

AD-A033 181

DECISIONS AND DESIGNS INC MCLEAN VA

F/G 5/10

ON THE PSYCHOLOGY OF EXPERIMENTAL SURPRISES: OUTCOME KNOWLEDGE --ETC(U)

AUG 76 B FISCHHOFF, P SLOVIC

N00014-76-C-0074

UNCLASSIFIED

NL

| OF |
AD
A033181



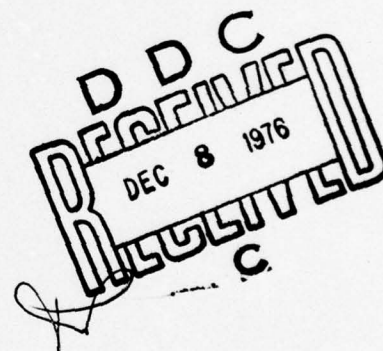
ADA033181

11

**ON THE PSYCHOLOGY
OF EXPERIMENTAL SURPRISES:
OUTCOME KNOWLEDGE AND THE JOURNAL
REVIEW PROCESS**

OREGON RESEARCH INSTITUTE

Baruch Fischhoff
Paul Slovic



ADVANCED
DECISION TECHNOLOGY
PROGRAM

CYBERNETICS TECHNOLOGY OFFICE
DEFENSE ADVANCED RESEARCH PROJECTS AGENCY
Office of Naval Research • Engineering Psychology Programs

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

The objective of the Advanced Decision Technology Program is to develop and transfer to users in the Department of Defense advanced management technologies for decision making.

These technologies are based upon research in the areas of decision analysis, the behavioral sciences and interactive computer graphics.

The program is sponsored by the Cybernetics Technology Office of the Defense

Advanced Research Projects Agency and technical progress is monitored by the Office of Naval Research — Engineering Psychology Programs. Participants in the program are:

Decisions and Designs, Incorporated
The Oregon Research Institute
Perceptronic, Incorporated
Stanford University
The University of Southern California

Inquiries and comments with regard to this report should be addressed to:

Dr. Martin A. Tolcott
Director, Engineering Psychology Programs
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

or

LT COL Roy M. Gulick, USMC
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

The views and conclusions contained in this document are those of the author(s) and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government. This document has been approved for public release with unlimited distribution.

TECHNICAL REPORT DDI-6

**ON THE PSYCHOLOGY OF EXPERIMENTAL SURPRISES:
OUTCOME KNOWLEDGE AND THE
JOURNAL REVIEW PROCESS**

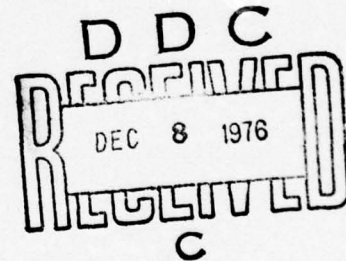
by

Baruch Fischhoff and Paul Slovic

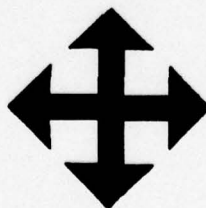
Sponsored by

Defense Advanced Research Projects Agency
Contract N00014-76-C-0074
Under Subcontract from
Decisions and Designs, Incorporated

August, 1976



ACCESSION for	
NTIS	White Section ☒
D C	Buff Section ☐
UNCLASSIFIED	☐
JUSTIFICATION	
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	Avail. and/or SPECIAL
A	



OREGON RESEARCH INSTITUTE
P.O. Box 3196
Eugene, Oregon 97403
(503) 484-2123

DISTRIBUTION STATEMENT A

Approved for public release;
Distribution Unlimited

SUMMARY

Overview

Three experiments show that in hindsight people systematically exaggerate the predictability of the results of scientific experiments. This judgmental bias has implications for the management of scientific research programs, the conduct of experimental research, and the review of scientific manuscripts.

Background

Much scientific research is conducted in two stages, the first of which is a pretest intended to see how viable the research design is and what sort of results it produces. After these initial results have been obtained, a decision must be made whether to continue with the full research project. Responsibility for the decision may lie either with the scientist conducting the study or a research manager. This decision can be made either before or after the pretest, with the corresponding questions being "If Result X is obtained, what should be done?" and "Given that Result X has been obtained, what should be done?" The answers to these two questions should be the same, as whether or not Result X is actually in hand should not affect its impact.

Findings

A series of three studies involving 463 people showed that this is typically not the case. Once the results of a pretest are in hand, they are viewed as much less surprising and much more likely to be replicated than they seemed in foresight. This finding was obtained with pretest results for a variety of different experiments. It appears that once a result has been reported, from even a sample of one, people feel that it more or less had to happen and that it is very likely to be obtained on reruns of the same experiment. In an effort to bring people's hindsightful perceptions of the meaning of pretest results more in line with their foresight perceptions, people told the result of the experiment were required to show how they would have explained the opposite pretest result, had it been obtained. This manipulation reduced the bias somewhat, but did not eliminate it.

