

AD-A008 580

HINDSIGHT/FORESIGHT: THE EFFECT OF
OUTCOME KNOWLEDGE ON JUDGMENT UNDER
UNCERTAINTY

Baruch Fischhoff

Oregon Research Institute

Prepared for:

Office of Naval Research
Advanced Research Projects Agency

7 April 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Oregon Research Institute Eugene, Oregon		2a. REPORT SECURITY CLASSIFICATION Unclassified	
		2b. GROUP	
3. REPORT TITLE Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty			
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Technical Report			
5. AUTHOR(S) (First name, middle initial, last name) Baruch Fischhoff			
6. REPORT DATE April 7, 1975		7a. TOTAL NO. OF PAGES 37	7b. NO. OF REFS 20
8a. CONTRACT OR GRANT NO. N00014-73-C-0438		9a. ORIGINATOR'S REPORT NUMBER(S)	
b. PROJECT NO. NR (197-026)			
c. ARPA Order No. 2449		9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)	
d.			
10. DISTRIBUTION STATEMENT This document has been approved for public release and sale; its distribution is unlimited.			
11. SUPPLEMENTARY NOTES		12. SPONSORING MILITARY ACTIVITY Office of Naval Research Code 455 Arlington, Virginia 22217	
13. ABSTRACT One major difference between historical and non-historical 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 over-estimated 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.			

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
JUDGMENT						
PREDICTION						
DECISION MAKING						
INFORMATION PROCESSING						
UNCERTAINTY						

OREGON RESEARCH INSTITUTE

**HINDSIGHT ≠ FORESIGHT
THE EFFECT OF OUTCOME KNOWLEDGE
ON JUDGMENT UNDER UNCERTAINTY**

Baruch Fischhoff

Oregon Research Institute

**Oregon Research Institute
Research Bulletin**

Volume 14 Number 13

1974

Hindsight ≠ Foresight:

The Effect of Outcome Knowledge on Judgment Under Uncertainty

Hindsight and foresight differ formally in the information available to the observer. The hindsightful judge possesses outcome knowledge, i.e., 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? and (b) how aware are people of the effects which outcome knowledge has on their perceptions? Answers to these questions will shed light on how people do, and how they might better, learn from history.

It should come as no surprise that historiographers and philosophers of history have discussed these questions in considerable detail. The centrality of outcome knowledge in determining the nature of historical judgment may be seen in the following remarks by Hexter (1961):

What is really different (with historical judgment) is that the historian knows what is going to happen next The historian who resolutely refused to use the insight that his own particular time gave him would not be superior to his fellows. He would be foolish, betraying a singular failure to grasp what history is." (p. 10)

It is evidently outcome knowledge which enables historical judges to "put events into perspective," and confidently second-guess the decisions of their predecessors.

What appears to be the historians' consensual answer to the question of what outcome knowledge does to judgment is succinctly expressed by Georges Florovsky (1968):

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." (p. 369)

If there is a tendency to see whatever is reported to have happened as having been relatively inevitable, an apt name might be "creeping determinism"--in contrast with philosophical determinism, the conscious belief that whatever happens has to happen. These two types of determinism are essentially independent. The philosophical determinist may believe that a reported event was inevitable, since whatever happens is, by definition, inevitable. He may, however, still be surprised by its occurrence. Indeed, he may well set for himself the task of researching the situation until its inevitable character becomes apparent. The creeping determinist may well be totally unaware of the raging debate over historical inevitability and free will. He perceives reported events as having been more or less bound to happen, simply as a matter of course.

Phenomena resembling creeping determinism have been noted by psychologists as well as historians. One example is Tversky and Kahneman's (1971) "law of small numbers," the belief that data which were observed more or less had to be observed. A second 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 (Lofland, 1969, Rosenhan, 1973; Schur, 1971). 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 which will increase their perceived predictability and avoidability (Walster, 1966). All of this evidence for creeping determinism is, however, either indirect, imprecise, unsystematic (anecdotal), or confounded by motivational and emotional issues.

The validity of "creeping determinism" as a description of the effects of outcome knowledge on judgment was directly tested in Experiment 1 below. Experiments 2 and 3 examined subjects' awareness of the effects of outcome knowledge on judgment.

Experiment 1

Method

Design. The six sub-experiments described below 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 sub-experiment, the Before group read a brief (150 word) description of an 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 which 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 the event description. In two of the sub-experiments subjects were also asked to indicate the relative extent to which they relied upon the passage and upon 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 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.

