ORGANIZATIONAL BEHAVIOR AND HUMAN PERFORMANCE 13, 1–16 (1975)

# "I Knew It Would Happen"

## Remembered Probabilities of Once–Future Things[1]

BARUCH FISCHHOFF AND RUTH BEYTH

*The Hebrew University of Jerusalem, Israel*

Judges who had estimated the likelihood of various possible outcomes of President Nixon's trips to Peking and Moscow were unexpectedly asked to remember, or reconstruct in the event that they had forgotten, their own predictions some time after the visits were completed. In addition, they indicated whether or not they thought that each event had in fact occurred. Remembered–reconstructed probabilities were generally higher than the originally assigned probabilities for events believed to have occurred and lower for those which had not (although the latter effect was less pronounced). In their original predictions, subjects overestimated low probabilities and underestimated high probabilities, although they were generally quite accurate. Judging by their reconstructed–remembered probabilities, however, subjects seldom perceived having been very surprised by what had or had not happened. These results are discussed in terms of cognitive "anchoring" and possible detrimental effects of outcome feedback.

Most judges engaged in predictive tasks are presumably interested in improving their own performance. A logical first step in this direction is to evaluate the accuracy of their own past predictions in the light of what has subsequently happened. In order to be evaluated, these predictions must, of course, either be remembered per se, reconstructed on the basis of what judges remember having known about the event at the time of the original prediction or estimated on the basis of the event's post facto likelihood. The effectiveness of the evaluation process depends in part upon the veracity of these remembered or reconstructed predictions. Little, if anything, however, is known about the extent of systematic or random error in remembered and reconstructed predictions.

For some time, we have been studying the differences between pre-

[1] We are indebted to Prof. D. Kahneman, Prof. A. Tversky, and Prof. P. Slovic for comments on a previous version of this article. Requests for reprints should be sent to Baruch Fischhoff, Oregon Research Institute, P. O. Box 3196, Eugene, OR 97403.

dictive and postdictive (post facto) judgment (Fischhoff, 1974). Some of our results suggest that remembered predictions may be consistently biased in a manner explored in the study reported here. In particular, we have found that events reported to have happened tend to be assigned higher postdictive than predictive probabilities, i.e., reporting an event's occurrence increases its perceived inevitability. This tendency was named "creeping determinism" as it expresses a tendency toward determinism which is nonetheless short of that advocated by theories of historical inevitability (Berlin, 1954; Carr, 1961).

To summarize briefly the more detailed discussion appearing in Fischhoff (1974), creeping determinism seems most readily understood through consideration of the demand characteristics of the retrospective judge's task. Typically, judges are called upon to predict the future and to "make sense" out of the past. Attempting to understand why a particular outcome occurred seems, among other things, to increase the salience of data and reasons which can be integrated into coherent explanatory patterns. Unintegratable data tend to be forgotten, deemphasized, or reinterpreted to fit the dominant explanation. Postdicted probabilities are estimated on the basis of such "updated" sets of event-descriptive data. Given this mode of outcome knowledge processing, it can readily be shown that judges using any one of a number of judgmental heuristics for probability estimation (Tversky & Kahneman, 1974), will produce postdictive probabilities higher than the corresponding predictive estimates. ·

Although the name creeping determinism has clear pejorative connotations, in many cases the postdictive probability of events which have happened is justifiably higher than the corresponding predictive probability. Consider sampling from an urn containing an unspecified proportion of red and blue balls. Of the first four balls drawn, two are red and two are blue. The fifth ball drawn is blue. Prior to the fifth drawing, the probability of a blue ball was 50%, following the drawing, that probability is properly evaluated as >50%, i.e., the postdicted probability is higher than the predicted probability. It is, however, our conviction that in real life such retrospective increases frequently constitute little more than facile reductions in the "surprisingness" of what has happened. Rather than reflecting some "wisdom of hindsight," they seem to reflect what might be called a "knew it all" attitude.

