

LEVEL II

(12)

TECHNICAL REPORT PTR-1091-80-5

FOR THOSE CONDEMNED TO STUDY THE PAST: REFLECTIONS ON HISTORICAL JUDGEMENT

DECISION RESEARCH

Baruch Fischhoff

May 1980

AD A088699

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution is unlimited.

DTIC
SEP 2 1980
A



**Command & Control
Decision & Forecasting
Systems Program**

CYBERNETICS TECHNOLOGY DIVISION
RESEARCH PROJECTS AGENCY

DDC FILE COPY

80 8 20 106

9 14
TECHNICAL REPORT PTR-1091-80-5

6
FOR THOSE CONDEMNED TO STUDY THE PAST:
REFLECTIONS ON HISTORICAL JUDGEMENT.

by
10 Baruch Fischhoff

Sponsored by
15
Defense Advanced Research Projects Agency
Under Contract MDA903-80-C-0194
✓ DARPA Order No. 3831

Under Subcontract from
Decisions and Designs, Inc.

11 May 1980 12 48

THE VIEWS AND CONCLUSIONS CONTAINED IN THIS DOCUMENT ARE THOSE OF THE AUTHOR AND SHOULD NOT BE INTERPRETED AS NECESSARILY REPRESENTING THE OFFICIAL POLICIES, EITHER EXPRESSED OR IMPLIED, OF THE DEFENSE ADVANCED RESEARCH PROJECTS AGENCY OR THE UNITED STATES GOVERNMENT.

DECISION RESEARCH
A Branch of Perceptronics
1201 Oak Street
Eugene, Oregon 97401
(503) 485-2400

DTIC
ELECTE
SEP 2 1980
A
D
DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

411215

LB

TABLES OF CONTENTS

	<u>Page</u>
SUMMARY	i
FIGURES	ii
ACKNOWLEDGMENT	iii
LOOKING FOR WISDOM	2
Informal Modeling	2
Formal Modeling	8
LOOKING FOR FOLLY	13
Focus on Failure	13
What Was the Problem?	14
Hindsight: Thinking Backwards?	18
LOOKING AT ALL	22
Why Look?	22
Have We Seen Enough?	28
CONCLUSION	30
Presentism	31
Methodism	31
Learning	32
Indeterminacy	33
FOOTNOTES	34
REFERENCES	35
DISTRIBUTION LIST	
DD FORM 1473	

Accession For	
NTIS General	☒
DDC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/ _____	
Availability Codes _____	
Dist.	Avail and/or Special
A	

SUMMARY

When anticipating the future and making decisions in the present, we are all prisoners of the past. Our personal or collective past tells us what factors are important to understand, how good our understanding is, and how many surprises to expect when making our plans. This dependence on the past is in large part justified; where else could one turn for wisdom and accumulated experience?

In trying to learn these lessons, our main, often only, tool is our own intellect. There has, however, been surprisingly little systematic study of the cognitive (or thought) processes involved in historical judgment, nor how people might be instructed to approach the past more efficaciously.

The present report provides a framework for studying historical judgment and describes the conclusions that may be drawn from psychological research and the historiographic literature, the musings of historians about their own craft. The cumulative picture suggests that the past does not yield its secrets readily. Some identifiable and perhaps correctable problems are: overinterpreting available evidence, unfairly second guessing historical actors, and exaggerating the predictability of future events for which analogs can be identified in the past. These judgmental biases can be found in lay as well as professional students of the past.

FIGURES

<u>Figure</u>		<u>Page</u>
Figure 1	Source: Jiler, 1972	4
Figure 2	Archetypal Patterns for the Co-occurrence of Historical Events	6

ACKNOWLEDGMENT

This research was supported by the Advanced Research
Projects Agency of the Department of Defense under contract
~~NDA-903-80-C-0194~~ ³⁸³¹
~~NO0014-79-312-0750~~ (ARPA Order No. ~~0194~~) under subcontract
from Decisions and Designs, Inc. to Perceptronics, Inc.

FOR THOSE CONDEMNED TO STUDY THE PAST

Benson (1972) has identified four reasons for studying the past: to entertain, to create a group (or national) identity, to reveal the extent of human possibility, and to develop systematic knowledge about our world, knowledge that may eventually improve our ability to predict and control. On a conscious level, at least, we behavioral scientists restrict ourselves to the last motive. In its pursuit, we do case studies, program evaluations and literature reviews. We even conduct experiments, creating artificial histories upon which we can perform our post mortems.

Three basic questions seem to arise in our retrospections: (a) Are there patterns upon which we can capitalize so as to make ourselves wiser in the future? (b) Are there instances of folly in which we can identify mistakes to avoid? (c) Are we really condemned to repeat the past if we don't study it? That is, do we really learn anything by looking backward?

Whatever the question we are asking, it is generally assumed that the past will readily reveal the answers it holds. Of hindsight and foresight, the latter appears as the troublesome perspective. One can explain and understand any old event if an appropriate effort is applied. Prediction, however, is acknowledged to be rather more tricky. The present essay investigates this presumption by taking a closer look at some archetypal attempts to tap the past. Perhaps its most general conclusion is that we should hold the past in a little more respect when we attempt to plumb its secrets. While the past entertains, ennobles, and expands quite readily, it enlightens only with delicate coaxing.

LOOKING FOR WISDOM

Informal Modeling

While the past never repeats itself in detail, it is often viewed as having repetitive elements. People make the same kinds of decisions, face the same kinds of challenges, and suffer the same kinds of misfortune often enough for behavioral scientists to believe that they can detect recurrent patterns. Such faith prompts psychometricians to study the diagnostic secrets of ace clinicians, clinicians to look for correlates of aberrant behavior, brokers to hunt for harbingers of price increases, and dictators to ponder revolutionary situations. Their search usually has a logic paralleling that of multiple regression or correlation. A set of relevant cases is collected and each member is characterized on a variety of dimensions. The resulting matrix is scoured for significant relationships that might aid us in predicting the future.

Usually, this process is conducted rather informally. One expression of informality is to avoid performing any calculations at all. Indeed, explicit calculation of relationships is very rare, not only among lay people, but even among those observers of passing events who offer general laws of history in their punditry. An obvious casualty of such informality is precision. If I went so far as to lay out all the data of interest before me, but failed to compute explicitly the correlation between number of siblings and GPA, I might be loath to describe the relationship with a stronger adjective than "high" or "negligible." If forced to give a number, I would probably give a rounded one like .5 or .2.

Adding another significant figure would represent misplaced precision, as I could not estimate the correlation so finely.

Imprecision is only a problem, however, if it produces errors large enough to threaten the validity of our conclusions. Often, that is not the case. For example, would a theory of social determinates of undergraduate success really look that different if the GPA-sibling correlation were .37 instead of .44? Probably not, and von Winterfeldt and Edwards (1973) have shown that moderate errors in estimating facts or values do little to change the expected value of decisions based upon them. Although we are proud of our ability to calculate, it may not be the chief benefit of our formal training and procedures.