From The age of reform, by E. L. Woodward,
Oxford University Press, London, 1938,
pp. 383-384.

The possible outcomes offered were: 1) British victory, 2) Gurka victory, 3) military stalemate with no peace settlement, 4) 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 upon 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 "data." 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. They dealt with: (A) the British-Gurka struggle cited above; (B) the near-riot in Atlanta, Georgia, in July, 1967, as described in the Kerner Commission Report on Civil Disorders

(National Advisory Commission, 1968); (C) and (D) clinical cases reported by Albert Ellis (1967). 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 in order 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 so as to be mutually exclusive and exhaustive. Although this is not the case for Event (A), pretests indicated that the four outcomes provided constituted an effective partition.

Subjects. Approximately equal numbers of subjects participated in each group in each sub-experiment. 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 Sheba. 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 statistics-less Hebrew University students.

Procedure. Questionnaires for the various experimental groups were distributed randomly. Subjects devoted 20-30 minutes 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 sub-experiment. The Before groups appear in the top (no outcome provided) line of the results for each event. The creeping determinism hypothesis

Insert Table 1 about here

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 sub-experiments 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 provided as what had really 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 ($p < .001$; sign test). Twenty-two of these differences were individually significant ($p < .025$; median test). The creeping determinism effect has, thus, been 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 100% to reported outcomes. 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 1.1 to 3.7 (mean 1.96). Thus, in the present sense, reporting an outcome's occurrence approximately doubles its perceived likelihood of occurrence.

Relevance judgments. Table 2 presents the mean relevance judgments for each datum in one sub-experiment. Inspection reveals that the relevance attributed to any datum is highly dependent upon 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 Number 11) was judged most relevant by subjects told of a British victory, and rather irrelevant by subjects told of a Gurka victory.

Insert Table 2 about here

A less impressionistic analysis on the effects of outcome knowledge in relevance judgment proceeded in the following manner. For each sub-experiment, a two-way (Outcome Reported x Datum Evaluated) fixed model ANOVA was performed on subjects' judgments of data relevance. As noted above, for several sub-experiments, the number of subjects varied over experimental groups. Rather than use one of the suggested (e.g., Winer, 1971) approximate ANOVA solutions for unequal N, which would have raised serious computational difficulties, the following procedure was adopted: The analysis was repeated three times to produce "maximum," "minimum," and "middle" solutions. For the former, 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.

They were: (a) In each sub-experiment, there was a significant Outcome Reported x Datum Evaluated interaction, reflecting the differential effect of outcome knowledge on relevance judgments discussed above and shown in Table 2. (b) Over the six sub-experiments, only two weak Outcome Reported effects emerged. Thus, there is no indication that the whole set of data may be more relevant for one outcome than another. (c) Datum Evaluated effects appear in all but one sub-experiment. They reflect data perceived to be either relevant or irrelevant whatever happens

(e.g., Hasting's decision to deal with the Gurkas "once and for all" [Datum No. 4] was universally perceived as relevant).

Reliance. Subjects in two of the sub-experiments 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 ($p < .05$; median test) more on the passage than did After subjects.

Discussion

Hindsight subjects consistently perceived reported outcomes (whether true or not in fact) as having been more likely to occur than did their foresightful counterparts. Outcome knowledge roughly doubled the odds favoring the reported event's occurrence. Interestingly, however, subjects almost never assigned 100% probability to the reported outcome. Evidently, they felt that in the light of the facts given in the description other outcomes were still possible (e.g., "The Gurkas had a 70% chance of winning, but the British still might have pulled it off"), consistent with creeping, but not philosophical, determinism.

Relevance judgments, too, were affected by which outcome, if any, was reported. Such hindsight-foresight differences in perceived data relevance have also been noted by historiographers observing the creeping determinism effect. Consider, for example, Tawney (1961): "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" (p. 177). Or, Wohlstetter (1962),

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. It comes to the observer embedded in an atmosphere of 'noise,' i.e., in the company of all sorts of information that is useless and irrelevant for predicting the particular disaster." (p. 387)

Knowing that something has happened clearly increases its perceived inevitability, as well as restructures judges' perceptions of what they know about it. How justified are these changes? It is hard to say, simply because there are no objective probabilities associated with unique events like the British-Gurka struggle. Consider another example: If someone claims that there was no chance (or a 7% chance or a 98.6% chance) of a thermonuclear war during the 1960's, who can prove him wrong? Indeed, the only wrong estimate is that it was 100% likely.