Verification of such suspicions is only possible in the relatively rare (for real-life judges) instances in which a well-defined model of the data-generating process is available. A model allows the calculation of predictive and postdictive probabilities, as well as actual data-diagnosticity. As our primary interest was the judgment of unique events, we turned our attention to an interesting side effect of creeping determinism whose non-normative status is readily established and whose consequences are of considerable interest in their own right. In particular, we found that judges appear to be generally incapable of assessing the changes in their judgments induced by possession of outcome knowledge. In a further experiment in the same series, subjects were provided with outcome knowledge regarding various events and asked to respond as they would have "had they not known what happened." These subjects responded more like subjects who knew what had happened than those who did not, i.e., they believed that without outcome knowledge they would have assigned significantly higher probabilities to events reported to have happened than did other, truly outcome-ignorant subjects.

Extrapolating these results of between-subject comparisons, we hypothesized that judges may also tend to remember having assigned higher probabilities than they actually did to events which they subsequently found to have happened (and vice versa for events which did not). That is to say, the remembered or reconstructed probability of an event will tend to be larger than the probability originally assigned to it if the event is believed to have occurred, and smaller if it is believed not to have occurred. The present study tests this hypothesis.

## METHOD

### *Design*

The effect of outcome knowledge on prediction recall-reconstruction was tested in the following fashion: Subjects estimated the probability of a number of events whose outcome would be known within a fixed period of time (Prediction). Some time after the time period had elapsed, these same subjects were asked to remember or reconstruct their own predictions as accurately as possible (Prediction Memory). No mention was made of the Prediction Memory task at the time of the original prediction. Finally, subjects indicated whether they thought that each event had or had not occurred on an Information questionnaire which was distributed immediately after the collection of the Prediction Memory questionnaire. The purpose of the Information questionnaire was to ascertain what each subject believed had happened. It was a fortuitous inclusion, as subjects frequently disagreed with one another and with "usually reliable" press reports. The order of the Prediction Memory and Information questionnaires was such as to obscure the purpose of the experiment. Reversing their order might be expected to heighten the hypothesized effect by increasing the salience of what had and had not occurred. Events used were possible outcomes of President Nixon's visits to China and the USSR in the first half of 1972.

*Subjects*

Participants in the present experiment were students in an Advanced Methodology class and an Introductory Psychology class at the Hebrew University of Jerusalem, and an Intermediate Statistics class at the University of the Negev, Beer Sheba, Israel. All responses were collected on questionnaires distributed in classrooms. Each class was visited twice, once to distribute the Prediction (Before) Questionnaire, and once to distribute the Prediction Memory and Information (After) Questionnaires. Due to absenteeism, only 55% of the subjects who completed Prediction questionnaires were present at the administration of the Prediction Memory and Information questionnaires. Prediction Memory and Information questionnaires were mailed to all subjects who had completed the Prediction questionnaires but had not been present for the second questionnaire administration (and who had provided addresses). Their responses are designated as Group V. Although most subjects were Hebrew-speaking, English versions of all questionnaires were available for those who requested them.

The five experimental groups were:

I. Predictions relating to the China trip made shortly before the visit (2.20.72; $N = 53$); recollection shortly after (3.5.72; $N = 29$). Subjects: Advanced Methodology class.
II. Predictions relating to the China trip made shortly before the visit (2.20.72; $N = 87$); recollection long after (6.11.72; $N = 41$). Subjects: Introductory Psychology class.
III. Predictions relating to the USSR trip made shortly before the visit (5.23.72; $N = 34$); recollection shortly after (6.6.72; $N = 26$). Subjects: students in Intermediate Statistics class.
IV. Predictions relating to the USSR trip long before the visit; recollection shortly after. Subjects and dates: same as Group II.
V. Predictions relating to the USSR trip long before the visit (2.20.72; $N = 52$); recollection long after (approximately 10.15.72 (week of mailing); $N = 23$). Subjects in Advanced Methodology and Introductory Psychology classes not present in class during the administration of After questionnaires.

*Instructions*

The following is an example of Prediction instructions:

> President Nixon is currently on the eve of his visit to China. The possible outcomes of this visit are still in doubt. Commentators have offered a number of possible outcomes, some of which are presented below. We would like to have you estimate the probability of each of these eventualities coming

to pass. That is to say, we would like you to give each outcome a probability value from 0–100%.

0%: There is no chance of the outcome happening.

100%: The outcome is certain to happen.

These instructions were appropriately adapted for each group.

