Accepting Error to Make Less Error

Hillel J. Einhorn

Center for Decision Research
University of Chicago

In this article I argue that the clinical and statistical approaches rest on different assumptions about the nature of random error and the appropriate level of accuracy to be expected in prediction. To examine this, a case is made for each approach. The clinical approach is characterized as being deterministic, causal, and less concerned with prediction than with diagnosis and treatment. The statistical approach accepts error as inevitable and in so doing makes less error in prediction. This is illustrated using examples from probability learning and equal weighting in linear models. Thereafter, a decision analysis of the two approaches is proposed. Of particular importance are the errors that characterize each approach: myths, magic, and illusions of control in the clinical; lost opportunities and illusions of the lack of control in the statistical. Each approach represents a gamble with corresponding risks and benefits.

On a recent program of Wall Street Week, the eminent economist Milton Friedman was being interviewed by Louis Rukeyser. Mr. Rukeyser, who is a cynic about clinical and statistical prediction, asked Friedman what he would do about the Federal Reserve Board, which has been an object of Friedman's criticism for many years. Without missing a beat, Friedman replied, "I would get rid of them." After expressing surprise, Rukeyser asked, "What would you use to replace them?" "A computer," responded Mr. Friedman. He then went on to explain that the money supply should be set by using a simple rule that is consistently applied. This would, he argued, provide for more stability and certainty in determining economic policy.

Whatever the merits of his argument, I feel confident (i.e., probability = .999) that the idea of replacing the Federal Reserve Board by a computer algorithm will seem absurd and dangerous to most people. Be that as it may, the point of this example is to illustrate that the clinical versus statistical prediction controversy is enduring and general. Not only is the controversy alive and well in economics, but the rapid growth in computer use will spread the conflict to new fields and intensify the battle where it already exists.
The purpose of this article is to understand why the controversy exists and persists. In what follows, I argue that the clinical and statistical approaches rest on quite different philosophical assumptions about the nature of error and the appropriate level of accuracy to be expected in prediction. To examine these issues, a case is made for each approach. Thereafter, a decision analysis is introduced to examine the costs and benefits of subscribing to each position.

THE CLINICAL APPROACH

Clinicians are determinists in their diagnostic activities. That is, symptoms, signs, and the like are viewed as manifestations of underlying causal processes that can be known in principle. Because much clinical reasoning involves diagnosis or backward inference (i.e., making inferences from effects to prior causes), the clinician, like the historian, has much latitude (or degrees of freedom) in reconstructing the past to make the present seem most likely (a kind of maximum likelihood approach, if you will). However, when engaging in prediction or forward inference, one is soon confronted with discrepancies between predicted and actual outcomes. Such discrepancies are often surprising, especially if the explanation for past behavior provides a coherent account of the facts. One is reminded of the unpleasant surprise that awaits the modeler who fits the data with many parameters only to find that the model cannot predict new cases. Thus, it is often the case that the power of post hoc explanation is matched by the paucity of predictive validity.

Given the fluency of causal reasoning, it is not difficult to construct reasons for why discrepancies in prediction occur. Indeed, in hindsight, it seems as if the outcomes could not have been otherwise (see Fishchoff, 1975, on "hindsight bias" as a form of creeping determinism). However, to what degree can (should) prediction errors be explained? It is at this point that the clinical and statistical approaches diverge, with the divergence having much to do with the meaning and significance of "random error."

Although the concept of randomness is complex and difficult to define (Lopes, 1982), it suggests an irreducible unpredictability and disorder of outcomes. The basic question then becomes: How much of behavior is random and how much is systematic? The answer depends on what is meant by randomness. For example, consider the random walk theory of stock market prices. Although most people have not heard of this theory, many have firsthand experience with its implications. Do stock prices follow a random walk? To date, the market remains difficult to predict (many have tried unsuccessfully). However, does that mean that it is impossible to do so? Imagine that there is a seven-way interaction that predicts price changes, but no one has yet induced it from the mass of complex and noisy data that is available. If there is hidden systematicity, one's gamble in searching for a predictive rule may pay off. On the other hand, such an interac-
tion may not exist, despite the fact that there are "experts" selling advice on what stocks to buy (are they simply selling "snake oil"?). Thus, if prediction error is due to our lack of knowledge and if randomness is only a label for our current ignorance, there are at least two reactions. The first is characteristic of the clinical approach; it says that the goal of perfect predictability, although difficult to attain, is not impossible. Moreover, this goal may be useful in itself because it can motivate the search for improved predictability via increased understanding of the causal texture of the environment (Tolman & Brunswik, 1935). The second reaction is characteristic of the statistical approach and emphasizes the possibility of a futile search for a Holy Grail. This is considered in greater detail later.

The importance of causal understanding, which is essential to the clinical approach, has other implications. Although the controversy between the clinical and statistical approaches centers on prediction, has too much attention been given to prediction per se? To illustrate, consider the following scenario:

Imagine that you lived several thousand years ago and belonged to a tribe of methodologically sophisticated cave-dwellers. Your methodological sophistication is such that you have available to you all present day means of the methodological arsenal—details of the principles of deductive logic, probability theory, access to computational equipment, etc. However, your level of substantive knowledge lags several thousand years behind your methodological sophistication. In particular, you have little knowledge about physics, chemistry or biology. In recent years, your tribe has noted an alarming decrease in its birth-rate. Furthermore, the tribe's statistician estimates that unless the trend is shortly reversed, extinction is a real possibility. The tribe's chief has accordingly launched an urgent project to determine the cause of birth. You are a member of the project team and have been assured that all means, including various forms of experimentation with human subjects, will be permitted to resolve this crucial issue. (Einhorn & Hogarth, 1982, p. 23)