More serious consequences of informality arise from the slippages in thinking it allows. Several recent case studies have shown that when their verbally stated assumptions are formalized, some of our more popular and accepted theories can be shown to contain internal contradictions (Coleman, 1960; Harris, 1976).

A more localized form of contradiction often buried in informal explanations is illustrated in Figure 1. Technical analysts spend their time exploring charts depicting the price movements of stocks, in the hopes of identifying precursors of past shifts in price, signs they hope to use in predicting future movements. Two of the many signs that analysts have identified are the formation of resistance to and support for future price increases. Yet a closer look shows that prior to the dramatic shifts at their respective ends, these two patterns were essentially identical. Thus, an undulating pattern neither predicts nor explains anything (given the present data), except in a tautological sense.

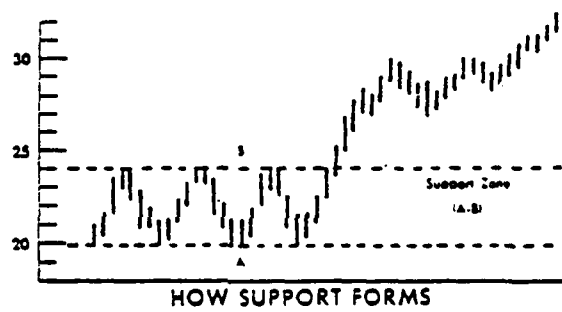
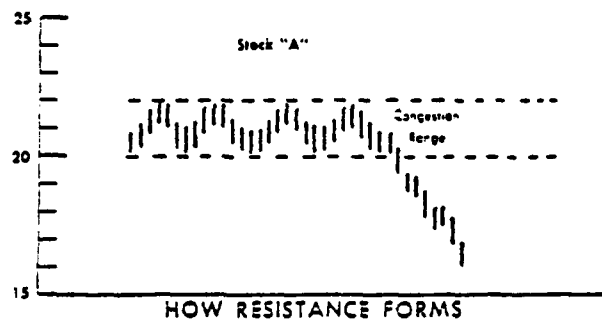


Figure 1
Source: Jiler, 1972

Closer to home, one can see the facility with which we are able to invoke contradictory "laws" of behavior to explain, predict or justify contrasting acts emerging from similar circumstances. "Haste makes waste" and "He who hesitates is lost" are such inconsistent explanations and admonitions. They make great sense when used alone and leave us looking foolish when presented together. When confronted with such an apparent contradiction, the natural defense is that "it all depends upon" Recognizing the need for such condition statements distinguishes science from undisciplined common sense. Progress might be measured by our ability to fill in the blank, wisdom by the frequency with which we remember those qualifications.

One barrier to discovering inconsistency is failing to realize the importance or relevance of inconsistent information. In their simplest form, laws or patterns of behavior or history can be presented by 2 x 2 tables of the type depicted in Figure 2. The rows might represent the occurrence or non-occurrence of one sort of event or personality characteristic (or its occurrence in large or small amounts); the columns represent something else. A predominance of entries in either diagonal represents a strong statistical relationship between the two variables; the absence of any off-diagonal entries indicates a logical relationship (i.e., E_1 iff E_2); entries in adjoining cells represent contradictory evidence (e.g., someone who is fat and happy contrasted with someone fat and sad or a happy-skinny and happy-fat pair).

Statisticians argue about the proper interpretation for various patterns of entries, as reflected in the great variety of correlation coefficients available. Lay people, on the

Example A

	E_1	$\bar{E}_1$
E_2	36	8
$\bar{E}_2$	5	23

Strong relationship

Example B

	E_1	$\bar{E}_1$
E_2	a	o
$\bar{E}_2$	o	a

Logical relationship

Example C

	Happy	Sad
Fat	a	b
Skinny	c	

Inconsistent pairs: a-b, b-c

Example D

	E_2
E_1	x

Psychologically relevant evidence

Figure 2
Archetypal Patterns for the Co-occurrence
of Historical Events

other hand, seem preoccupied with the upper-left-hand corner, representing co-occurrence of the two events or traits (or whatever) in question. When testing logical relationships, their predilection is to ask questions whose answers cannot falsify the hypothesis (Wason & Johnson-Laird, 1972). When assessing relationships in observed but untallied data, their attention is drawn to those cases in which rain followed cloud-seeding or some psychic's prediction preceded a major event or diagnosed paranoids perceived menacing eyes in inkblots. While the existence of many such upper-left-hand cases may give the impression of a recurrent pattern, in principle it tells little about the nature of the relationship between two phenomena or variables or traits (Ward & Jenkins, 1965).

Since one could always define the variables so as to put any co-occurrence in the favored cell, what determines which cell is attended to? One natural determinant is linguistic. Even though, as Sherlock Holmes demonstrated, one can sometimes learn much from the failure of a dog to bark, "history is by and large, a record of what people did, not of what they failed to do" (Carr, 1961, p. 126). A second determinant is our own expectations; we see and seek out (Shaklee & Fischhoff, 1979) what we expect to see and tend to miss or discount or avoid co-occurrences falling in the other cells. Chapman and Chapman (1969) have shown that people may see in data anticipated relationships that are not even there.

The representativeness of the sample of events upon which we base our conclusions is further compromised by the foibles of our own memories. In addition to focusing on expected occurrences, our recall processes may be biased in

other ways, say, toward recent events or those with lurid details.

Thus, there is everything to be said for being as explicit as possible in one's analysis of past events. A biased glance backward may be worse than none at all.

Formal Modeling

Scientific training is designed to help us avoid such mistakes. We use consistent schemes for characterizing cases and computational routines that include all relevant data. Rather than be satisfied with the gist of what was happening, we often develop specific formulae to account for past behavior.

The Daily Racing Form, for example, offers the earnest handicapper some one hundred pieces of information on each horse in any given race. The handicapper with a flair for data processing might commit to some computer's memory the contents of a bound volume of the Form and try to derive a formula predicting speed as a weighted sum of scores on various dimensions. For example:

$$\hat{y} = b_1x_1 + b_2x_2 + b_3x_3 \quad (1)$$

where $\hat{y}$ is our best guess at a horse's speed, x_1 is its percentage of victories in previous races, x_2 is its jockey's percentage of winning races, and x_3 is the weight it will carry in the present race. Assuming that standardized scores¹ are used, the b_i reflect the importance of the different factors. If $b_1 = 2b_2$, then a given change in the horse's percentage of wins affects our speed prediction twice as much as an

equivalent change in jockey's percentage, because past performances have proved twice as sensitive to x_1 as x_2 .

Sounds easy, but there are a thousand pitfalls. One emerges when the predictors (x_i) are correlated, as might (and in fact does) happen were winning horses to draw winning jockeys or vice versa. In such cases of multicollinearity, each variable has some independent ability to explain past performance and the two have some shared ability. When the weights are determined, that shared explanatory capacity will somehow be split between the two. Typically, that split renders the weights (b_i) uninterpretable with any degree of precision. Thus the regression equation cannot be treated as a theory of horse racing, showing the importance of various factors.