The following is an example of Prediction Memory instructions:

> As you remember, about two weeks ago, on the eve of President Nixon's trip to China, you completed a questionnaire by providing probabilities for the occurrence of a number of possible outcomes of the trip. We are presently interested in the relation between the quality of people's predictions and their ability to remember their predictions. For this reason, we would like to have you fill out once again the same questionnaire which you completed two weeks ago, giving the *same probabilities which you gave then.* If you cannot remember the probability which you then assigned, give the probability which you would have given to each of the various outcomes on the eve of President Nixon's trip to China.

The answer sheets of the Prediction and Prediction Memory questionnaires differed only in the order of the possible outcomes. This was done to prevent the intrusion of possible incidental memory (e.g., a subject might just happen to recall what he predicted for the first item of the Prediction questionnaire, or that his last three estimates had been 0%).

Instructions for the Information questionnaires read:

> One of our hypotheses is that memory for probability judgments is influenced by information relating to what actually did or did not happen. The enclosed questionnaire is identical to that which you just completed, except that this time we would like you to indicate whether or not each event occurred. Beside each possible outcome you will find the four following possibilities:
>
> A: I believe that the event occurred and was publicized.
> B: I believe that the event did not occur.
> C: I believe that the event occurred and was not publicized.
> D: I don't know.
>
> For each possible outcome of the visit, please circle that possibility which best suits the information at your disposal. Circle only one possibility. This is not a test of your political knowledge, and consultation with your neighbor is only liable to distort the results.

Students who had not filled out Before questionnaires, but happened to be present at the administration of the After questionnaires, were asked to produce reconstructed probabilities, giving "the probabilities which you would have given had you been asked on the eve of President Nixon's visit to China (the USSR)." Sixty-four subjects gave such postdictions of the China trip outcomes (Groups VII, VIII) and 27 for the USSR trip outcomes (Group VI).

*Outcomes*

Fifteen possible outcomes of each trip were presented. They were chosen so as to (1) cover most areas of potential activity (especially those of interest to our subjects), and (2) elicit a wide range of probability values. Representative events[2] are:

China: (1) The USA will establish a permanent diplomatic mission in Peking, but not grant diplomatic recognition; (2) President Nixon will meet Mao at least once; (3) President Nixon will announce that his trip was successful;

USSR: (1) A group of Soviet Jews will be arrested attempting to speak with President Nixon; (2) The USA and the USSR will agree to a joint space program.

## RESULTS

Each subject produced two probabilities for each of 15 possible outcomes: one before the relevant trip, $p_1$, and one after, $p_2$, as well as an answer for the knowledge of outcome question (A, B, C, or D). Thus, for each outcome, it could be determined whether each subject's responses supported the hypothesis about the relation between prediction memory and outcome knowledge (+), contradicted the hypothesis (−), or were irrelevant to the hypothesis (0). Response sets were evaluated:

(+) if $p_1 < p_2$ and the subject reported A or C (event happened),
if $p_1 > p_2$ and the subject reported B (event did not happen);
(−) if $p_1 < p_2$ and the subject reported B (event did not happen),
if $p_1 > p_2$ and the subject reported A or C (event happened);
(0) if $p_1 = p_2$ and the subject reported A, B, C, or D,
if the subject reported D and any $p_1$ and $p_2$,
if $p_1 = 100\%$ and the subject reported A or C (the natural ceiling of the probability measure makes it impossible for $p_2$ to be higher than $p_1$ in accordance with the hypothesis),
if $p_1 = 0$ and the subject reported B (corresponding floor effect).

Inclusion of the relatively few instances in which subjects reported (A or C; $p_1 = 100$; $p_2 < 100$) or (B; $p_1 = 0$; $p_2 > 0$) would not have appreciably altered the results presented below and appears to be an unduly conservative policy.

Only 8.3% of all Information responses fell in Category C. As the response patterns for Category C (events which happened but which had not been published) were quite similar to those for Category A (events

[2] Details on the individual events used and subjects' responses to them may be obtained from the authors.

which had happened and had been publicized), the two categories were combined to obtain more stable estimates. Category A–C refers, then, to all events judged to have happened, whether publicized or not.