In many well-defined situations, the wisdom of some increase in the postdicted probability of reported outcomes is readily apparent. Consider the task of having to predict the outcome of successive draws from an urn containing an unspecified proportion of red and blue balls. Two red and two blue balls have been drawn. The fifth ball drawn is blue. Prior to the fifth draw, the best estimate of the probability of a blue ball being drawn was 50%. Following the drawing, that probability is properly evaluated as having been greater than 50%. That is to say, knowing that an outcome has occurred (a blue ball has been drawn) increases its probability of occurrence. Analogously, consider a judge who has in his lifetime encountered four British-Gurka type struggles, two of which were won by the "British." Upon learning of a "Gurka" victory, he may properly

update that outcome's predictive (Before) probability of 50% to a higher postdictive (After) probability. The latter example also demonstrates the justifiability of outcome knowledge affecting relevance judgments. Hearing of a Gurka victory may, for example, alert a judge to 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 anecdotal observations presented above (e.g., Florovsky, Schur, Tawney), however, all suggest that this is not the case. They indicate that what passes for the wisdom of hindsight often contains heavy doses of sophistry; that the perceived inevitability of reported outcomes is imposed upon, rather than legitimately inferred from, the available evidence. Tversky and Kahneman (1971) have empirically demonstrated the extent to which research psychologists exaggerate the likelihood of results which they have obtained.

As described in these accounts, the retrospective exaggeration of likelihood is a largely unconscious process evoked by receipt of outcome knowledge. Subjects' degree of awareness of the effects of outcome knowledge on their perceptions is 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 2Method

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 know 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 Negev participated.

Procedure. Questionnaires were randomly distributed to a single group of subjects. Each subject received one version of each of the four 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 group from Experiment 1. The entries in each row will be called a "profile." They indicate the probabilities subjects believe that they would have assigned to the outcomes had they not known "what really happened."

Insert Table 3 about here

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 ($p > .10$; 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 ($p < .002$; sign test). The median absolute difference between corresponding cells was 3.7% for

After (ignore)-After, and 6.4% for After (ignore)-Before ($p < .001$; Mann-Whitney U test).

Relevance judgments. If After (ignore) subjects are able to ignore outcome knowledge, the outcome report which they received should have no effect on their reconstructed relevance judgments. Instead, however, these relevance judgments clearly reflected the outcomes which 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 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 ($p < .0005$) Outcome Reported x Datum Evaluated interactions.

Interestingly, for 128 of the 184 individual data 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 they 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 which 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 ($p < .002$; 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 ($p > .10$; Mann-Whitney 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.

Insert Table 4 about here

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 which they attributed on the outcome knowledge which they were to ignore produced significant ($p < .01$) Outcome Reported x 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 which they themselves only saw in hindsight.

Projection. Comparing Tables 1 and 4, and 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 (other) 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 ($p < .05$; 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 which 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 study (Fischhoff & Beyth, 1975), 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). Two weeks to six 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.

The results showed that subjects remembered having given higher probabilities than they actually had to events believed to have occurred and lower probabilities to events which hadn't. Their original predictions showed considerable over-estimation of low probabilities, i.e., too many events which they judged to be extremely unlikely or impossible did occur. The probability judgments which 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 virtually any reasonable technique to produce subjective probability estimates. For example, consider Tversky and Kahneman's recent (1974) compendium of heuristics for producing subjective probabilities. One such heuristic is "representativeness," by which outcomes are deemed likely according to the extent that they "represent" the dominant features of the situation which produced them. The process often evokes but a simple associative matching. Assimilation of outcome knowledge should certainly increase the perceived "fit" between reported outcomes and the situations which preceded them. 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 ways in which receipt of outcome knowledge may restructure the judgmental task. For example, outcome knowledge might tend to reverse judges' temporal perspective and encourage the production of scenarios which proceed backward in time, from the outcome to the situation of the event description. Such scenario retrodiction may effectively obscure the ways in which events might not have

taken place, much as solving a maze backward can obscure the ways in which one might have gotten lost entering from the beginning.

A second restructuring explanation could apply to judges using a third Tversky-Kahneman (1974) heuristic. The "anchoring and adjustment" heuristic directs judges to perform estimation tasks by initially selecting some particularly salient response value and then "adjusting" up or down from there. Typically, these adjustments are inadequate, and the judge remains anchored in his initial value. In retrodiction, judges may tend to adjust downward from 100% for reported outcomes and upward from 0% for unreported outcomes. Creeping determinism would result from remaining overly anchored in these initial values.