Implications

If we underestimate the surprisingness of scientific results, we may also underestimate how much we have learned from them and overestimate how much we ourselves know without the benefit of such research. The practical implications of this judgmental bias depend on the judge's role in the research world. It could encourage a critic of research expenditures to say "What do we need these studies for? They are only telling us things we already know." It could lead a research manager to ask himself "Why did I decide to go ahead with that project when its results were so easily foretold?" It could lead a scientist to curtail a research program after receiving results from a pretest,

without realizing how statistically unstable such results are. It could lead the editor or reviewers of scientific journals to reject manuscripts because the results they report seem inevitable. Indeed, this paper begins with a selection of such reasons for rejection culled from the files of one of the most prestigious of psychological journals.

No proven way of overcoming this bias is known at present, and the article urges extreme caution in assessing the surprisingness of research results.

TABLE OF CONTENTS

	Page
SUMMARY	ii
TABLES	vi
ACKNOWLEDGMENTS	vii
INTRODUCTION	1
EXPERIMENT I	4
Method	4
Design	4
Stimuli	4
Virgin rat	5
Hurricane seeding	5
Gosling imprinting	5
Y-test	6
Instructions	6
Foresight	7
Hindsight	7
Subjects	8
Results	8
Discussion	10
EXPERIMENT II	11
Method	11
Results	11
Discussion of Experiments I and II	13
EXPERIMENT III	14
Method	14
Design	14
Stimuli	15
Procedure	16
Subjects	16
Results	16
GENERAL DISCUSSION	16

REFERENCES

20

DISTRIBUTION LIST

22

DD 1473

25

TABLES

	Page
Mean Probabilities for Experiments I and II	9
Mean Ratings for Experiment III	17

ACKNOWLEDGMENTS

Support for this research performed by Oregon Research Institute was provided by the Advanced Research Projects Agency of the Department of Defense and was monitored under Contract N00014-76-C-0074 with the Office of Naval Research under subcontract from Decisions and Designs, Inc.

We would like to thank Michael Posner for providing the anonymous quotations at the beginning of this article and Barbara Combs, Lewis Goldberg, and Sarah Lichtenstein for comments on an earlier draft.

ON THE PSYCHOLOGY OF EXPERIMENTAL SURPRISES:
OUTCOME KNOWLEDGE AND THE JOURNAL REVIEW PROCESS

INTRODUCTION

Consider the following excerpts from critiques of manuscripts submitted for editorial review.

"The present experiment does not tell us much that is new."

"None of the results appear terribly surprising. The author has used an elephant gun to kill a flea."

"I find the willingness of the author to obtain this result in yet another context slightly depressing."

"The reaction of the readers of (this journal) to this paper would be one of: 'of course'."

"What is clear is that (the authors) had in me a reader whose prior was on the order of .95 or more. By how much could they increase it?"

"I must apologize to you and the manuscript's author for the delay in responding to the manuscript. Part of my problem was in deciding why I could not recommend publication of a study with which I found no flaws. The paper is well written, the studies were well designed and conducted, and I do not feel that reading the paper was a waste of my time. Nevertheless, I could not escape the feeling that the paper merely shows to be false a hypothesis one can hardly take seriously to begin with."

In each of these examples, the reviewer recommended rejecting an article that was technically competent because the results appeared too predictable and unsurprising. Certainly this is a legitimate selection

criterion. Results that are wholly predictable, and thus fail to increase our scientific knowledge, hardly rate readers' time and valuable journal space.

Presumably, no one is better qualified to make this sort of judgment than the same reviewers who are selected for their ability to evaluate methodological competence, clarity of presentation, adequacy of literature review, and so on. Yet, judging the predictability of results requires a rather special kind of competence. The reviewer must ignore the results he or she has just seen reported and ascertain how likely they seemed before the experiment was performed. Some recent findings by Fischhoff (1975a; 1975b) and Fischhoff and Beyth (1975) suggest that such judgments may be problematic. Their results showed that reporting the outcome of a historical event increases the perceived likelihood of that outcome, and that people underestimate the effect of outcome knowledge on their perceptions. As a result, people believe that they would have seen in foresight the relative inevitability of the reported outcome which, in fact, was only apparent in hindsight. Thus, they exaggerate the predictability of reported outcomes.

It seems plausible that similar effects might occur when reviewing the results of scientific research. Once we hear experimental findings, we may tend to feel as though we "knew all along" that it would come out that way. If this is the case, then reviewers may systematically exaggerate the predictability of the findings they evaluate and as a result be unduly severe in their criticism.

The experiments reported below examine this possibility. Subjects in Experiments I and II read descriptions of a number of studies from

different scientific disciplines, each of which had two possible outcomes. Foresight subjects were told that a single subject was about to be tested in each experiment. For both possible outcomes, they were asked to indicate the probability that that outcome would be obtained on a specified number of additional replications, if it were obtained on the first subject. Hindsight subjects were told that the first subject had already been tested and had produced one of the possible outcomes. They were asked how likely it was that this outcome would be obtained on the same specified number of additional replications.