For each judge, the numbers of hypothesis-consistent (+) and hypothesis-inconsistent (−) responses were counted within each Information category (A–C and B). If the number of (+)'s was greater than the number of (−)'s in a category, the subject was considered to be "hypothesis-supporting" for that category. The number of hypothesis-supporting and nonsupporting subjects in each experimental group appear in Table 1. The total number of (+)'s and (−)'s for each subject was also counted and similarly evaluated. The number of hypothesis-supporting and nonsupporting subjects were compared by a sign test and the result translated to a normal variate to facilitate comparison (a negative sign indicates that majority of subjects were nonhypothesis-supporting).

### *Main Effect*

In general, the results presented in Table 1 provide support for the notion that receipt of outcome knowledge may be associated with systematic biases in prediction recollection and reconstruction. For two-thirds of the subjects, misremembered and misreconstructed probabilities generally erred in the anticipated direction ($z = +3.12$; sign test). Three-quarters generally remembered or reconstructed having assigned higher probabilities than they actually had to events which they believed had happened (A–C) ($z = +5.12$; sign test). However, only 57% generally reported lower $p_2$ for events believed not to have happened (B) ($z = +1.08$; sign test). The difference in the proportions of hypothesis-supporting subjects for events which were and were not perceived to have occurred was significant ($z = +2.52$). Most of this difference arose from Groups II and IV (composed of the same subjects responding to different stimuli) where 60% of subjects tended to reconstruct–remember higher probabilities for events perceived not to have happened.

Several determinants of effect size may be ascertained. One is the period of time which elapsed between the estimation and memory tasks. Regarding events believed to have happened, in Groups II, IV, and V, where 3–6 months separated the tasks, some 84% of subjects evidenced the predicted bias; compared with 67% for Groups I and III, where but 2 weeks elapsed ($z$ for difference in proportions $= +1.95$). For events perceived not to have occurred, this trend is reversed, owing largely to the negative result with Groups II and IV. For Groups I and III (short time period), 64% of subjects supported the hypothesis; for Groups II, IV, and V, 51% ($z$ for difference $= +1.29$).

Another determinant is the a priori likelihood ($p_1$) of events, evidently

TABLE 1

NUMBER OF GENERALLY HYPOTHESIS-SUPPORTING (+) AND NONSUPPORTING (−) SUBJECTS FOR EACH EXPERIMENTAL GROUP AND INFORMATION RESPONSE CATEGORY (A–C, B, OVERALL)

| Group I ($N$ = 29)<br>China: shortly before, shortly after | | | |
|---|---|---|---|
| | A–C | B | overall |
| + | 17 | 15 | 17 |
| − | 7 | 8 | 11 |
| $z$ | +1.84 | +1.25 | +.94 |
| %+ | 70.8 | 65.2 | 60.7 |

| Group II ($N$ = 41)<br>China: shortly before, long after | | | |
|---|---|---|---|
| | A–C | B | overall |
| + | 30 | 15 | 26 |
| − | 7 | 22 | 14 |
| $z$ | +3.62 | −.99 | +1.74 |
| %+ | 81.1 | 40.5 | 65.0 |

| Group III ($N$ = 26)<br>USSR: shortly before, shortly after | | | |
|---|---|---|---|
| | A–C | B | overall |
| + | 14 | 13 | 15 |
| − | 8 | 8 | 8 |
| $z$ | +1.06 | +.87 | +1.25 |
| %+ | 63.6 | 61.9 | 65.2 |

| Group IV ($N$ = 41)<br>USSR: long before, shortly after | | | |
|---|---|---|---|
| | A–C | B | overall |
| + | 30 | 15 | 19 |
| − | 5 | 24 | 13 |
| $z$ | +4.04 | −1.28 | +.88 |
| %+ | 85.7 | 38.5 | 59.4 |

| Group V ($N$ = 23)<br>USSR: long before, long after | | | |
|---|---|---|---|
| | A–C | B | overall |
| + | 17 | 15 | 18 |
| − | 3 | 6 | 4 |
| $z$ | +2.51 | +1.76 | +2.88 |
| %+ | 85.0 | 71.4 | 81.8 |

| Groups I, II, III, V combined ($N$ = 119)<br>Groups I, III, IV, V combined ($N$ = 119)[a] (in parentheses) | | | |
|---|---|---|---|
| | A–C | B | overall |
| + | 78 (78) | 58 (58) | 76 (69) |
| − | 25 (23) | 44 (46) | 37 (36) |
| $z$ | +5.12 (+5.37) | +1.29 (+1.08) | +3.58 (+3.12) |
| %+ | 75.7 (77.2) | 56.9 (55.8) | 67.2 (65.7) |