A more modest theoretical goal would simply be to determine which factors are and which factors are not important, on the basis of how much each adds to our understanding of y . The logic here is that of stepwise regression; additional variables are added to the equation as long as they add something to its overall predictive (or explanatory) power. Yet even this minimalistic strategy can run afoul of multicollinearity. If many reflections of a particular factor (e.g., different aspects of breeding) are included, their shared explanatory ability may be divided up into such small pieces that no one aspect makes a "significant" contribution.

Of course, these nuances may be of relatively little interest to handicappers as long as the formula works well enough to help them somewhat in beating the odds. We scientist types, however, want wisdom as well as efficacy

from our techniques. It is hard for us to give up interpreting weights. Regression procedures not only express, but also produce, understanding (or, at least, results) in a mechanical, repeatable fashion. Small wonder then that they have been pursued doggedly despite their limitations. One of the best documented pursuits has been in the study of clinical judgment. Clinical judgment is exercised by a radiologist who sorts X rays of ulcers into "benign" and "malignant," by a personnel officer who chooses the best applicants from a set of candidates, or by a crisis center counselor who decides which callers threatening suicide are serious. In each of these examples, the diagnosis involves making a decision on the basis of a set of cues or attributes. When, as in these examples, the decision is repetitive and all cases can be characterized by the same cues, it is possible to model the judge's decision-making policy statistically. One collects a set of cases for which the expert has made a summary judgment (e.g., benign, serious) and then derives a regression equation, like (1), whose weights show the importance the judge has assigned to each cue.

Two decades of such policy-capturing studies persistently produced a disturbing pair of conclusions: (a) simple linear models, using a weighted sum of the cues, did an excellent job of postdicting judges' decisions, although (b) the judges claimed that they were using much more complicated strategies (Goldberg, 1968, 1970; Slovic & Lichtenstein, 1971). A commonly asserted form of complexity is called "configural" judgment, in which the diagnostic meaning of one cue depends upon the meaning of other cues (e.g., "that tone of voice makes me think 'not suicidal' unless the call comes in the early hours of the morning").

Two reasons for the conflict between measured and reported judgment policies have emerged from subsequent research, each with negative implications for the usefulness of regression modeling for "capturing" the wisdom of past decisions. One was the growing realization that combining enormous amounts of information in one's head, as required by such formulae, overwhelms the computational capacity of anyone but an idiot savant. A judge trying to implement a complex strategy simply would not be able to do so with great consistency. Indeed, it is difficult to learn and use even a non-configural, weighted sum, decision rule when there are many cues or unusual relationships between the cues and predicted variable (Slovic, 1974).

The second realization that has emerged from clinical judgment research is that simple linear models are extraordinarily powerful predictors. As long as one can identify and measure the attributes relevant to an individual, one can mimic his or her decisions to a large degree with simple models bearing no resemblance to actual cognitive processes. That is, under very general conditions, one can misspecify weights and even combinations rules and still do a pretty good job of predicting decisions (Dawes, 1979). Thus, whatever people are doing will look like the application of a simple linear model. In Hoffman's (1960) term, such models are paramorphic in that they reproduce the input-output relations of the phenomena they are meant to describe without any guarantee of fidelity to the underlying processes.

Empirically discovering an analytical result by Wilks (1938), Dawes and Corrigan (1974) showed that considerable predictive success is possible without almost any modeling at all. All one has to do is to identify the variables (or

attributes) to which a decision maker attends and decide whether they are positively or negatively related to the decision criterion. If these variables are expressed in standard units, they can be given unit weights (+1 or -1, as appropriate). Such a unit weighted model will, under very general conditions, predict decisions as well as a full-blown regression model does.

Thus, a simple substantive theory indicating what variables people care about when making decisions may be all one needs to make pretty good predictions of their behavior. If some signs encourage a diagnosis or decision and others discourage it, simply counting the number of encouraging and discouraging signs will provide a pretty good guess at the individual's behavior. The result, however, will be a more modest theory than one can derive by flashy regression modeling.

Obviously, some factors are more important than others. Therefore, a theory using importance weights should be more faithful to reality than one using unit weights. However, any unreliability or misspecification of those weights due to poor procedure or multicollinearity reduces their usefulness very quickly. Indeed, models using poorly conceived or executed weighting schemes may succeed in spite of rather than because of their increased sophistication (Fischhoff, Goitein & Shapira, in press). Thus, while the past seems to be right out there to be understood, our standard statistical procedures don't always tell us what we want to know. If not used carefully, they may mislead us, leaving us less wise than when we started. It is tempting to embrace highly complicated theories in their entirety without realizing that their power comes from very simple underlying notions, rather than from having captured the essence of the past.

LOOKING FOR FOLLY

Focus on Failure

Searching for wisdom in historic events requires an act of faith, belief in the existence of recurrent patterns waiting to be discovered. Searching for wisdom in the behavior of historical characters requires a somewhat different act of faith, confidence that our predecessors knew things we don't know. The first of these faiths is grounded in philosophy; it distinguishes those who view history as a social science, not an ideographic study of unique events. The second of these faiths is grounded in charity and modesty. It distinguishes those who hope to see further by standing on the shoulders of those who came before and those satisfied with standing on their faces. Aphorisms like "those who do not study history are condemned to repeat it" suggest that the latter faith is relatively rare.

An active search for folly is, of course, not without merit. Not only do individuals for whom things do not go right often have a lot of explaining to do, but such explanations are crucial to learning from their experience. By seeing how things went wrong, we hope to make them go right in the future. The quest for misfortunes to account for is hardly difficult. The eye, journalist, and historian are all drawn to disorder. An accident-free drive to the store or a reign without wars, depressions or earthquakes are for them uneventful.

Although it has legitimate goals, focus on failure is likely to mislead us by creating a distorted view of the

prevalence of misfortune. The perceived likelihood of events is determined in part by the ease with which they are imagined and remembered (Tversky & Kahneman, 1973). Belaboring failures should, therefore, disproportionately enhance their perceived frequency in the past (and perhaps future).

It is also likely to promote an unbalanced appraisal of our predecessors' performance. The muckracker in each of us is drawn to stories of welfare cheaters or the "over-regulation" of particular environmental hazards (e.g., the Occupational Safety and Health Administration's infamous standard for a workplace toilet-seat design). We tend to forget, though, that any fallible, but not diabolical, decision-making system produces errors of both kinds. For every cheater garnering undeserved benefits, there are one or several or a fraction of cheatees, denied their rights by the same imperfect system. In fact, the two error rates are tied in a somewhat unintuitive fashion dependent upon the accuracy of judgment and the total resources available, i.e., the percentage of eligible indigent or hazards that can be treated (Einhorn, 1978). Before rushing to criticize the welfare system for allowing a few cheaters, we should consider whether or not there might not be too few horror stories of that type, given the ratio of errors of commission to errors of omission.

In general, there is a good chance of being misled when we examine in isolation decisions that only "work out" on a percentage-wise basis.

What Was the Problem?