The assimilation or outcome knowledge processing explanations seem preferable to these restructuring explanations in their ability to readily account for the underestimation of creeping determinism found in Experiments 2 and 3 and the Fischhoff and Beyth (1975) study. "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 had some effect on their perceptions still face the unenviable task of having to reconstruct the foresightful state of mind which preceded receipt of outcome knowledge. The accuracy of their reconstructed probability estimates depends, of course, upon the accuracy of these reconstructions. The relevance judgments in Experiments 2 and 3 reflect their inability to accomplish that reconstruction. Judges who remain anchored in their present, outcome knowledge-laden state of mind should produce reconstructed judgments bearing the mark of creeping determinism.

None of the present results indicate that creeping determinism is affected by either the truth or the likelihood of the reported outcome. It remains to be seen if this is generally the case. Sherif and Howland (1961) have

suggested that highly inconsistent or unintegratable data tend to be contrasted or discounted rather than assimilated. If this is the case, then extremely unlikely (inconsistent) outcome reports should not produce creeping determinism and subsequent unawareness. Other possible determinants of the extent of creeping determinism are discussed in Fischhoff and Beyth (1975) and Fischhoff (1974).

Implications. It has frequently been found that people are unduly confident in their predictive abilities (e.g., Alpert & Raiffa, Note 1; Fischhoff & Beyth, 1975; Kahneman & Tversky, 1973). Thus, there are typically many outcomes whose occurrence or non-occurrence should constitute relative "surprises" for them. With receipt of outcome knowledge, however, their "surprisingness" vanishes or at least diminishes. In this light, failure to ignore outcome knowledge may be seen to hold 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 to say, that we possess hindsightful foresight. Failure to adequately perceive the surprises which the past holds and has held for us can, however, seriously impair our ability to judge that 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 pre-outcome 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. A more appropriate lesson from the experience might be that the data at his disposal are quite indeterminate, and that he should be ready

for a substantial number of surprises and make his plans accordingly. If, however, all events which occur appear to have been fairly likely, there is little reason to prepare for surprises.

Experiment 3 suggests that there are similar processes at work in second-guessing the decisions of others. Consider looking backward at an historical figure whose plans have gone awry because of unanticipated events. In retrospect, the eventualities which foiled his plan seem to have been relatively likely. Whereas his failure may actually reflect his misfortune in encountering an unavoidable surprise, it appears to be a matter of incompetence, folly, or worse. Yet in many situations where information is limited, occasional surprises--and resulting failures--are inevitable. It seems to be 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. A classic case in point is the demotion of Admiral Kimmel, Naval C. O. at Pearl Harbor, following the Japanese attack in World War II. After reviewing extensive documentary evidence, the historian Roberta Wohlstetter (1962) comes to the conclusion that it is unreasonable to expect any American military or political figure to have anticipated the attack, in the light of the information at his disposal beforehand. She concludes that the lesson to be learned from investigating the disaster is to "accept the fact of uncertainty and learn to live with it. Since no magic will provide certainty, our plans must work without it" (p. 401).¹

Finally, the judge who looks at the past and finds that it holds relatively few surprises for him is essentially denying that he has very much to learn from it. If he has invested resources in obtaining outcome knowledge about some event and upon receipt feels that he really did know that it would happen, those resources seem wasted. When judges attempt to understand past events, they implicitly test the hypotheses or rules they use to interpret and anticipate the world around them. A past which is inordinately barren of surprises provides an inordinately weak test of the hypotheses applied to it (Popper, 1965). The judge who perceives a relatively surprise-free past may feel little compulsion to change the hypotheses which guided him in viewing that past. Thus, the very outcome knowledge which gives him the feeling that he understands what the past was all about may prevent him from learning anything from it.

Reference Notes

1. Alpert, M., & Raiffa, H. Training probability assessors: A progress report. Unpublished manuscript, Harvard Business School, 1969.