Group VI ($N$ = 27)
USSR: no before, shortly after

| | A–C | B | overall |
|---|---|---|---|
| + | 20 | 10 | 16 |
| − | 4 | 14 | 8 |
| $z$ | +3.10 | −.61 | +1.43 |
| %+ | 83.3 | 41.7 | 66.7 |

Group VII ($N$ = 27)
China: no before, long after

| | A–C | B | overall |
|---|---|---|---|
| + | 16 | 13 | 14 |
| − | 5 | 13 | 12 |
| $z$ | +2.23 | .00 | +.19 |
| %+ | 76.2 | 50.0 | 53.8 |

Group VIII ($N$ = 37)
China: no before, long after

| | A–C | B | overall |
|---|---|---|---|
| + | 18 | 22 | 24 |
| − | 13 | 12 | 13 |
| $z$ | +.71 | +1.54 | +1.64 |
| %+ | 58.1 | 64.7 | 64.9 |

Groups VI, VIII combined ($N$ = 64)
Groups VII, VIII combined ($N$ = 64)[b] (in parentheses)

| | A–C | B | overall |
|---|---|---|---|
| + | 38 (34) | 32 (35) | 40 (38) |
| − | 17 (18) | 26 (25) | 21 (25) |
| $z$ | +2.69 (+2.10) | +.66 (+1.16) | +2.33 (+1.51) |
| %+ | 69.1 (65.4) | 55.2 (58.3) | 65.6 (60.3) |

[a] As Groups II and IV consist of the same subjects responding to different questions, they were not combined.
[b] As Groups VI and VII consist of the same subjects responding to different questions, they were not combined.

reflecting the floor and ceiling imposed by the natural upper and lower limits of the probability measure. These limits might be expected to attenuate the present effect in a fairly straightforward fashion. For unlikely events, which generally did not happen, $p_1$ may have been so low that there was little "room" (given random fluctuations and the slight regression effect noted below) for $p_2$ to be consistently lower when the event did not happen (and conversely for likely events). Thus, the less extreme the initial probability, the more "room" there is for the anticipated change and the stronger the effect which may be expected. Considering the number of hypothesis-supporting responses as a measure of the size of the effect for individual events, we find a substantial quadratic (inverted-$U$) relationship between effect size and median $p_1$ of individual events ($F(2,60) = 9.304$, $p < .0005$).

The size and nature of the effect may be further understood by comparing typical (median) probabilities remembered and reconstructed for the various possible $p_1$ values. Figure 1 presents this information for A–C and B events separately. For A, B, and C events combined (not shown), the regression line of median $p_2$ values on $p_1$ is $\tilde{y} = 9.4 + .87x$ ($r = .99$; $df = 19$; $p < .0005$). The fact that the slope is less than one

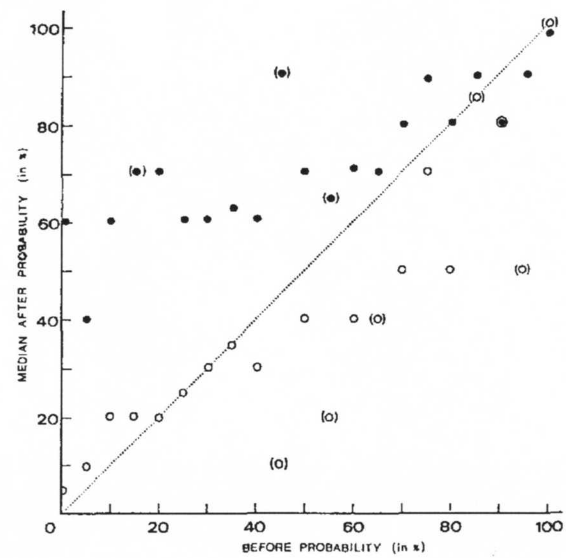


FIG. 1. Median After probabilities of events assigned Before probability of $X\%$; presented separately for events perceived to have happened (darkened circles) and perceived not to have happened (open circles). Parentheses indicate median determined by five or fewer judgments.