There are other contexts in which errors in the small may look different when some larger context is considered. For

example, we are taught that scientific theories should roll over dead once any inconsistent evidence is present. As a result, we are quick to condemn the folly of scientists who persist in their theories despite having been "proved" wrong. Kuhn (1962), however, argued that such local folly might be consistent with more global wisdom in the search for scientific knowledge. Others (e.g., Feyerabend, 1975; Lakatos, 1970) have, in fact, extolled the role of disciplined anarchy in the growth of understanding and doubted the possibility of wisdom emerging from orderly adherence to any one favored research method. They argue that obstinate refusal to look at contrary evidence or to abandon apparently disconfirmed theories is often necessary to scientific progress.

The \$125 million dollar settlement levied against Ford Motor Company in the Pinto case made the company's decision to save a few dollars in the design of that car's fuel tank seem like folly. Yet in purely economic terms, a guaranteed saving of, say \$50 on each of one million Pintos makes the risk of a few large law suits seem like a more reasonable gamble. Since the judgment in this well-publicized suit was reduced to \$6 million upon appeal, the company may actually be ahead in strict economic terms, despite having had worst come to worst. Where the company may be faulted is in seeing one larger context (the number of cars on which it would save money), but not another (the non-economic consequences of its decision). It seems not to have realized the impact that adverse publicity would have on Ford's image as a safety-conscious auto maker, or on prices for used Pintos (although that price was borne by Pinto owners, not producers). Similarly, one may be charitable with NASA for losing the gamble that it would cost less to attempt to rescue Skylab should it begin to descend than to install correcting rockets. It may be harder, though,

to excuse the agency's decisions to threaten unnecessarily the lives of earthlings, by which "NASA started a game of Russian roulette. Even if no one is hurt, the United States loses. Civilized people do not throw rocks from tall buildings even if the odds are good that no one will be hurt" (New York Times, July 8, 1979).

If reprobation is the name of the game, a mistake is a mistake. Yet, if one is interested in learning from the experience of others, it is quite important to determine what problem they were attempting to solve. Upon careful examination, many apparent errors prove to represent deft resolution of the wrong problem. For example, if they are to be criticized at all, Ford and NASA might be held guilty of tactical wisdom and strategic folly (or perhaps of putting institutional health over societal well-being).

This distinction is important, not only for evaluating the past, but also for knowing what corrective measures need to be taken in the future. Usually, tactical mistakes are easier to correct than strategic misunderstandings. Once we've properly characterized a situation, there may be a "book," recording conventional wisdom as accumulated through trial-and-error experience, or at least formulae for optimally combining the information at our disposal (Hexter, 1971). Baseball managers, for example, may either know that it has proven successful to have the batter sacrifice with a runner on first and no one out in a close game or else have the statistics needed to calculate how to "go with the percentages." These guides are, however, unhelpful or misleading if the real problem to be solved is maintaining morale (the runner has a chance to lead the league in stolen bases) or aiding the box

office (the fans need to see some swinging). Studies of surprise attacks in international relations reveal that surprised nations have often done a good job of playing by their own book, but have misidentified the arena in which they were playing (Ben Zvi, 1976; Lanir, 1978). In a sense, they were reading the wrong book; the better they read, the quicker they met their demise.

One reason for the difficulty posed by strategic problems is that they must be "thought through" analytically, without the benefit of cumulative (statistical) experience. A second limitation is that misconceptions are often widely shared within a decision-making group or community. One is consulted on decisions only after one has completed the catechism in the book. Recurrent pieces of advice for institutions interested in avoiding surprises are (a) set up several separate analytical bodies in order to provide multiple, independent looks at a problem or (b) appoint one member to serve as "devil's advocate" for unpopular points of view (Janis, 1972). In practice, the first strategy may fail because shared misconceptions make the groups very like one another, creating redundancy rather than pluralism (Chan, 1979). The second fails because advocates either bow to group pressure or are ostracized if they take their unpopular positions seriously, even when those "extreme" positions do not drastically challenge group preconceptions.

Failure to distinguish between tactical and strategic decisions can also create an undeserved illusion of savvy. Banks and insurance companies are usually considered to be extremely rational and adroit in their decision-making processes. Yet a closer look reveals that this reputation comes from their success in making highly repetitive, tactical decisions in which they almost can't lose. Home mortgages and life insurance

policies are issued on the basis of conservative interpretations of statistical tables acquired and adjusted through massive trial-and-error experience. These institutions' ventures into more speculative decisions requiring analytical, strategic decisions suggest that they are no smarter than the rest of us. Commercial banks lost large sums of money in the 1960's through unwise investments in real estate investment trusts; a similarly minute percentage of their overall decisions in the 1970's has chained the U.S. economy to the future of semi-solvent Third World countries to whom enormous (\$60+ billion) loans have been made. (Although this linkage may be for the long-range good of humanity, that wasn't necessarily the problem the banks were solving.) The slow and erratic response of life insurance companies to changes in the economics of casualty insurance and their almost haphazard, non-analytical methods for dealing with many non-routine risks should leave the rest of us feeling not so stupid when compared with these vaunted institutions.

Hindsight: Thinking Backwards?

Assuming that we know what has happened and what problem an individual was trying to solve, we are then in a position to exploit the wisdom of our own hindsight in explaining and evaluating his or her behavior. Upon closer examination, however, the advantages of knowing how things turned out may be oversold (Fischhoff, 1975). In hindsight, people consistently exaggerate what could have been anticipated in foresight. They not only tend to view what has happened as being inevitable, but also to view it as having appeared "relatively inevitable" before it happened. People believe that others should have been able to anticipate events much

better than was actually the case. They even misremember their own predictions so as to exaggerate in hindsight what they knew in foresight (Fischhoff & Beyth, 1975).

As described by historian Georges Florovsky (1969):

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).

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

One collary tendency is to telescope the rate of historical processes, exaggerating the speed with which "inevitable" changes are consummated (Fischer, 1970). For example, people may be able to point to the moment when the latifundia were doomed, without realizing that they took two and a half centuries to disappear. Another is the tendency to remember people as having been much more like their current selves than was actually the case (Yarrow, Campbell & Burton, 1970). A third may be seen in Barraclough's (1972) critique of the historiography of the ideological roots of Naziism. Looking back from the Third Reich, one can trace its roots to the writings of many authors whose writings one could not have projected Naziism. A fourth is to imagine that

the participants in a historical situation were fully aware of its eventual importance ("Dear Diary, The Hundred Years' War started today," Fischer, 1970). A fifth is the myth of the critical experiment, unequivocally resolving the conflict between two theories or establishing the validity of one. In fact, "the crucial experiment is seen as crucial only decades later. Theories don't just give up, since a few anomalies are always allowed. Indeed, it is very difficult to defeat a research programme supported by talented and imaginative scientists" (Lakatos, 1970, pp. 157-8).

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

Consider decision makers who have been caught unprepared by some turn of events and who try to see where they went wrong by recreating their pre-outcome knowledge state of mind. If, in retrospect, the event appears to have seemed relatively likely, they can do little more than berate themselves for not taking the action that their knowledge seems to have dictated. They might be said to add the insult of regret to the injury inflicted by the event itself. When second-guessed by a hindsightful observer, their misfortune appears as incompetence, folly, or worse.

