agrarian problem in the sixteenth century*. New York: Franklin, 1961.
- 11 **Wohlstetter R.** *Pearl Harbor: warning and decision*. Stanford, Calif.: Stanford University Press, 1962.
- 12 **Fischhoff B, Beyth R.** “I knew it would happen”—remembered probabilities of once-future things. *Organizational Behav Human Performance* 1975;**13**:1–16.
- 13 **Tversky A, Kahneman D.** Judgment under uncertainty: heuristics and biases. *Science* 1974;**185**:1124–31.
- 14 **Fischhoff B.** Hindsight: thinking backward? *Oregon Res Inst Res Monogr* 1974;**14**: No. 1.
- 15 **Fischhoff B, Beyth R.** Failure has many fathers (review of “Victims of groupthink” by I Janis). *Oregon Res Inst Res Bull* 1974;**14**: No 17.

COMMENTARY

FOR THOSE CONDEMNED TO LIVE IN THE FUTURE

“In situations where information is limited and indeterminate, occasional surprises—and resulting failures—are inevitable. It is both unfair and self-defeating to castigate decision makers who have erred in fallible systems without admitting to that fallibility and doing something to improve the system”.³ (page 298)

A common goal of many of the people concerned with the “error problem” in medicine is ultimately to improve the system. However, there is a great debate about the best strategy for accomplishing this goal. The extreme poles in this debate might be caricatured as the *error elimination strategy*¹ and the *safety management strategy*.² The error elimination strategy tends to rely heavily on hindsight. This strategy tries to reconstruct the history of events in order to identify the “causes” of the errors. It is believed that, by systematically eliminating the causes of error, the system is made increasingly safer. The safety management strategy tends to rely more on foresight. This strategy tries to integrate past experiences to better understand the evolving work ecology. This includes trying to anticipate the functional constraints that shape the opportunities and risks associated with work and the information that might best specify those constraints to decision makers. This approach believes that making the relevant constraints more salient to decision makers is the most promising direction for increasing safety.

Fischhoff’s work^{3,4} on the “hindsight bias” suggests that either strategy is vulnerable to errors, and re-reading this important work should provide a healthy dose of humility for people on both sides of the debate. Woods *et al*⁵ summarized this vulnerability as follows:

“Given knowledge of outcome, reviewers will tend to *simplify* the problem-solving situation that was actually faced by the practitioner. The dilemmas, the uncertainties, the tradeoffs, the attentional demands, and double binds faced by practitioners may be missed or under-emphasized when an incident is viewed in hindsight. . . . Possessing knowledge of the outcome, because of the hindsight bias, trivializes the situation confronting the practitioners and makes the correct choice seem crystal clear.” (page 7–8)

The error elimination strategy is particularly vulnerable because this approach depends on the ability to accurately reconstruct the past in order to identify causal chains. In fact, Fischhoff⁴ suggests that the very notion of “causality” may be a symptom of hindsight bias in which an “outcome seems a more or less inevitable outgrowth of the reinterpreted situation” in light of hindsight (page 343). This is one reason why the safety management strategy prefers to focus on “constraints” rather than “causes”. However, the safety management strategy is also vulnerable to hindsight bias in that judgments about the natural salience of information are also affected by the hindsight bias—such that the natural salience of relevant information may be overestimated in the light of hindsight. Despite this clear danger, I tend to believe that the safety management strategy, while fallible, provides the best way forward for improving system safety.

One lesson of Fischhoff’s work is that the human memory system is not designed to accurately reconstruct the past (as is explicitly assumed by much of the research on memory which measures memory solely in terms of its ability to accurately remember past events), but rather the human memory system is designed to adapt to the future. This adaptation involves “making sense” of the past in order to better anticipate the future. This is clearly not a perfect system:

“‘Making sense’ out of what we are told about the past is, in turn, so natural that we may be unaware that outcome knowledge has had any effect on us. Even if we are aware of there having been an effect, we may still be unaware of exactly what it was. In trying to reconstruct our foresightful state of mind, we will remain anchored in our hindsightful perspective, leaving the reported outcome too likely looking.”³ (page 343)