Regression lines:

| | |
|---|---|
| ● Perceived to have happened (A–C) | $\tilde{y} = 54 + .37x$; $r = .80$; $df = 19$, |
| ○ perceived not to have happened (B) | $\tilde{y} = 7 + .63x$; $r = .85$; $df = 19$, |
| □ all events (not shown) | $\tilde{y} = 9 + .87x$; $r = .99$; $df = 19$. |

may be interpreted as a mild regression toward the mean effect to the extent that the $p_2$ are measures of the original probability estimate, rather than response memories. A regression effect would mitigate against the research hypothesis, resulting in higher $p_2$ for unlikely events (which tended not to occur), and lower $p_2$ for likely events (which tended to occur). Quite a different picture is obtained by considering A–C and B events separately. The two separate regression lines are highly distinct. For events perceived to have occurred, $p_2$ tended to be higher than $p_1$ for all but the largest $p_1$ values ($\tilde{y} = 54 + .37x$; $r = .80$; $df = 19$; $p < .0005$). For events perceived not have happened, $p_2$ tended to be lower than $p_1$ for all but the smallest $p_1$ values ($\tilde{y} = 7.0 + .63x$; $r = .85$; $df = 19$; $p < .0005$). One summary description is that remembered–reconstructed probability estimates for A–C and B events "regressed" about highly distinct means. At the other extremes, few A–C events were perceived to have been very unlikely ($x = 0\%$ intercept equal to 54%); few B events were perceived to have been very likely ($x = 100\%$ intercept equal to 70%).

### *After-Only Subjects*

All of the subjects considered above (Groups I–V) explicitly stated their predictions regarding the various trip outcomes ($p_1$). It might be wondered whether this act improved their memories for cue configurations and the inferences drawn from the $p_1$ and consequently reduced the vulnerability of their reconstructions to systematic biases. A partial answer to this question may be derived from the $p_2$ responses of those subjects merely asked to reconstruct the predictions which they would have provided had they been asked prior to the trips (Groups VI–VIII). In the absence of $p_1$ responses for these After-Only subjects, their reconstructed probabilities ($p_2$) were compared with the median a priori probabilities ($p_1$) given by the other (Before and After) subjects, on the assumption that these probabilities were close to what they would have responded had they been asked earlier. This mode of analysis is, of course, somewhat less sensitive than the strictly within-subject analysis reported above. Its power is also reduced by smaller sample size. Nevertheless, it is instructive that essentially the same results were derived (see Table 1). Over 60% of these After-Only subjects generally supported the hypothesis; about two-thirds did so for events perceived to have occurred, and somewhat over half for events perceived not to have happened. Thus, there is little evidence that expressly stating predictions reduces the vulnerability of reconstructions to systematic biases of the type under consideration here.

*Surprisingness*

If a "surprise" is defined as the occurrence of an unlikely event or the nonoccurrence of a likely event, one result of the bias considered here is to reduce the "surprisingness" of the past; the occurrence of an event increases its reconstructed probability and makes it less surprising than it would have been had the original probability been remembered. The surprisingness of a set of events in the light of predictions may be ascertained by evaluating the percentage of events assigned various probabilities perceived to occur. For a perfectly calibrated set of judgments, $X\%$ of those events assigned $X\%$ probability of occurrence would actually occur. The percentage of events assigned $X\%$ probability which were perceived to have occurred was calculated separately for the $p_1$ and $p_2$ responses of Before and After subjects, and for the $p_2$ responses of After-Only subjects. These results appear in Fig. 2. Due to subjects' tendency to use round numbers and the very large quantities of data needed to obtain stable estimates of occurrence rates, only 12 probability categories were used: 0–4%, 5–9%, 10–19%, 20–29%, 30–39%, 40–49%, 50–59%, 60–69%, 70–79%, 80–89%, 90–99%, and 100%. Roughly equal numbers of the 1921 Before, 1909 Before with After, and 832 After-Alone predictions fall into each category.