In situations where information is limited and indeterminate, occasional surprises and resulting failures are

inevitable. It is both unfair and self-defeating to castigate decision makers who have erred in fallible systems, without admitting to that fallibility and doing something to improve the system. According to historian Roberta Wohlstetter (1962), the lesson to be learned from American surprise at Pearl Harbor is that we must "accept the fact of uncertainty and learn to live with it. Since no magic will provide certainty, our plans must work without it" (p. 401).

When we attempt to understand past events, we implicitly test the hypotheses or rules we use both to interpret and to anticipate the world around us. If, in hindsight, we systematically underestimate the surprises that the past held and holds for us, we are subjecting those hypotheses to inordinately weak tests and presumably, finding little reason to change them. Thus, the very outcome knowledge which gives us the feeling that we understand what the past was all about may prevent us from learning anything from it.

Protecting ourselves against this bias requires some understanding of the psychological processes involved in its creation. It appears that when we receive outcome knowledge, we immediately make sense out of it by integrating it into what we already know about the subject. Having made this reinterpretation, the reported outcome now seems a more or less inevitable outgrowth of the reinterpreted situation. "Making sense" out of what we're told about the past is, in turn, so natural that we may be unaware of outcome knowledge having had any effect on us. Even if we are aware of there having been an effect, we may still be unaware of exactly what it was. In trying to reconstruct our foresightful state of mind, we will remain anchored in our hindsightful perspective, leaving the reported outcome too likely looking.

As a result, merely warning people about the dangers of hindsight bias has little effect (Fischhoff, 1977). A more effective manipulation is to force oneself to argue against the inevitability of the reported outcomes, that is, try to convince oneself that it might have turned out otherwise. Questioning the validity of the reasons you have recruited to explain its inevitability might be a good place to start (Koriat, Lichtenstein & Fischhoff, 1980; Slovic & Fischhoff, 1977). Since even this unusual step seems inadequate, one might further try to track down some of the uncertainty surrounding past events in their original form. Are there transcripts of the information reaching the Pearl Harbor Command prior to 7 am on December 7? Is there a notebook showing the stocks you considered before settling on Waltham Industries? Are there diaries capturing Chamberlain's view of Hitler in 1939? An interesting variant was Douglas Freeman's determination not to know about any subsequent events when working on any given period in his definitive biography of Robert E. Lee (Commager, 1965). Although admirable, this strategy does require some naive assumptions about the prevalence of knowledge regarding who surrendered at Appomattox.

LOOKING AT ALL

Why Look?

Study of the past is predicated on the belief that if we look, we will be able to discern some interpretable patterns. Considerable research suggests that this belief is well founded. People seem to have a remarkable ability to find some order or meaning in even randomly produced data. One of the most

familiar examples is the gamblers' fallacy. Our feeling is that in flipping a fair coin, four successive "heads" will be followed by a "tail" (Lindman & Edwards, 1961). Thus in our minds, even random processes are constrained to have orderly internal properties. Kahneman and Tversky (1972) have suggested that of the 32 possible sequences of 6 binary events only one actually looks "random."

Although the gamblers' fallacy is usually cited in the context of piquant but trivial examples, it can also be found in more serious attempts to explain historical events. For example, after cleverly showing that Supreme Court vacancies appear more or less at random (according to a Poisson process), with the probability of at least one vacancy in any given year being .39, Morrison (1977) claimed that:

[President] Roosevelt announced his plan to pack the Court in February, 1937, shortly after the start of his fifth year in the White House. 1937 was also the year in which he made his first appointment to the Court. That he had this opportunity in 1937 should come as no surprise, because the probability that he would go five consecutive years without appointing one or more justices was but .08, or one chance in twelve. In other words, when Roosevelt decided to change the Court by creating additional seats, the odds were already eleven to one in his favor that he would be able to name one or more justices by traditional means that very year (pp. 143-4).

However, if vacancies do appear at random, then this reasoning is wrong. It assumes that the probabilistic process

creating vacancies, like that governing coin flips, has a memory and a sense of justice, as if it knows that it is moving into the fifth year of the Roosevelt presidency and that it "owes" FDR a vacancy. However, on January 1, 1937, the past four years were history, and the probability of at least one vacancy in the coming year was still .39 (Fischhoff, 1978).

Feller (1968) offers the following anecdote involving even higher stakes: Londoners during the blitz devoted considerable effort to interpreting the pattern of German bombing, developing elaborate theories of where the Germans were aiming (and when to take cover). However, when London was divided up into small, contiguous geographic areas, the frequency distribution of bomb-hits per area was almost a perfect approximation of the Poisson distribution. Natural disaster constitutes another category of consequential events where (threatened) lay people see order when experts see randomness (Kates, 1962).

One secret to maintaining such beliefs is failure to keep complete enough records to force ourselves to confront irregularities. Historians acknowledge the role of missing evidence in facilitating their explanations with comments like "the history of the Victorian Age will never be written. We know too much about it. For ignorance is the first requisite of the historian--ignorance which simplifies and clarifies, which selects and omits, with placid perfection unattainable by the highest art" (Strachey, 1918, preface).

Even where records are available and unavoidable, we seem to have a remarkable ability to explain or provide a causal interpretation for whatever we see. When events are

produced by probabilistic processes with intuitive properties, random variation may not even occur to us as a potential hypothesis. For example, the fact that athletes chastized for poor performance tend to do better the next time out fits our naive theories of reward and punishment. This handy explanation blinds us to the possibility that the improvement is due instead to regression to those players' mean performance (Furby, 1973; Kahneman & Tversky, 1973).

Fama (1965) has forcefully argued that the fluctuations of stockmarket prices are best understood as reflecting a random walk process. Random walks, however, have even more unintuitive properties than the binary processes to which they are formally related (Carlsson, 1972). As a result, we find that market analysts have an explanation for every change in price, whether purposeful or not. Some explanations, like those shown in Figure 1, are inconsistent;² others seem to deny the possibility of any random component, for example, that ultimate fudge factor, the "technical adjustment."

The pseudo-power of our explanations can be illustrated by analogy with regression analysis. Given a set of events and a sufficiently large or rich set of possible explanatory factors, one can always derive postdictions or explanations to any desired degree of tightness. In regression terms, by expanding the set of independent variables one can always find a set of predictors with any desired correlation with the independent variable. The price one pays for overfitting is, of course, shrinkage, failure of the derived rule to work on a new sample of cases. The frequency and vehemence of methodological warnings against overfitting suggest that correlational overkill is a bias that is quite resistant to even extended professional training (for references, see Fischhoff and Slovic, in press).