Thus, both groups were asked to assess the probability of a number of future replications, conditional on the outcome obtained from a first subject. Formally, these conditional probabilities should be the same for subjects in both groups. We hypothesized, however, that hindsight subjects, told the outcome obtained on the first subject, would exaggerate its inevitability and thus the probability that it would be replicated on future trials. Foresight subjects, we believed, would be less sanguine about the prospects of successful replications. One reason for such an effect is that foresight is a perspective conducive to seeing how an experiment could go either way, whereas in hindsight, we may be so intent on explaining the reported result that we can no longer see how the experiment could, in past or future, have gone otherwise.

For the sake of conceptual clarity, it may be valuable to consider the relationship of this experiment to Tversky and Kahneman's (1971, 1973; also Kahneman & Tversky, 1973) finding that people are insensitive to the amount and quality of the information on which their judgments are based. In particular, people are willing to make very confident predictions on the basis of very limited samples. These results

predict that both foresight and hindsight subjects will have undue confidence in the replicability of the outcome obtained from the first subject in each study. They do not predict different degrees of overconfidence for foresight and hindsight subjects.

EXPERIMENT I

Method

Design

Subjects received brief descriptions of experiments drawn from biology, psychology and meteorology which they were told either would soon be conducted (foresight) or had recently been conducted (hindsight). Foresight subjects were told that two outcomes were possible with the first subject in each experiment, while hindsight subjects were told that one of those two possible outcomes had been obtained. Foresight subjects were asked to: (a) assign a probability to each of the possible first-subject outcomes; (b) explain why each outcome might occur; and (c) estimate the probability that each of the two possible outcomes would be replicated in all, some, or none of a fixed number of replications if it were obtained with the initial subject. Hindsight subjects were asked to: (a) explain why the reported outcome had occurred; and (b) estimate the probability that it would be obtained in all, some, or none of the replications. The dependent variable for all groups was the conditional probability of replicating the outcome of the initial experiment.

Stimuli

The descriptions of the four experiments along with the possible

outcomes considered and the number of replications were as follows:¹

Virgin rat: Several researchers intend to perform the following experiment. They will inject blood from a mother rat into a virgin rat immediately after the mother rat has given birth. After the injection, the virgin rat will be placed into a cage with the newly born baby rats, after removal of the actual mother.

Outcomes used: (a) The virgin rat exhibited maternal behavior; (b) the virgin rat failed to exhibit maternal behavior. Subjects estimated the probability of the initial result being replicated with all, some, or none of 10 additional virgin rats.

Hurricane seeding: A team of government meteorologists recently seeded a tropical storm, which had reached hurricane status, with large quantities of silver-iodide crystals (the same type of crystals are used to seed clouds in attempts to produce rain).

Outcomes used: (a) The hurricane increased in intensity; (b) the hurricane decreased in intensity. Subjects estimated the probability of the initial result being replicated in all, some, or none of six additional hurricanes.

Gosling imprinting: A goose egg was placed in a sound-proof, heated box from time of laying to time of cracking. Approximately two days before it cracked, the experimenter began to intermittently play sounds of ducks quacking into the box. On the day after birth,

¹ For stylistic purposes, the tenses of the verbs used in these descriptions varied between experiments and between hindsight and foresight versions of the same experiment. Fischhoff (1976) has found that the tense used in describing events has no effect on their perceived likelihood.

the gosling was placed on a smooth floor equidistant from a duck and a goose, each of which was in a wire cage. The gosling was observed for two minutes.

Outcomes used: (a) The gosling approached the caged duck; (b) the gosling approached the caged goose. Subjects estimated the probability of the initial result being replicated with all, some, or none of 10 additional goslings.

Y-test: In the pretest of an experiment that she intends to run in the future, an experimenter placed a four-year-old child in front of an easel with a large Y on it, with a dot in the lower left-hand third. The child was then taken around to the back of the easel where he saw another Y. He was asked to draw a dot in the "same position" on that Y as the one he had just seen.

Outcomes used: (a) The child placed a dot in Area A [the lower left-hand third]; (b) the child placed a dot in Area B [the lower right-hand third]. Subjects estimated the probability that the initial result would be replicated with one additional child.

The hurricane seeding experiment was loosely based on Howard, Matheson and North (1972); the imprinting study on Grier, Counter and Shearer (1967); the Y-test on Smothergill, Hughes, Timmons and Hutko (1975). The virgin rat study was invented.

The virgin rat experiment was presented to one set of foresight and hindsight groups. The other three experiments were presented together to a second set of foresight and hindsight groups.

Instructions

All subjects received the same general instruction: "The following questionnaire concerns your scientific intuitions. We'd like to

ask you a number of questions about possible results of several experiments in different areas which have recently been conducted, or will be in the near future. We thank you for your cooperation." Each experiment appeared on a separate page, with the description at the top. Questions were presented in the following format (using the virgin rat example):

Foresight: 1.a. What is the probability that the virgin rat will exhibit maternal behavior? _____ Why do you think that this might happen?

b. What is the probability that the virgin rat will not exhibit maternal behavior? _____ Why do you think that this might happen?

2. If the virgin rat does exhibit maternal behavior, what is the probability that in a replication of this experiment with ten additional virgin female rats:

a. All will exhibit maternal behavior? _____

b. Some will exhibit maternal behavior? _____

c. None will exhibit maternal behavior? _____

(Note: These three probabilities should total 100%.)