A considerably smaller proportion of unlikely events ($p \leq 30\%$) and a slightly larger proportion of likely events were perceived to have oc-

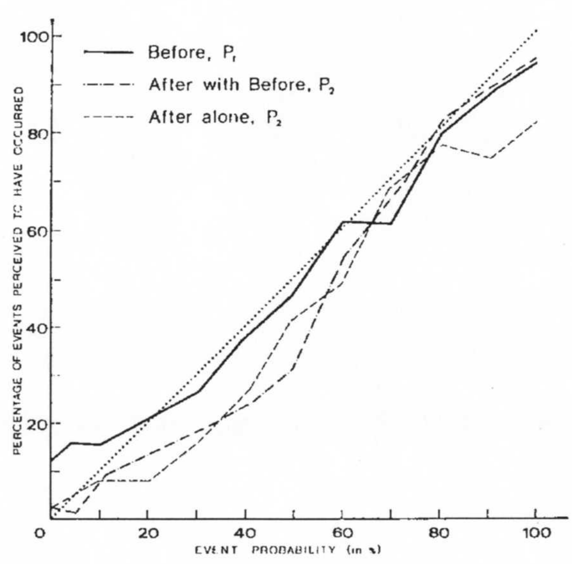


FIG. 2. Percentage of events perceived to have occurred of those assigned $X\%$ probability of occurrence.

curred in retrospect ($p_2$) than in the light of $p_1$. That is to say, subjects reconstructed–remembered having been less surprised by the events which did and did not occur in the course of President Nixon's trip than they really should have been (judging by their own predictions). The original predictions were, in general, quite well calibrated, except with regard to unlikely events where they met too many substantial surprises; 10% of the events assigned 0% probability of occurrence were perceived to have occurred, as well as 16% of those assigned 5% probability. In contrast, in the light of $p_2$, there were too *few* big surprises. Very few events with remembered–reconstructed probabilities less than 30% were perceived to have occurred. All groups somewhat underestimated very likely probabilities ($90\% \leq p \leq 100\%$), i.e., encountered too *few* likely occurrences. Thus, although very few events which happened had low reconstructed probabilities, there were still some events which did not happen with high reconstructed probabilities. This is consistent with the differential effect obtained with events which did and did not happen.

## DISCUSSION

Why are remembered probabilities biased in the manner shown above? Two explanations, each applicable to a different tactic which subjects might adopt in retrieving $p_1$, seem particularly attractive. Both reflect the notion of judgmental "anchoring and adjustment." As described in Slovic (1972, p. 16), "In this process, a natural starting point is used as a first approximation to the judgment, an anchor so to speak. This anchor is then adjusted to accommodate the implications of additional information. Typically, the adjustment is a crude and imprecise one which fails to do justice to the importance of additional information." Given the original creeping determinism results (Fischhoff, 1974), it may be assumed that After judges have a mental set, or "state of mind," in which reported outcomes tend to appear more likely than they did before their occurrence.

The prediction memory judge intent upon retrieving $p_1$ may try to do so by first retrieving his own previous (Before) state of mind, and then reestimating $p_1$. That is to say, he might ask himself, "Considering what I knew then, how likely did the event seem?" He may, however, find himself so "anchored" in his present (After) state of mind that his previous state is beyond retrieval, i.e., his adjustment is inadequate. The probability value which he produces from this underadjusted state of mind ($p_2$) will tend to lie between what he presently believes (his postdicted probability) and what he originally believed ($p_1$). That is to say, $p_2$ will tend to be higher than $p_1$ for events reported to have happened, lower for events reported not to have happened. If, for example, he judges the likelihood of events by his ability to build scenarios leading to their occurrence (Tversky & Kahneman, 1973), he may find scenarios of occurrence more available in reconstruction.

An alternative, although related, approach to retrieving $p_1$ is to take the postdictive probability of the outcome as an anchor and to adjust upward or downward from there, as seems appropriate. However valid the perceived reasons for adjustment, the combination of creeping determinism and underadjustment would lead to the effect studied here. One may, for example, find it difficult to imagine how he could ever have imagined that things could work out otherwise.