References

- Ellis, A. Psychosexual and marital problems. In L. A. Berg & L. A. Pennington (Eds.), An introduction to clinical psychology. New York: Ronald, 1966.
- Fischhoff, B. Aspects of historical judgment: The effects of temporal perspective and outcome knowledge on judgments under uncertainty. Unpublished doctoral dissertation, The Hebrew University, Jerusalem, Israel, 1974.
- Fischhoff, B. Hindsight: Thinking backward? Oregon Research Institute Research Monograph, 1974, 14, No. 1.
- Fischhoff, B., & Beyth, R. Failure has many fathers. (Review of Victims of groupthink by Irving Janis). Oregon Research Institute Research Bulletin, 1974, 14, 17.
- Fischhoff, B., & Beyth, R. "I knew it would happen"--remembered probabilities of once-future things. Organizational Behavior and Human Performance, 1975, in press.
- Florovsky, G. The study of the past. In R. H. Nash (Ed.), Ideas in history, (Vol. II). New York: Dutton, 1969.
- Hexter, J. Reappraisals in history. Evanston: Northwestern University Press, 1961.
- Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.
- Lofland, R. Deviance and identity. Englewood Cliffs: Prentice-Hall, 1969
- National Advisory Commission on Civil Disorders Report. New York: Bantam, 1968.
- Popper, K. Conjectures and refutations. New York: Harper, 1968.
- Rosenhan, D. On being sane in insane places. Science, 1973, 79, 250-252.

- Schur, E. Labelling deviant behavior. New York: Harper & Row, 1971.
- Sherif, H., & Hovland, O. Social judgment: Assimilation and contrast effects in communication and attitude change. New Haven: Yale University Press, 1961.
- Tawney, R. H. The agrarian problem in the sixteenth century. New York: Franklin, 1961.
- Tversky, A., & Kahneman, D. Belief in the law of small numbers. Psychological Bulletin, 1971, 76, 105-110.
- Tversky, A., & Kahneman, D. Judgment under uncertainty: heuristics and biases. Science, 1974, 185, 1124-1131.
- Walster, E. Assignment of responsibility for an accident. Journal of Personality and Social Psychology, 1966, 3, 73-79.
- Winer, B. J. Statistical principles in experimental design (2nd ed.). New York: McGraw-Hill, 1971.
- Wohlstetter, R. Pearl Harbor: Warning and decision. Stanford: Stanford University Press, 1962.
- Woodward, E. L. Age of reform. London: Oxford University Press, 1938.

Footnotes

Preparation of this report was supported by the Advanced Research Projects Agency of the Department of Defense (ARPA Order No. 2449) and was monitored by ONR under Contract No. N00014-73-C-0438 (NR 197-076). The research reported constitutes part of a doctoral dissertation submitted to The Hebrew University of Jerusalem. I am indebted to Amos Tversky, Daniel Kahneman, Paul Slovic, Ruth Beyth, and Sarah Lichtenstein for their contributions to this project. Reprints are available from the author at Oregon Research Institute, P. O. Box 3196, Eugene, Oregon 97403.

¹Elaboration on this point as well as speculation on how hindsight can be improved may be found in Fischhoff (1974) and Fischhoff and Beyth (1974).

Table 1
Mean Probabilities Assigned (Exp. 1)

		Outcome Evaluated:				
		1 ^a	2	3	4	N
Event A: British - Gurka struggle English-speaking Subjects	Outcome Provided:					
	None	33.8	21.3	32.3	12.3	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	20	
		Outcome Evaluated:				
		1 ^a	2	3	4	N
Event A: British - Gurka struggle Hebrew-speaking Subjects	Outcome Provided:					
	None	29.4	23.5	31.7	12.4	17
	1	33.3	20.9	32.2	13.6	18
	2	20.0	41.3	30.3	8.3	18
	3	16.9	31.6	38.3	13.3	20
4	18.2	24.6	28.2	28.9	17	
		Outcome Evaluated:				
		1 ^a	2	3	4	N
Event B: Near-riot in Atlanta Subjects with Knowledge of Statistics	Outcome Provided:					
	None	11.2	30.8	43.8	14.2	20
	1	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	20	
		Outcome Evaluated:				
		1 ^a	2	3	4	N
Event B: Near-riot in Atlanta Subjects without Knowledge of Statistics	Outcome Provided:					
	None	11.8	26.1	44.0	18.3	19
	1	18.8	33.5	28.3	19.5	18
	2	11.3	35.3	33.9	19.6	15
	3	4.0	22.7	50.9	22.4	16
4	14.5	37.9	21.2	26.9	18	