3. Identical to Question 2 except that it begins "If the virgin rat does not exhibit maternal behavior . . ."

Hindsight (after being told either that the initial virgin rat exhibited maternal behavior or that it failed to exhibit maternal behavior):

1. Why do you think that this happened?

2. What is the probability that in a replication of this experiment with ten additional virgin female rats:

- a. All will exhibit maternal behavior? _____
- b. Some will exhibit maternal behavior? _____
- c. None will exhibit maternal behavior? _____

(Note: These three probabilities should total 100%.)

Subjects

All 184 subjects were paid volunteers who responded to an ad in the University of Oregon student newspaper. The present task was the first of several performed during a two-hour session. Group size varied from 24 to 37.

Results

The first and third columns of Table 1 present the mean probability of replication assigned by the foresight and hindsight groups in Experiment I. The italicized rows of Table 1 (rows 1, 6, 7, 10, 13, 16, 19, and 23) present the mean judged probability of the initial outcome's being obtained on all subsequent replications. In six of eight cases (two from each of four experiments), this probability was significantly larger for hindsight than foresight subjects. Thus, subjects told that an experiment had "worked" once in the past found its working consistently in the future more likely than those asked "if it works once, how likely is it to work again consistently?" For the three experiments with multiple replications (virgin rat, hurricane seeding, gosling imprinting), the mean probability of an initial outcome's always being replicated was .383 for the foresight group and .546 for the hindsight group; the mean probabilities of its never being replicated were .187 and .095, respectively.

Table 1

Mean Probabilities for Experiments I and II

Experiment	Initial Result	Result of Replication	Foresight		Hindsight	
			Two-Alternative		One-Alternative	
			Experiment I	Experiment II	Experiment I	Experiment II
1. Virgin Rat	a. Shows maternal behavior	1. All show maternal behavior	.295 ^a	.269 ^b	.436 ^c	.473 ^b
		2. Some show maternal behavior	.419	.497	.489	.397
		3. None show maternal behavior	.289***	.234	.073*	.130***
	b. Fails to show maternal behavior	4. All show maternal behavior	.095	.088	.087	.110
		5. Some show maternal behavior	.334	.324	.239	.314
		6. None show maternal behavior	.570	.589	.666	.575
2. Hurricane Seeding	a. Intensity increases	7. All increase	.291***	.307	.557***	.347.
		8. Some increase	.494*	.521	.367*	.511
		9. None increase	.215***	.171	.077**	.142
	b. Intensity decreases	10. All weaken	.337*	.315	.470	.502**
		11. Some weaken	.472	.491	.399	.375
		12. None weaken	.191	.194	.131	.123*
3. Gosling Imprinting	a. Approaches goose	13. All approach goose	.403*	.334	.728*	.588*
		14. Some approach goose	.409	.512	.218*	.340
		15. None approach goose	.188**	.153	.054	.071**
	b. Approaches duck	16. All approach duck	.401	.399	.417	.331
		17. Some approach duck	.456	.436	.432	.554
		18. None approach duck	.146	.165	.150	.115
4. Y-test	a. Places dot in region A	19. Places in region A	.292*	.298	.579***	.382*
		20. Places in region B	.134	.115	.119	.180
		21. Places in region C	.574**	.585	.304**	.436**
	b. Places dot in region B	22. Places in region A	.260	.286	.245*	.292
		23. Places in region B	.160***	.149*	.363	.305***
		24. Places in region C	.579***	.463*	.393	.388***

* $p < .05$ ** $p < .01$ *** $p < .001$ significance of differences by one-tailed t-tests

a = significance tests refer to differences from hindsight-one alternative (column 3)

b = significance tests refer to differences from foresight-two alternatives (column 1)

c = significance tests refer to differences from hindsight-two alternatives (column 4)

Note: Sample size varies from 24 to 41.

Discussion

Why should these formally equivalent conditional probabilities be judged differently by hindsight and foresight subjects? Two possibilities occur to us. One, suggested in the introduction, is that hindsight subjects unduly concentrate their attention on the reported outcome, thereby failing to see how the initial experiment, could have gone the other way. A second possibility is that the conditional judgments that the foresight subjects make ("if it were to work once, what is . . . ?") are quite difficult and confusing. Thus, when they attempt to consider the possible occurrence of two different outcomes, they may be unable to devote to either the attention given by hindsight to their one alternative. As a result, foresight subjects may be unable to properly assess the impact which the result from the first subject should have on the perceptions. In summary, the "availability of reasons" explanation attributes the discrepancy to hindsight subjects' failure to consider the feasibility of alternative outcomes. The "conditionality" explanation attributes the effect to the inability of foresight subjects to consider multiple contingencies. Both may be true.