The differential effect with A–C and B events was an unexpected and interesting result meriting further attention. One possible explanation is that reports of nonoccurrence tend to have a smaller and/or more readily ignored (eliminated) impact on the judge's "state of mind" than reports of occurrence. If as Carr (1961, p. 26) claims, "History is by and large a record of what people did and not of what they failed to do," reports of nonoccurrence may tend not even to be noticed. Possibly, distinguishing between "events reported not to have happened" and "events not reported which have not happened," as we distinguished between A and C events, would sharpen the analysis. A supplementary explanation relevant to these particular stimulus materials arises from the fact that the Nixon trips were noted more for what did not happen than what did. Whatever their symbolic and long-range significance, there were fewer substantive results than many observers anticipated. Such nonoccurrences as observers did note may have included many acknowledged "surprises." If After judges remembered the surprisingness of these nonoccurrences and their memories erased the tendency for $p_2$ to be lower than $p_1$, then random fluctuations, along with the slight regression effect, would have produced many instances of $p_2$ higher than $p_1$. An additional situation-specific consideration is the fact that none of the outcomes could have happened had the trips been canceled, a real possibility at the time of $p_1$ estimation. In retrospect, however, the doubt which surrounded the trips may be unavailable and the likelihood of contingent outcomes enhanced.

The "surprisingness" of a set of events might be defined as the extent to which unlikely events are perceived to occur and likely events not to occur. For a judge evaluating his own performance, the surprisingness of a set of events is an indicator of his degree of understanding of those events. For the judge with perfect knowledge of a set of determinate events, there will be no "surprises," as he assigns 100% and 0% to A–C and B events, respectively. The more surprising a set of events is perceived to be, the greater the negative feedback and impetus to learn from experience which it presumably provides.[3]

In this light, the above results reflect a retrospective reduction in the surprisingness of the events judged, a reduction which also constitutes a tendency to convert negative feedback to positive. Although a causal link has not been established, it seems reasonable to speculate that once distorted in memory, knowledge of unexpected outcomes may actually encourage ineffective predicting instead of compelling the judge to improve his prediction-producing mechanisms. The judge who is insufficiently aware of the surprises the past held for him, and of the need to improve his performance, seems likely to continue being surprised by what happens in the future. Figure 2 offers the contrast of a relatively surprise-free past ($p_2$) with a relatively surprise-full future ($p_1$); although, of course, here judgments of the future temporally preceded those of the past. The "inertia effect" reported by Geller and Pitz (1968) is one case in which judges' conversion of negative feedback to positive is detrimental to learning.

Consider also a judge who has been caught unprepared by some turn of events. Looking backward, he "remembers" attributing a greater likelihood to the event before its occurrence than he actually did attribute to it. He may conclude that he, more or less, "knew that it was going to happen," but wasn't ready for it when it did, and that in the future he'll do better. If, for example, $p_1 = 30\%$ and $p_2 = 50\%$, he might decide that next time he'll be doubly ready for any 50% likely event, which would leave him unprepared for the occurrence of a similarly likely ($p_1 = 30\%$) event. Had he remembered his own prediction, a, possibly the, operative lesson to be learned from the incident is that the data at his disposal is quite indeterminate and that he should be ready for a substantial number of surprises. As Wohlstetter (1962, p. 401) noted in concluding her study of the surprise attack at Pearl Harbor, "We have to accept the fact of uncertainty and learn to live with it. No magic, in code or otherwise, will provide certainty. Our plans must work without it."

[3] An implicit assumption throughout this discussion is that, to be effective, learning from experience must be at least partially conscious. A case might be made that what is important for learning in the present context is that postdictive probabilities be in order, and not that predictive probabilities be remembered and the reasons for the prediction–postdiction difference recognized. Considering the evidence available, we believe both that this is usually not the case, and that postdictive probabilities are generally not in order, for reasons mentioned in the text and in Fischhoff (1974).

## REFERENCES

Berlin, I. *Historical inevitability.* London: Oxford University Press, 1954.

Carr, E. H. *What is history?* London: Macmillan, 1961.

Fischhoff, B., *Aspects of historical judgment.* Unpublished doctoral dissertation, 1974, Hebrew University of Jerusalem, Israel.

Geller, E. S., & Pitz, G. F. Confidence and decision speed in the revision of opinion. *Organizational behavior and human performance,* 1968, **3,** 190–201.

Slovic, P. From Shakespeare to Simon: Speculations, and some evidence, about man's ability to process information. *Oregon research institute research monograph,* 1972, **12,** No. 2.

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

Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive psychology,* 1973, **5,** 207–232.

Wohlstetter, Roberta. *Pearl Harbor: Warning and decision.* Stanford, CA: Stanford University Press, 1962.

Received: December 4, 1973