Event D: George in Therapy

		Outcome Evaluated:				
		1	2 ^a	3	4	N
Outcome Provided:	None	27.4	26.9	39.4	6.3	17
	1	33.6	20.8	37.8	8.0	18
	2	22.4	41.8	28.9	7.1	18
	3	20.5	22.3	50.0	7.3	20
	4	30.6	19.5	37.7	12.3	17

Event C: Mrs. Dewar in Therapy

		Outcome Evaluated:				
		1	2	3	4 ^a	N
Outcome Provided:	None	26.6	15.8	23.4	34.4	19
	1	43.1	13.9	17.3	25.8	13
	2	26.5	23.2	13.4	36.9	17
	3	30.6	14.1	34.1	21.3	16
	4	21.2	10.2	22.6	46.1	17

Note. Outlined cells are those with the After probabilities of repeated outcomes.

^a Indicates true outcome.

Table 2

Mean Data Relevance Judgments for Event A, Experiment 1 (Hebrew-speaking subjects)

Datum Number:	1	2	3	4	5	6	7	8	9	10	11	12
Outcome Reported:												
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 Win	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 Win	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												
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
Stalemate												
No 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

Table 3

Mean Probabilities Assigned by Subjects, responding "as if you did not know what happened" [Exp. 2]

Event A: British-Gurka struggle		Outcome evaluated:				
		1	2	3	4	N
Outcome provided:	None	29.4	23.5	34.7	12.4	17
	1	29.8	27.4	24.9	18.4	20
	2	38.0	21.7	19.7	20.7	15
	3	22.1	31.8	31.9	14.3	18
	4	18.1	32.9	28.9	21.2	18

Event B: Near-riot in Atlanta		Outcome evaluated:				
		1	2	3	4	N
Outcome provided:	None	11.3	29.0	43.9	16.3	39
	1	24.6	27.0	28.3	19.8	17
	2	9.0	41.5	36.4	13.1	21
	3	6.3	24.5	43.5	25.8	20
	4	13.3	20.3	36.5	24.0	20

Event C: Mrs. Dewar in therapy		Outcome evaluated:				
		1	2	3	4	N
Outcome provided:	None	26.6	15.8	23.4	34.4	19
	1	36.4	10.2	16.1	37.4	19
	2	24.7	28.8	15.5	31.9	19
	3	25.1	13.7	34.9	26.4	15
	4	18.3	12.3	21.8	52.8	20

Event D: George in therapy		Outcome evaluated:				
		1	2	3	4	N
Outcome provided:	None	26.4	26.9	39.4	6.3	17
	1	41.8	16.5	35.3	6.5	17
	2	24.6	35.9	32.4	7.0	18
	3	18.3	20.4	57.3	4.0	20
	4	21.0	21.1	38.4	19.6	18

Note: In each case, the "none" results are taken from the corresponding Before (no outcome) group in Exp. 1 (subjects who actually responded not knowing what happened).

Table 4

Mean Probabilities Assigned by Subjects responding "as did other students who did not know what happened" [Exp. 3]

Event A: British-Gurka struggle

		Outcome evaluated:				
		1	2	3	4	N
Outcome provided:	None	26.4	24.5	29.5	19.5	21
	1	39.4	22.4	20.3	18.8	17
	2	18.8	42.6	20.3	20.0	17
	3	31.1	21.2	26.6	20.0	22
	4	28.2	21.9	23.7	26.2	17

Event B: Near-riot in Atlanta

		Outcome evaluated:				
		1	2	3	4	N
Outcome provided:	None	11.6	24.0	41.8	23.2	20
	1	15.0	24.7	36.5	23.8	17
	2	13.2	36.0	35.2	14.6	18
	3	4.8	22.5	51.1	21.6	19
	4	12.3	26.4	38.4	22.8	16

Event C: Mrs. Dewar in therapy

		Outcome evaluated:				
		1	2	3	4	N
Outcome provided:	None	19.6	15.9	24.0	40.5	21
	1	20.3	20.0	28.3	31.4	18
	2	31.9	23.3	14.8	30.0	18
	3	30.6	12.5	26.9	30.1	16
	4	12.5	20.4	22.6	44.4	19

Event D: George in therapy

		Outcome evaluated:				
		1	2	3	4	N
Outcome provided:	None	30.7	22.4	39.2	7.8	19
	1	46.0	15.3	30.0	8.7	15
	2	22.5	36.6	34.1	6.9	16
	3	19.8	14.8	57.7	7.8	17
	4	23.5	18.3	40.3	17.8	16