Experiment II tests these hypotheses by replicating Experiment I with the following differences: (a) Foresight subjects were required to consider the probability of replicating only one of the possible outcomes; (b) hindsight subjects were required to explain not only why the reported outcome happened, but also "Had the experiment worked out the other way, how would you explain it?" These one-alternative foresight subjects should be able to devote the same undivided attention to their one possible outcome that the one-alternative hindsight

subjects in Experiment I could devote to their one reported outcome. If the conditionality hypothesis is correct, they should respond more like one-alternative hindsight subjects than the two-alternative foresight subjects in Experiment I. According to the "availability-of-reasons" hypothesis, hindsight subjects forced to consider why the unreported outcome might have occurred should respond like foresight subjects.

EXPERIMENT II

Method

Experiment II was identical to Experiment I except for two changes. The first was that foresight subjects estimated the probability of replicating only one of the two possible outcomes for each experiment. They were asked either "If the experiment works, how likely is that result to be replicated?" or "If the experiment doesn't work, how likely is that result to be replicated?" The two-alternative foresight group in Experiment I answered both these questions. Second, the two-alternative hindsight group of Experiment II was asked not only "Why did the experiment work out this way?" but also "Had the experiment worked out the other way, how would you explain it?" Like the one-alternative hindsight subjects of Experiment I, they estimated the probability of replication only for the reported outcome. One hundred and fifty-one subjects were recruited in the same manner as in Experiment I.

Results

Columns 2 and 4 of Table 1 present the mean probabilities from Experiment II. Comparing columns 1 and 2, we see that the responses

of one- and two-alternative foresight subjects were generally indistinguishable. Reducing the number of alternatives considered did not systematically increase the perceived probability of replicating the initial outcome. In only 6 of 24 cases was the one-alternative foresight mean closer to the one-alternative hindsight mean than to the two-alternative foresight mean. Thus, there is no evidence that attentional problems are responsible for the hindsight-foresight discrepancy.

The second manipulation, forcing two-alternative hindsight subjects to consider how the first trial of the experiment could have turned out otherwise, produced a marked difference. A comparison of column 4 with columns 1 or 2 reveals a substantial hindsight effect for 5 of the 8 outcomes considered (all but 1b, 2a and 3b). The size of the effect, however, was reduced. For four of the eight outcomes (1a, 2a, 3a, 4a), the mean probability of consistently replicating the reported outcome was significantly lower for two-alternative than for one-alternative hindsight subjects. In general, the means of the two-alternative hindsight lie between those for the one-alternative hindsight and both foresight groups. Although not inconsistent with the conditionality hypothesis, these results strongly support the "availability-of-reasons" hypothesis.

Further evidence of the effect of reason availability on probability judgments was sought by looking at those two-alternative hindsight subjects unable to supply reasons for one of the two alternative outcomes. Subjects who could not think of one reason why the unreported outcome might have happened found replication of the reported outcome slightly more likely than did other subjects (mean difference

= .021); subjects who could not think of one reason for the reported outcome found replication much less likely than did other subjects (mean difference = .211).

Discussion of Experiments I and II

Although these results seem to support the "availability-of-reasons" account of the hindsight-foresight discrepancies, the evidence is inconclusive. The possibility remains that conditional tasks, however structured, cause difficulties. There appear to be few, if any, empirical studies germane to this problem. Aside from its theoretical interest, the question of conditional judgments has significant applied implications. If we are to engage effectively in contingency planning, we must be able to assess, in advance, the impact which receipt of various possible data may have on our perceptions. If these conditional judgments are inaccurate, the plans based on them may appear grossly inappropriate when dimly foreseen contingencies do arise (Brown, 1976).

If the hindsight effect found in Experiments I and II afflicts researchers, it may constitute an important impediment to scientific progress. When planning an experiment, investigators, like our foresight subjects, may be able to see that various results are possible and that they should not put undue confidence in results from a few initial subjects. However, once a (any?) result has been obtained on pilot trials, they may throw caution to the wind and view that result as highly likely and easily replicable. As a result they may reduce the size and power of the ensuing sample--a step whose consequences have been noted by Cohen (1969) and by Tversky and Kahneman (1971).

As reviewers, this bias may lead us to denigrate worthy contributions, believing, like those reviewers cited at the beginning of this paper, that they are trivial, obvious and foreseeable. Before speculating further on the implication of this hindsight bias on the journal review process, let us consider some evidence acquired in a setting more closely resembling that process. Experiment III replicates Experiments I and II in a journal review format.

EXPERIMENT III

Method

Design

Subjects were asked to read and evaluate scientific manuscripts in a manner similar to that of professional reviewers. Hindsight reviewers received manuscripts with introduction, method, and results sections. For foresight subjects, the results section was missing. Each manuscript was composed so that there were two possible outcomes for the study in question. There were two separate hindsight groups, each receiving one of the possible outcomes presented as if it had actually happened.

Subjects were asked to evaluate the manuscripts on seven 7-point scales, two of which were designed to be sensitive to hindsight-foresight differences. One was surprisingness of results: hindsight subjects assessed the surprisingness of the reported outcome; foresight subjects assessed how surprising each of the two possible outcomes would seem were they obtained. The second sensitive question was stability of results: hindsight subjects assessed the likelihood that the reported results would be obtained in an exact replication of the same

experiment; foresight subjects answered the same question for each of the possible results. The remaining five scales were used as fillers and to test for other possible changes between hindsight and foresight. They referred to: clarity of the introduction, clarity of the research question, clarity of the method, adequacy of the method to test the research question, and personal interest in the study.