One way of thinking of an overfitted theory is like a suit tailored so precisely to one individual in one particular pose that it will not fit anyone else or even that same individual in the future or even in the present if new evidence about him comes to light (e.g., he lets out his breath to reveal a pot-belly). An historian who had built an air-tight case accounting for all available evidence in explaining how the Bolsheviks won might be in a sad position were the USSR to release suppressed documents showing that the Mensheviks were more serious adversaries than had previously been thought. The price investment analysts pay for overfitting is their long-run failure to predict any better than market averages (Dreman, 1979)--although the cynic might say that they actually make their living through the generation of hope (and commissions).³

Overfitting works because of capitalization on chance fluctuations. If measurement is sufficiently fine, two cases differing on one variable will also differ on almost any other variables one chooses to name. As a result, one can calculate a non-zero (actually, in this case, perfect) correlation between the two variables and derive an "interesting" substantive theory. Processes analogous to this two-dimensional case work with any m observations in the n -space defined by our set of possible explanatory concepts.

In these examples, the data are fixed and undeniable, while the set of possible explanations is relatively unbounded; one hunts until one finds an explanation that fits. Another popular form of capitalization on chance leaves the set of explanations fixed (usually at one candidate) and sifts through data until supporting evidence is found. While the crasser

forms of this procedure are well known, others are more subtle and even somewhat ambiguous in their characterization. For example, you run an experiment and fail to receive an anticipated result. Thinking about it, you note an element of your procedure that might have mitigated the effect of the manipulated variable. You correct that; again no result, but again a possible problem. Finally, you (or your subjects) get it right and the anticipated effect is obtained. Now, is it right to perform your statistical test on that n'th sample (for which it shows significance) or the whole lot of them? Had you done the right experiment first, the question wouldn't even have arisen. Or, as a toxicologist, you are "certain" that exposure to Chemical X is bad for one's health, so you compare workers who do and do not work with it in a particular plant for bladder cancer, but still no effect. So you try intestinal cancer, emphysema, dizziness, ... , until you finally get a significant difference in skin cancer. Is that difference meaningful? Of course, the way to test these explanations or theories is by replication on new samples. That step, unfortunately, is seldom taken and often not possible for technical or ethical reasons (Tukey, 1977).

Related complications can arise even with fixed theories and data sets. Diaconis (1978) notes the difficulty of evaluating the surprisingness of ESP results, even in the rare cases in which they have been obtained in moderately supervised settings, because the definition of the sought event keeps shifting. "A major key to B.D.'s success was that he did not specify in advance the result to be considered surprising. The odds against a coincidence of some sort are dramatically less than those against any prespecified particular one of them" (p. 132).⁴

Tufte and Sun (1975) discovered that the existence or non-existence of bellwether precincts depends upon the creativity and flexibility allowed in defining the event (for what office, in what elections, how good is good, are precincts that miss consistently to be included?). They are commonly believed to exist because we have an uncommonly good ability to find a signal even in total noise.

Have We Seen Enough?

Given that we are almost assured of finding something interpretable when we look at the past, our next question becomes "have we understood it?" The hindsight research described earlier suggests that we are not only quick to find order, but also poised to feel that we knew it all along in some way, or would have been able to predict the result had we been asked in time. Indeed, the ease with which we discount the informativeness of anything we are told makes it surprising that we ever ask the past, or any other source, many questions. This tendency is aggravated by tendencies (a) not to realize how little we know or are told, leaving us unaware of what questions we should be asking in search of surprising answers (Fischhoff, Slovic & Lichtenstein, 1977, 1978) and (b) to draw far-reaching conclusions from even small amounts of unreliable data (Kahneman & Tversky, 1973; Tversky & Kahneman, 1971).

Any propensity to look no further is encouraged by the norm of reporting history as a good story, with all the relevant details neatly accounted for and the uncertainty surrounding the event prior to its consummation summarily buried, along with any confusion the author may have felt (Gallie, 1964; Nowell-Smith, 1970). Just one of the secrets

to doing this is revealed by 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).⁵

Although an intuitively appealing goal, the construction of coherent narratives exposes the reader to some interesting biases. A completed narrative consists of a series of somewhat independent links, each fairly well established. The truth of the narrative depends upon the truth of the links. Generally, the more links there are, the more detail in each link, the less likely the story is to be correct in its entirety. However, Slovic, Fischhoff and Lichtenstein (1976) have found that adding detail to an event description can increase its perceived probability of occurrence, evidently by increasing its thematic unity. Bar-Hillel (1973) found that people consistently exaggerate the probability of the conjunction of a series of likely events. For example, her subjects generally preferred a situation in which they would receive a prize if seven independent events each with a probability of .90 were to occur to a situation in which they would get the same prize if a fair coin fell on "heads." The probability of the compound event is less than .50, whereas the probability of the single event is .50. In other words, uncertainty seems to accumulate at much too slow a rate.

What happens if the sequence includes one or a few weak or unlikely links? The probability of its weakest link should set an upper limit on the probability of an entire narrative. Coherent judgments, however, may be compensatory, with the coherence of strong links "evening out" the incoherence of weak links. This effect is exploited by attorneys who bury

the weakest link in their arguments near the beginning of their summations and finish with a flurry of convincing, uncontestable arguments.

Coles (1973) presents a delicious example of the overall coherence of a story obscuring the unlikelihood of its links: Freud's most serious attempt at psychohistory was his biography of Leonardo DeVinci. For years, Freud had sought the secret to understanding Leonardo, whose childhood and youth were basically unknown. Finally, he discovered a reference by Leonardo to a recurrent memory of a vulture touching his lips while he was in the cradle. Noting the identity of the Egyptian hieroglyphs for "vulture" and "mother" and other circumstantial evidence, Freud went on to build an imposing and coherent analysis of Leonardo. While compiling the definitive edition of Freud's works, however, the editor discovered that the German translation of Leonardo's recollection (originally in Italian) which Freud had used was in error, and that it was a kite and not a vulture which had stroked his lips. Despite having the key to Freud's analysis destroyed, the editors decided that the remaining edifice was strong enough to stand alone. As Hexter (1971) observed, "Partly because writing bad history is pretty easy, writing very good history is rare" (p. 59).

CONCLUSION

What general lessons can we learn about the study of the past, beyond the fact that understanding is more elusive than may often be acknowledged?

Presentism

Inevitably, we are all captives of our present personal perspective. We know things that those living the past did not. We use analytical categories (e.g., feudalism, Hundred Years War) that are meaningful only in retrospect (Brown, 1974). We have our own points to prove when interpreting a past which is never sufficiently unambiguous to avoid the imposition of our ideological perspective (Degler, 1976). Historians do "play new tricks on the dead in every generation" (Becker, 1935).

There is no proven antidote to presentism. Some partial remedies can be generalized from the discussion of how to avoid hindsight bias when second-guessing the past. Others appear in almost any text devoted to the training of historians. Perhaps the most general messages seem to be (a) knowing ourselves and the present as well as possible; "the historian who is most conscious of his own situation is also most capable of transcending it" (Croce, quoted in Carr, 1961, p. 44); and (b) being as charitable as possible to our predecessors; "the historian is not a judge, still less a hanging judge" (Knowles, quoted in Marwick, 1970, p. 101).