However, the “biases” in judging the past may have positive as well as negative implications when projected into the future. Consider the quote from the analysis by Dominguez⁶ of the conversion decision in the context of laparoscopic cholecystectomy shown in box 1. Perhaps the error was an inevitable result of the uncertainties associated with surgery. And perhaps this surgeon is castigating himself too severely, given the inevitability of errors in this type of environment. However, I suspect that this surgeon will move into the future with a greater awareness of the potential for danger and that he will be a much better (safer) surgeon as a consequence. In this sense, the adaptation process may involve a migration toward the boundaries of safety. The consequence of crossing a boundary (an error) may be an overcorrection in favor of caution (in this sense it is a bias—incorrectly feeling that he should have seen the error coming). While clearly not optimal in a statistical sense, this may lead to a system that sacrifices in favor of safety! Such a system—one that errs in the direction of caution—may well be more likely to survive in an uncertain world than one that optimizes around a particular history [that might reflect both real and imagined (luck) constraints].

Box 1 Quote from Dominguez⁶

“I would be trying to shoot an intraoperative cholangiogram before I’d go ahead and clip that but then again that’s just my own bias from my own previous experience from having a ductile injury. In that girl, [she] had a fairly acute disease, wasn’t quite as bad looking as this but everything was fine until 5 days post-op when she came back to the office just still puking her guts out. And I’d just destroyed her hepatic duct, her common hepatic duct, because I hadn’t realized where we were and that was an error on my part and I had been fooled by the size of her cystic duct. The stone, it had been a good size stone, it had worked its way down chronically to the cystic duct enlarging it so that it looked like the infundibulum of the gallbladder and then at the common duct junction I thought the common duct was the cystic duct so I went ahead and clipped it, divided and then started cauterizing. Well when you cauterize up through there you’ve got the hepatic duct line right behind it and I eventually burned that part. If you talk to any other surgeons who’ve had that kind of an injury, I mean I lost sleep for several nights over that. It’s one of those things that haunt you and you hate it, you just hate it.”

Fischhoff’s research suggests that it is impossible to reconstruct the past accurately. Any approach to the medical error problem that depends on an accurate reconstruction of the past is therefore doomed to fail. However, it is also important to note that the past is only an imperfect predictor of the future. Even a perfect memory of past events will not allow an unambiguous projection of the future. A system that studies the past with an eye to the future coupled with a healthy dose of humility and caution may therefore provide the best path forward. Today, the safety management strategy reflected in the cognitive systems engineering approach^{2,7} offers the best hope to a medical system that is destined to live in the future.

J M Flach

Department of Psychology, Wright State University, Dayton, OH 45435, USA; john.flach@wright.edu

REFERENCES

- 1 Senders J, Moray N. *Human error: cause, prediction, and reduction*. Hillsdale, NJ: Erlbaum, 1991.
- 2 Rasmussen J, Svendung I. *Proactive risk management in a dynamic society*. Raddningsverket, Sweden: Swedish Rescue Services Agency, 2000.
- 3 Fischhoff B. Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty. *J Exp Psychol: Human Perception and Performance* 1975;1:288–99.
- 4 Fischhoff B. For those condemned to study the past: heuristics and biases in hindsight. In: Kahneman D, Slovic P, Tversky A, eds. *Judgments under uncertainty: heuristics and biases*. Cambridge: Cambridge University Press, 1982: 335–51.
- 5 Woods DD, Johannesen LJ, Cook RI, et al. *Behind human error: cognitive systems, computers, and hindsight*. (CSERIAC SOAR 94-01) Wright-Patterson, AFB, OH: Crew Systems Ergonomics Information Analysis Center, 1991.
- 6 Dominguez C. *First, do no harm: expertise and metacognition in laparoscopic surgery*. Dissertation, Wright State University, Dayton, OH, 1996.
- 7 Rasmussen J, Pejtersen AM, Goodstein LP. *Cognitive systems engineering*. New York: Wiley, 1994.