Stimuli

Three experiments from diverse areas of psychology were used.² One, called "Scientific ambiguity and attitudinal conflict," described an unpublished experiment which we had recently completed. In that experiment, subjects first indicated their position on several environmental issues, including nuclear power; some time later, they were asked to guess whether an ambiguous statement about nuclear power was offered by an opponent or proponent of nuclear power. We had hypothesized that people would interpret ambiguous statements as supporting their own positions--but we were wrong.

The second and third studies were "elaborated" versions of the gosling imprinting and Y-test studies used in Experiments I and II. No hypothesis was advanced for either of these studies.

These studies were chosen to be unfamiliar, yet comprehensible, without prior knowledge of the area. They were written to show that there were two possible outcomes, each of which could conceivably be obtained.

² Copies of these descriptions and the accompanying questionnaires are available upon request.

Procedure

Subjects were told about the review process for scientific manuscripts and then asked to perform a task similar to that of actual reviewers. They read the three studies in the order given above, evaluating each before going on to the next.

Subjects

One hundred twenty-eight paid subjects participated, responding to an advertisement in the University of Oregon student newspaper. They were assigned to the foresight group or one of the two hindsight groups according to their preference for experimental date and hour.

Results

If these reviewers are susceptible to a hindsight bias, the hindsight subjects should find the reported results less surprising and more likely to be replicated than the foresight subjects anticipated they would appear. Table 2 presents the relevant group means for the two outcomes used for each of the three experiments. In five of the six cases, hindsight subjects found the reported outcome less surprising and more replicable than did foresight subjects; in three of six cases, this difference was statistically significant. There were no systematic differences on the five filler questions.

GENERAL DISCUSSION

Experiment III did not allay our concern. As reviewers, hindsight bias may lead us to denigrate worthy contributions, believing, like those reviewers cited at the beginning of this paper, that they are trivial, obvious, and foreseeable. An extreme measure to counter this tendency would be to institute a system of deaf review, forcing

Table 2

Mean Ratings for Experiment III

<u>Experiment</u>	<u>Outcome</u>	<u>Surprisingness of Results (1 = very surprising)</u>		<u>Stability of Results (1 = very likely to replicate)</u>	
		<u>Foresight</u>	<u>Hindsight</u>	<u>Foresight</u>	<u>Hindsight</u>
Ambiguity	Confirm	5.57	5.42	3.09	2.76
	Hypothesis				
	Disconfirm	3.02	4.67***	4.41	2.91***
	Hypothesis				
Gosling Imprinting	Follow	4.05	4.28	2.77	2.81
	Goose				
	Follow	3.66	4.62*	3.11	2.93
	Chicken				
Y-Test	Place dot	5.34	6.04*	2.75	1.73**
	on A				
	Place dot	3.23	3.69	3.52	2.80
	on C				

one-tailed tests:

* $p < .05$ ** $p < .01$ *** $p < .001$

the reviewer to make some evaluations without hearing the results. Under this system, the reviewer receives the introduction and method sections of an article without the results. After reading these sections, he or she makes a written prediction of the outcome of the experiment. Once this prediction has been returned to the editor, the remainder of the manuscript is sent. The written record can alert both editor and reviewer to what results could be anticipated. For such a scheme to be successful, authors would have to give no hint of their results in the introduction and method sections.

Given the demands currently placed on journal editors, reviewers, and authors, this cumbersome proposal has little chance of being accepted. For many, possibly most, manuscripts, such a procedure would also be unnecessary. For these manuscripts are rejected on technical grounds independent of their informativeness. Only if a submission is methodologically competent, tolerably written, and sent to the proper journal, does its fate depend on its scientific substance. Thus, hindsight bias, when present, will primarily affect a subset of the best submitted manuscripts.

How do we protect these manuscripts from unfair rejection? The reduced hindsight effect with the two-alternative hindsight group in Experiment II suggests one solution: Have reviewers provide reasons pointing to the result that was not obtained. Such a proposal seems implicit in the editorial policy recently set forth by one APA journal:

The author and reader of a research report should both feel it possible to make a convincing case that the results of the reported research could have been interesting if they had come out differently from those reported. Publication

in JPSP is inappropriate when it is not possible to imagine any reasonable basis for finding results other than the ones reported. (Greenwalt, 1976, p. 4)

There is, however, no guarantee that this simple procedure will do the job. In Experiment II the debiasing was only partial, and substantial hindsight effects were still obtained with five of the eight outcomes. More research is clearly needed. Until the extent of this bias is known and techniques for eliminating it are developed, we might do well to reject or at least reduce the importance of informativeness-surprisingness as a criterion in the manuscript review process.