Methodism

In addition to the inescapable prison of our own time, we often further restrict our own perspective by voluntarily adopting the blinders that accompany strict adherence to a single scientific method. Even when used judiciously, no one method is adequate for answering many of the questions we put to the past. Each tells us something and misleads us somewhat. When we do not know how to get the right answer to a question,

an alternative epistemology is needed: use as broad a range of techniques or perspectives as possible, each of which enables us to avoid certain kinds of mistakes. This means a sort of interdisciplinary cooperation and respect different from that encountered in most attempts to comingle two approaches. Matches or mismatches like psychohistory too often are attempted by advocates insensitive to the pitfalls in their adopted fields (Fischhoff, in press). Hexter (1971) describes the historians involved in some such adventures as "rats jumping aboard intellectually sinking ships" (p. 10).

Learning

Returning to Benson, if we want the past to serve the future, we cannot treat it in isolation. The rules we use to explain the past must also be those we use to predict the future. We must cumulate our experience with a careful eye to all relevant tests of our hypotheses. One aspect of doing this is compiling records that can be subjected to systematic statistical analysis; a second is keeping track of the deliberations preceding our own decisions, realizing that the present will soon be past and that a well-preserved record is the best remedy to hindsight bias; a third is to make predictions which can be evaluated. One disturbing lesson from Three Mile Island is that it is not entirely clear what that ostensibly diagnostic event told us about the validity of the Reactor Safety Study (U.S. Nuclear Regulatory Commission, 1975) which attempted to assess the risks from nuclear power; a fourth is to get a better idea of the validity of our own feelings of confidence, insofar as confidence in present knowledge controls our pursuit of new information and interpretation (Fischhoff, Slovic & Lichtenstein, 1977). Thus, we want to structure our lives so as to facilitate learning.

Indeterminacy

To the end, though, there may be no answers to many of the questions we are posing. Some are ill-formed. Others just cannot be answered with existing or possible tools. As much as we would like to know "how the pros do it," there may be no way statistically to model experts' judgmental policies to the desired degree of precision with realistic stimuli. Our theories are often of "such complexity that no single quantitative work could even begin to test their validity" (O'Leary et al., 1974, p. 228). When groups we wish to compare on one variable also differ on another, there is no logically sound procedure for equating them on that nuisance variable (Meehl, 1970). When we have tried many possible explanations on a fixed set of data, there is no iron-clad way of knowing just how many degrees of freedom we have used up, just how far we have capitalized on chance (Campbell, 1975). When we use multiple approaches, the knowledge they produce never converges neatly. In the end, we may have to adopt Trevelyan's philosophical perspective that "several imperfect readings of history are better than none at all" (cited in Marwick, 1970, p. 57).

FOOTNOTES

1. To standardize scores on a particular variable, one subtracts the mean of all scores from each score and then divides by the standard deviation. The result is a set of scores with a mean of 0 and standard deviation of 1.

2. One of my favorite contrasts is that when the market rises following good economic news, it is said to be responding to the news; if it falls, that is explained by saying that the good news had already been discounted.

3. A friend once took a course in reading form charts from a local brokerage. Each session involved the teaching of 10-12 new cues. When the course ended, 8 sessions and 83 cues later, the instructor was far from exhausting his supply.

4. Diaconis continues, "To further complicate any analysis, several such ill-defined experiments were often conducted simultaneously, interacting with one another. The young performer electrified his audience. His frequently completely missed guesses were generally regarded with sympathy, rather than doubt; and for most observers they seemed only to confirm the reality of B. D.'s unusual powers."

5. Such strategies may affect the spirit as well as the mind, by subjectively enhancing the strength and stability of the status quo and reducing its apparent capacity for change (Marković, 1970).

REFERENCES

- Bar-Hillel, M. On the subjective probability of compound events. Organizational Behavior and Human Performance, 1973, 9, 396-406.
- Barraclough, G. Mandarins and Nazis. New York Review of Books, 1972, 19(6), 37-42.
- Becker, C. Everyman his own historian. American Historical Review, 1935, 40, 221-236.
- Benson, L. Toward the scientific study of history: Selected essays. Philadelphia: Lippincott, 1972.
- Ben Zvi, A. Hindsight and foresight: A conceptual framework for the analysis of surprise attacks. World Politics, 1976, 28, 381-395.
- Brown, E. A. R. The tyranny of a construct: Feudalism and historians of medieval Europe. American Historical Review, 1974, 79, 1063-1088.
- Campbell, D. T. "Degrees of freedom" and the case study. Comparative Political Studies, 1975, 8, 178-193.
- Carlsson, G. Random walk effects in behavioral data. Behavioral Science, 1972, 17, 430-437.
- Carr, E. H. What is history? London: Macmillan, 1961.
- Chan, S. The intelligence of stupidity: Understanding failures in strategic warning. American Political Science Review, 1979, 73, 171-180.
- Chapman, L. J. & Chapman, J. P. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. Journal of Abnormal Psychology, 1969, 74, 271-280.
- Coleman, J. S. The mathematical study of small groups. In H. Solomon (Ed.), Mathematical thinking on the measurement of behavior. Glencoe, Illinois: Free Press, 1960.

- Coles, R. Shrinking history. New York Review of Books, Part I in Feb. 22, 1973, 20, pp. 15-21 and Part II in Mar. 8, 1973, 20, pp. 25-29.
- Commager, H. S. The nature and study of history. Columbus: C. E. Merrill, 1965.
- Dawes, R. M. The robust beauty of improper linear models in decision making. American Psychologist, 1979, 34, 571-582.
- Dawes, R. M. & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81(2), 95-106.
- Degler, C. N. Why historians change their minds. Pacific Historical Review, 1976, 48, 167-189.
- Diaconis, P. Statistical problems in ESP research. Science, 1978, 201, 131-136.
- Dremen, D. Contrarian investment strategy. New York: Random House, 1979.
- Einhorn, H. J. Decision errors and fallible judgment: Implications for social policy. In K. R. Hammond (Ed.), Judgment and decision in public policy formation. Denver, Colorado: Westview Press, 1978.
- Fama, E. F. Random walks in stock market prices. Financial Analysts Journal, 1965, 21, 55-60.
- Feller, W. An introduction to probability theory and its applications. (3rd edition) Vol. 1. New York: Wiley, 1968.
- Feyerabend, P. Against method. New York: NLB, 1975.
- Fischer, D. H. Historians' fallacies. New York: Harper & Row, 1970.
- 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. Perceived informativeness of facts. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 349-358.
- Fischhoff, B. Intuitive use of formal models. A comment on Morrison's "Quantitative Models in History." History and Theory, 1978, 17, 207-210.
- Fischhoff, B. No man is a discipline. In J. Harvey (Ed.), Cognition, social behavior and the environment. Hillsdale, N.J.: Erlbaum, in press.
- Fischhoff, B. & Beyth, R. "I knew it would happen":--Remembered probabilities of once-future things. Organizational Behavior and Human Performance, 1975, 13, 1-16.
- Fischhoff, B., Goitein, B. & Shapira, Z. The experienced utility of expected utility approaches. In N. Feather (Ed.), Expectancy, incentive and action. Hillsdale, N.J.: Erlbaum, in press.
- Fischhoff, B. & Slovic, P. A little learning. . .: Confidence in multicue judgment. Attention and Performance, VIII, in press.
- Fischhoff, B., Slovic, P. & Lichtenstein, S. Knowing with certainty: The appropriateness of extreme confidence. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 552-564.
- Fischhoff, B., Slovic, P. & Lichtenstein, S. Fault trees: Sensitivity of estimated failure probabilities to problem representation. Journal of Experimental Psychology: Human Perception and Performance, 1978, 4, 330-344.
- Florovsky, G. The study of the past. In R. H. Nash (Ed.), Ideas of history (Vol. 2). New York: Dutton, 1969.
- Furby, L. Interpreting regression toward the mean in developmental research. Developmental Psychology, 1973, 8, 172-179.

