Learning from the past looks easy, but there are a thousand pitfalls.

For Those Condemned to Study the Past: Reflections on Historical Judgment

Baruch Fischhoff

"I often think it odd that history should be so dull, for a great deal of it must be invention."
—Catherine Morlund

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: (1) Are there patterns upon which we can capitalize so as to make ourselves wiser in the

My thanks to Lita Furby, Lewis Goldberg, Sarah Lichtenstein, and Paul Slovic for their perceptive comments on earlier drafts.

future? (2) Are there instances of folly in which we can identify mistakes to avoid? (3) Are we really condemned to repeat the past if we do not 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 more 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

Formal 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.

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:

$$\tilde{y} = b_1x_1 + b_2x_2 + b_3x_3 \qquad (1)$$

where $\tilde{y}$ 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. When scores are standardized (by subtracting the mean and then dividing by the standard deviation), 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 to 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 (such as 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 cue, 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 (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: (1) simple linear models, using a weighted sum of the cues, did an excellent job of postdicting judges' decisions, although (2) the judges claimed that they were using much more complicated strategies (Goldberg, 1968, 1970; Slovic and 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 (for example, "that tone of voice makes me think 'not suicidal' unless the call comes in the early hours of the morning").

Two reasons for 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 combination 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 reasonably 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 fair 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, and Shapira, in press). Thus, while the past seems to be right out there to be understood, our standard statistical procedures do not 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

Searching for wisdom in historic events requires an act of faith, a 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 do not 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. Idioms 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.

Assuming that we know what has happened, we are then in a position to exploit the wisdom of our own hindsight in explaining and evaluating the past behavior of others. On 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 and 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 corollary 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 large landed estates (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, and Burton, 1970). A third may be seen in Barraclough's (1972) critique of the historiography of the ideological roots of Nazism. Looking back from the Third Reich, one can trace its roots to the writings of many authors from whose writings one could not have projected

Nazism. 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 do not 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. 157–158).

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 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 about 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 are 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 outcome, that is, try to convince oneself that it might have turned out otherwise. Questioning the validity of the reasons recruited to explain its inevitability might be a good place to start (Koriat, Lichtenstein, and Fischhoff, in press; Slovic and 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 A.M. 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 and 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 thirty-two possible sequences of six 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. 143–144].

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).

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 and 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 are inconsistent: for example, 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. Other explanations seem to deny the possibility of any random factor—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 post-dictions 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 (for instance, he lets out his breath to reveal a pot-belly). An historian who had built an airtight 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, in press)—although the cynic might say that they actually make their living through the generation of hope (and commissions).

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 variable 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 would not 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).

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 (1) 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, and Lichtenstein, 1977, 1978) and (2) to draw far-reaching conclusions from even small amounts of unreliable data (Kahneman and Tversky, 1973; Tversky and 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).

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 obscurring 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 the 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 editor's 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 in the past did not. We use analytical categories (such as 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 (1) 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 (2) 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 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, and Lichtenstein, 1977). Thus we want to structure our lives so as to facilitate learning.

Indeterminacy. In 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 text their validity" (O'Leary and others, 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 ironclad 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).

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.

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

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, Ohio: Merrill, 1965.

Dawes, R. M. "The Robust Beauty of Improper Linear Models in Decision Making." *American Psychologist,* 1979, *34,* 571-582.

Dawes, R. M., and 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.

Dreman, D. *Contrarian Investment Strategy* New York: Random House, in press.

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 ed) Vol. 1. New York: Harper & Row, 1968

Fischer, D. H. *Historian's Fallacies* New York: Harper & Row, 1970.

Fischhoff, B. "Hindsight ≠ 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., and 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., and 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., and Slovic, P. "A Little Learning . . . : Confidence in Multicue Judgment." *Attention and Performance,* VIII, in press.
Fischhoff, B., Slovic, P., and 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., and 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 and 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 Versus Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inference." *Psychological Bulletin,* 1970, *73,* 422-432.
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.
Kahneman, D., and Tversky, A. "Subjective Probability: A Judgment of Representativeness." *Cognitive Psychology,* 1972, *3,* 430-454.
Kahneman, D., and Tversky, A. "On the Psychology of Prediction." *Psychological Review,* 1973, *80,* 237-251.
Kates, R. W. *Hazard and Choice Perception in Flood Plain Management.* Research Paper No. 78. Chicago: University of Chicago, Department of Geography, 1962.
Koriat, A., Lichtenstein, S., and Fischhoff, B. "Reasons for Confidence." *Journal of Experimental Psychology: Human Learning and Memory,* in press.
Lakatos, I. "Falsification and Scientific Research Programmes." In I. Lakatos and A. Musgrave (Eds.), *Criticism and the Growth of Scientific Knowledge.* Cambridge: Cambridge University Press, 1970.
Lindman, H. G., and Edwards, W. "Supplementary Report: Unlearning the Gamblers' Fallacy." *Journal of Experimental Psychology,* 1961, *62,* 630.
Marwick, A. *The Nature of History.* London: Macmillan, 1970.
Meehl, P. E. "Nuisance Variables and the Ex Post Facto Design." In M. Radner and 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: State University of New York Press, 1970.
O'Leary, M. K., Coplin, W. D., Shapiro, H. B., and Dean, D. "The Quest for Relevance." *International Studies Quarterly,* 1974, *18,* 211-237.
Slovic, P. "Hypothesis Testing in the Learning of Positive and Negative Linear Functions." *Organizational Behavior and Human Performance,* 1974, *11,* 368-376.

Slovic, P., and Fischhoff, B. "On the Psychology of Experimental Surprises" *Journal of Experimental Psychology Human Perception and Performance,* 1977, *3,* 544-551.
Slovic, P., Fischhoff, B , and Lichtenstein, S. "Cognitive Processes and Societal Risk Taking." In J. S. Carroll and J. W. Payne (Eds.), *Cognition and Social Behavior.* Hillsdale, N.J.: Erlbaum, 1976.
Slovic, P., and 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: Putnam's, 1918.
Tawney, R. H. *The Agrarian Problem in the Sixteenth Century* New York: Franklin, 1961.
Tufte, E. R., and 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., and Kahneman, D "The Belief in the 'Law of Small Numbers.'" *Psychological Bulletin,* 1971, *76,* 105-110.
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, Calif.: Stanford University Press, 1962.
Yarrow, M., Campbell, J D., and Burton, R. V. "Recollections of Childhood: A Study of the Retrospective Method" *Monographs of the Society for Research in Child Development,* 1970, *35* (5)

Baruch Fischhoff is research scientist at Decision Research, a branch of Perceptronics, 1201 Oak Street, Eugene, Oregon 97401.
His primary interests include the psychology of individual judgment and decision making, subjective aspects of formal decision making procedures, the management of technological hazards, and historical judgment.