REFERENCES

- Brown, R. V. Modeling subsequent acts for decision analysis. McLean, Va.: Decisions and Designs, Inc., 1975.
- Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1969.
- Fischhoff, B. Hindsight: Thinking backward? Psychology Today, 1975, 3, 4.
- Fischhoff, B. Hindsight $\neq$ foresight: The effect of outcome knowledge on judgment under uncertainty. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 288-299.
- Fischhoff, B. The effect of temporal setting on likelihood estimates. Organizational Behavioral and Human Performance, 1976, 15, 180-194.
- Fischhoff, B., & Beyth, R. "I knew it would happen"--remembered probabilities of once-future things. Organizational Behavior and Human Performance, 1975, 13, 1-16.
- Greenwald, A. G. An editorial. Journal of Personality and Social Psychology, 1976, 33, 1-7.
- Grier, J.B., Counter, S. A., & Shearer, W. M. Pre-natal auditory imprinting in chickens. Science, 1967, 155, 1692-1693.
- Howard, R. A., Matheson, J. E., & North, D. W. The decision to seed hurricanes. Science, 1972, 176, 1191-1202.
- Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.
- Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 237-251.
- Smothergill, D. W., Hughes, F. P., Timmons, S. A., & Hutko, P. Spatial visualizing in children. Developmental Psychology, 1975, 11, 4-13.

Tversky, A., & Kahneman, D. The belief in the "law of small numbers."

Psychological Bulletin, 1971, 76, 105-110.

Research Distribution List

Department of Defense

Assistant Director (Environment and Life Sciences)

Office of the Deputy Director of Defense
Research and Engineering (Research and
Advanced Technology)

Attention: Lt. Col. Henry L. Taylor
The Pentagon, Room 3D129
Washington, DC 20301

Office of the Assistant Secretary of Defense (Intelligence)

Attention: CDR Richard Schlaff
The Pentagon, Room 3E279
Washington, DC 20301

Director, Defense Advanced Research Projects Agency

1400 Wilson Boulevard
Arlington, VA 22209

Director, Cybernetics Technology Office

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Director, Program Management Office

Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209
(two copies)

Administrator, Defense Documentation Center

Attention: DDC-TC
Cameron Station
Alexandria, VA 22314
(12 copies)

Department of the Navy

Office of the Chief of Naval Operations (OP-987)

Attention: Dr. Robert G. Smith
Washington, DC 20350

Director, Engineering Psychology Programs (Code 455)

Office of Naval Research
800 North Quincy Street
Arlington, VA 22217
(three copies)

Assistant Chief for Technology (Code 200)

Office of Naval Research
800 N. Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 230)

800 North Quincy Street
Arlington, VA 22217

Office of Naval Research

Naval Analysis Programs (Code 431)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research

Operations Research Programs (Code 434)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (Code 436)

Attention: Dr. Bruce McDonald
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research

Information Systems Program (Code 437)
800 North Quincy Street
Arlington, VA 22217

Office of Naval Research (ONR)

International Programs (Code 1021P)
800 North Quincy Street
Arlington, VA 22217

Director, ONR Branch Office

Attention: Dr. Charles Davis
536 South Clark Street
Chicago, IL 60605

Director, ONR Branch Office

Attention: Dr. J. Lester
495 Summer Street
Boston, MA 02210

Director, ONR Branch Office

Attention: Dr. E. Glove and Mr. R. Lawson
1030 East Green Street
Pasadena, CA 91106
(two copies)

Dr. M. Bertin

Office of Naval Research
Scientific Liaison Group
American Embassy — Room A-407
APO San Francisco 96503

Director, Naval Research Laboratory

Technical Information Division (Code 2627)
Washington, DC 20375
(six copies)

Director, Naval Research Laboratory (Code 2029)

Washington, DC 20375
(six copies)

Scientific Advisor
Office of the Deputy Chief of Staff
for Research, Development and Studies
Headquarters, U.S. Marine Corps
Arlington Annex, Columbia Pike
Arlington, VA 20380

Headquarters, Naval Material Command
(Code 0331)
Attention: Dr. Heber G. Moore
Washington, DC 20360

Headquarters, Naval Material Command
(Code 0344)
Attention: Mr. Arnold Rubinstein
Washington, DC 20360

Naval Medical Research and Development
Command (Code 44)
Naval Medical Center
Attention: CDR Paul Nelson
Bethesda, MD 20014

Head, Human Factors Division
Naval Electronics Laboratory Center
Attention: Mr. Richard Coburn
San Diego, CA 92152

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Naval Personnel Research and Development
Center
Management Support Department (Code 210)
San Diego, CA 92152

Naval Personnel Research and Development
Center (Code 305)
Attention: Dr. Charles Gettys
San Diego, CA 92152

Dr. Fred Muckler
Manned Systems Design, Code 311
Navy Personnel Research and Development
Center
San Diego, CA 92152

Human Factors Department (Code N215)
Naval Training Equipment Center
Orlando, FL 32813

Training Analysis and Evaluation Group
Naval Training Equipment Center
(Code N-00T)
Attention: Dr. Alfred F. Smode
Orlando, FL 32813

Department of the Army

Technical Director, U.S. Army Institute for the
Behavioral and Social Sciences
Attention: Dr. J.E. Uhlaner
1300 Wilson Boulevard
Arlington, VA 22209