- Gallie, W. B. Philosophy and the historical understanding.
London: Chatto & Windus, 1964.
- Goldberg, L. R. Simple models or simple processes? Some
research on clinical judgments. American Psychologist,
1968, 23, 483-496.
- Goldberg, L. R. Man vs. model of man: A rationale, plus some
evidence, for a method of improving on clinical
inferences. Psychological Bulletin, 1970, 73, 422-432.
- Harris, R. J. The uncertain connection between verbal theories
and research hypotheses in social psychology. Journal
of Experimental Social Psychology, 1976, 12, 210-219.
- Hexter, J. H. The history primer. New York: Basic Books, 1971.
- Hoffman, P. J. The paramorphic representation of clinical
judgment. Psychological Bulletin, 1960, 47, 116-131.
- Janis, I. Victims of groupthink. Boston: Houghton Mifflin,
1972.
- Jiler, W. How charts can help you in the stock market. New
York: Trendline, 1962.
- 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.
- Kates, R. W. Hazard and choice perception in flood plain
management. Chicago: University of Chicago,
Department of Geography, Research Paper No. 78, 1962.
- Koriat, A., Lichtenstein, S. & Fischhoff, B. Reasons for
confidence. Journal of Experimental Psychology:
Human Learning and Memory, 1980, 6, 107-118.
- Kuhn, T. The structure of scientific revolution. Chicago:
University of Chicago Press, 1962.

- Lakatos, I. Falsification and scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), Criticism and the growth of scientific knowledge. Cambridge: Cambridge University Press, 1970.
- Lanir, Z. Critical reevaluation of the strategic intelligence methodology. Tel Aviv: Center for Strategic Studies, Tel Aviv University, 1978.
- Lindman, H. G. & Edwards, W. Supplementary report: Unlearning the gamblers' fallacy. Journal of Experimental Psychology, 1961, 62, 630.
- Marković, M. Social determinism and freedom. In H. E. Kiefer and M. K. Munitz (Eds.), Mind, science and history. Albany: SUNY Press, 1970.
- Marwick, A. The nature of history. London: Macmillan, 1970.
- Meehl, P. E. Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), Minnesota Studies in the Philosophy of science. Minneapolis: University of Minnesota Press, 1970.
- Morrison, R. J. Franklin D. Roosevelt and the Supreme Court: An example of the use of probability theory in political history. History and Theory, 1977, 16, 137-146.
- Nowell-Smith, P. H. Historical explanation. In H. E. Kiefer and M. K. Munitz (Eds.), Mind, science and history. Albany: SUNY Press, 1970.
- New York Times. What to do about Skylab. July 8, 1979.
- O'Leary, M. K., Coplin, W. D., Shapiro, H. B. & Dean, D. The quest for relevance. International Studies Quarterly, 1974, 18, 211-237.
- Shaklee, H. & Fischhoff, B. Strategies of information search in causal analysis. Decision Research Report 79-1, 1979.
- Slovic, P. Hypothesis testing in the learning of positive and negative linear functions. Organizational Behavior and Human Performance, 1974, 11, 368-376.

- Slovic, P. & Fischhoff, B. On the psychology of experimental surprises. Journal of Experimental Psychology: Human Perception and Performance, 1977, 3, 544-551.
- Slovic, P., Fischhoff, B. & Lichtenstein, S. Cognitive processes and societal risk taking. In J. S. Carroll and J. W. Payne (Eds.), Cognition and social behavior. Potomac, Md.: Erlbaum, 1976.
- Slovic, P. & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.
- Strachey, L. Eminent victorians. New York: G. Putman's Sons, 1918.
- Tawney, R. H. The agrarian problem in the sixteenth century. New York: Franklin, 1961.
- Tufte, E. R. & Sun, R. A. Are there bellwether electoral districts? The Public Opinion Quarterly, 1975, 39, 1-18.
- Tukey, J. W. Some thoughts on clinical trials, especially problems of multiplicity. Science, 1977, 198, 679-690.
- Tversky, A. & Kahneman, D. The belief in the "law of small numbers": Psychological Bulletin, 1971, 76, 105-110.
- Tversky, A. & Kahneman, D. Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 1973, 4, 207-232.
- von Winterfeldt, D. & Edwards, W. Evaluation of complex stimuli using multi-attribute utility procedures. Engineering Psychology Laboratory Technical Report 011313-2-T. Ann Arbor, Michigan: University of Michigan, 1973.
- Ward, W. C. & Jenkins, H. M. The display of information and the judgment of contingency. Canadian Journal of Psychology, 1965, 19, 231-241.

- Wason, P. C. & Johnson-Laird, P. N. Psychology of reasoning: Structure and content. London: Batsford, 1972.
- Wilks, S. S. Weighting systems for linear functions of correlated variables when there is no dependent variable. Psychometrika, 1938, 3, 23-40.
- Wohlstetter, R. Pearl Harbor: Warnings and decision. Stanford, California: Stanford University Press, 1962.
- Yarrow, M., Campbell, J. D. & Burton, R. V. Recollections of childhood: A study of the retrospective method. Monographs of the Society for Research in Child Development, 1970, 35(5).

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER PTR-1091-80-5	2. GOVT ACCESSION NO. AD-A088 699	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) For those condemned to study the past: Reflections on historical judgment		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Baruch Fischhoff		8. CONTRACT OR GRANT NUMBER(s) NO0001-77-012-0130 MDA-903-80-C-0194
9. PERFORMING ORGANIZATION NAME AND ADDRESS Decision Research A Branch of Perceptronics 1201 Oak Street, Eugene, Oregon 97401		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS ARPA Order No. 0000 3831
11. CONTROLLING OFFICE NAME AND ADDRESS Defense Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, Virginia 22217		12. REPORT DATE Aug 1980
		13. NUMBER OF PAGES 41
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) 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 none		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) historical judgment decision making		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The present report provides a framework for studying historical judgment and describes the conclusions that may be drawn from psychological research and historiographic literature, the musings of historians about their own craft. The cumulative picture suggests that the past does not yield its secrets readily. Some identifiable and perhaps correctable problems are: overinterpreting available evidence, unfairly second guessing historical actors, and exaggerating the predictability of future events for which analogs can be identified in the past. These judgmental biases can be found in lay as well as professional students of the past.		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)