Director, Individual Training and Performance
Research Laboratory
U.S. Army Institute for the Behavioral and
and Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Director, Organization and Systems Research
Laboratory
U.S. Army Institute for the Behavioral and
Social Sciences
1300 Wilson Boulevard
Arlington, VA 22209

Department of the Air Force

Air Force Office of Scientific Research
Life Sciences Directorate
Building 410, Bolling AFB
Washington, DC 20332

Robert G. Gough, Major, USAF
Associate Professor
Department of Economics, Geography and
Management
USAF Academy, CO 80840

Chief, Systems Effectiveness Branch
Human Engineering Division
Attention: Dr. Donald A. Topmiller
Wright-Patterson AFB, OH 45433

Aerospace Medical Division (Code RDH)
Attention: Lt. Col. John Courtright
Brooks AFB, TX 78235

Other Institutions

The Johns Hopkins University
Department of Psychology
Attention: Dr. Alphonse Chapanis
Charles and 34th Streets
Baltimore, MD 21218

Institute for Defense Analyses
Attention: Dr. Jesse Orlansky
400 Army Navy Drive
Arlington, VA 22202

Director, Social Science Research Institute
University of Southern California
Attention: Dr. Ward Edwards
Los Angeles, CA 90007

Perceptronics, Incorporated
Attention: Dr. Amos Freedy
6271 Variel Avenue
Woodland Hills, CA 91364

Director, Human Factors Wing
Defense and Civil Institute of
Environmental Medicine
P.O. Box 2000
Downsville, Toronto
Ontario, Canada

Stanford University
Attention: Dr. R.A. Howard
Stanford, CA 94305

Montgomery College
Department of Psychology
Attention: Dr. Victor Fields
Rockville, MD 20850

General Research Corporation
Attention: Mr. George Pugh
7655 Old Springhouse Road
McLean, VA 22101

Oceanautics, Incorporated
Attention: Dr. W.S. Vaughan
3308 Dodge Park Road
Landover, MD 20785

Director, Applied Psychology Unit
Medical Research Council
Attention: Dr. A.D. Baddeley
15 Chaucer Road
Cambridge, CB 2EF
England

Department of Psychology
Catholic University
Attention: Dr. Bruce M. Ross
Washington, DC 20017

Stanford Research Institute
Decision Analysis Group
Attention: Dr. Allan C. Miller III
Menlo Park, CA 94025

Human Factors Research, Incorporated
Santa Barbara Research Park
Attention: Dr. Robert R. Mackie
6780 Cortona Drive
Goleta, CA 93017

University of Washington
Department of Psychology
Attention: Dr. Lee Roy Beach
Seattle, WA 98195

Eclectech Associates, Incorporated
Post Office Box 179
Attention: Mr. Alan J. Pesch
North Stonington, CT 06359

Hebrew University
Department of Psychology
Attention: Dr. Amos Tversky
Jerusalem, Israel

Dr. T. Owen Jacobs
Post Office Box 3122
Ft. Leavenworth, KS 66027

Tel-Aviv University
Interdisciplinary Center for
Technological Analysis and
Forecasting
Attention: Dr. Nava Pliskin
Ramat-Aviv, Israel

**Director, International Institute
for Applied Systems Analysis**
Schloss Laxenburg, Austria

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle)		5. TYPE OF REPORT & PERIOD COVERED
On the Psychology of Experimental Surprises: Outcome Knowledge and the Journal Review Process.		Technical Rept.
6. AUTHOR(s)		7. PERFORMING ORG. REPORT NUMBER
Baruch Fischhoff Paul Slovic		DDI-6
8. CONTRACT OR GRANT NUMBER(s)		9. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
Prime Contract No.: N00014-76-C-0074		
Subcontract No.: 75-036-0712		
10. PERFORMING ORGANIZATION NAME AND ADDRESS		11. REPORT DATE
Oregon Research Institute P. O. Box 3196 Eugene, Oregon 97403		August, 1976
12. CONTROLLING OFFICE NAME AND ADDRESS		13. NUMBER OF PAGES
Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Va. 22209		32
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report)
Office of Naval Research 800 North Quincy Street Arlington, VA 22217		Unclassified
15a. DECLASSIFICATION/DOWNGRADING SCHEDULE		
16. DISTRIBUTION STATEMENT (of this Report)		
Approved for public release; distribution unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
Support for this research performed by Oregon Research Institute was provided by the Advanced Research Projects Agency of the Department of Defense and was monitored under Contract N00014-76-C-0074 with the Office of Naval Research under subcontract from Decisions and Designs, Inc.		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
Foresight Hindsight Bias Probability of replication "Availability of reasons" hypothesis Conditionality hypothesis (This report examines whether		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)		
To merit publication a study must be informative; the reader must not feel that the results could have been confidently predicted prior to conducting the study. Earlier studies of the psychology of hindsight have shown that people overestimate the predictability of past events. Do they also overestimate the predictability of experimental results? Three experiments strongly suggest that reviewers may overestimate the degree to which they could have predicted the reported results of experiments. The implications of this bias for the journal review process, and some ways in which it might be alleviated, are discussed.		

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE

25

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

390664