The foregoing story illustrates the following points. First, the goal of prediction is to provide guidance for taking action. Therefore, prediction is intimately tied to decision making and should be evaluated within this context. Indeed, one might find small consolation in being able to predict accurately when the tribe will become extinct. Second, when decisions are based on predictions, the determination of forecast accuracy is problematic because outcomes are a function of predictions and actions (Einhorn & Hogarth, 1978). For example, if the president takes strong antirecession measures based on predictions of an economic slowdown, how is one to evaluate the accuracy of the forecast? Consider the outcome of no recession. This could result from an incorrect forecast and a useless action or from an accurate forecast and a highly effective action. Similarly, a recession could indicate an accurate forecast with an ineffective action or an inaccurate forecast with an action that causes the malady it is intended to prevent. Some actions are taken to counteract the prediction of undesirable events, but other actions cause the very outcomes that are predicted. For example:
People in a small town hear a rumor that the banks are about to fail. They think that if this forecast is accurate, they had better withdraw their money as soon as possible. Accordingly, they go to the banks to close their accounts (those sceptical of the forecast see many people withdrawing money and either take this as a sign that the rumor is true or foresee the consequences of waiting too long, thus joining the crowd in either case). By the end of the day the banks have failed, thereby confirming the rumor. (Einhorn & Hogarth, 1982, p. 24)

Note that awareness of such self-fulfilling prophecies is often low and can lead to overconfidence in predictions that are of low or even zero accuracy (see Einhorn, 1980; Einhorn & Hogarth, 1978).

Third, in order to understand the relations between predictions, actions, and outcomes, one needs a causal model of the process. In this regard, the clinical approach, with its emphasis on diagnosis and causal understanding, is important. Moreover, the role of clinical judgment in the development of models and the determination of relevant variables has been sadly neglected. Consider the following conclusion from the literature on clinical versus statistical prediction stated by Dawes and Corrigan (1974): “the whole trick is to decide what variables to look at and then to know how to add” (p. 105). Assuming we can add, how do we decide what variables to look at? Such decisions must rest on some implicit or explicit theory of the phenomenon that allows one to distinguish relevant from irrelevant factors. Therefore, prediction depends on backward inference, which involves both the forming of hypotheses to interpret the past and the choosing of relevant from irrelevant variables in that interpretation.

THE STATISTICAL APPROACH

Although the clinical approach rests on the worthy and optimistic goal of perfect predictability, it is a goal that can have negative consequences, which are noted later. The statistical approach, on the other hand, accepts error. This acceptance can occur in several ways. First, one may believe that the world is inherently uncertain. In this case, probabilistic knowledge is the best we can hope for, and random error cannot be reduced by greater knowledge. Second, one can maintain determinism at the level of the physical world but believe that our knowledge of that world will always be fragmentary and hence uncertain. In this case, randomness is due to ignorance, but the goal of perfect predictability is abandoned as being too unrealistic. Third, the use of any equation or algorithm, with its limited number of variables and mechanical combining rule, can never capture the richness and full complexity of the phenomenon it is meant to predict (recall Meehl's discussion of “broken leg cues,” 1954, p. 25). Thus, models are simplifications of reality that must lead to errors in prediction (cf. Chapinis, 1961).
Let us now consider how the acceptance of error can lead to less error. To do so, recall the research on probability learning done several years ago (e.g., Edwards, 1956; Estes, 1962). In these studies, either a red or green light is illuminated on each of a number of trials, and subjects are asked to predict which light will go on. If the prediction is correct, subjects are given a cash payoff; if the prediction is wrong, there is no payoff. However, unbeknownst to the subject, the lights are programmed to go on according to a binomial process with a given proportion of red and green (e.g., 60% red and 40% green). Thus, the process is random, although subjects do not know this. The major result of these experiments is something called “probability matching”; that is, subjects respond to the lights in the same proportion as they occur. For example, in the aforementioned case, subjects predict red 60% of the time and green 40%. The expected payoff for such a strategy can be calculated as follows: Because the subject predicts red on 60% of the trials and red occurs on 60%, the subject will be correct (and receive the payoff) on 36% of the trials. Similarly for green; the subject predicts green on 40% of the trials, and green occurs on 40%. Hence, 16% of the trials will be correctly predicted. Therefore, over both predictions, subjects will be correct on $36\% + 16\% = 52\%$ of the trials. Now consider how well subjects would do by using a simple rule that said: Always predict the most likely color. Note that such a strategy accepts error; however, it also leads to 60% correct predictions (i.e., I always say red, and red occurs 60% of the time). Inasmuch as 60% is greater than 52%, subjects would make more money if they accepted error and consistently used a simple rule. Indeed, such a rule maximizes their wealth in this situation. However, most are trying to predict perfectly and are engaged in a futile attempt to see patterns in the data that are diagnostic of the (nonexistent) rule that they believe determines the onset of the lights. (The analogy to the stock market is noted without further comment.)

Another example of accepting error to make less error comes from the work on equal or unit weights in linear models (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975). Many people are surprised that equally weighted linear regression models can outpredict models with “optimal” weights on new cases. The reason for the surprise is that we often believe that the weights for the variables are not equal. Thus, the use of equal weights deliberately introduces error into the model. However, there is a benefit from such a procedure, namely that equal weighting protects one against a reversal of the relative weighting of the variables on the basis of poor data. Thus, if $X_1$ and $X_2$ have a true relative weighting of 2:1, equal weights protect one from data showing that the weight for $X_2$ is larger than that for $X_1$. Therefore, if data are of sufficiently poor quality, seeking error can lead to less error in prediction.

Although the idea of trade-offs among errors may be new to some, there are several more mundane advantages of the statistical approach that nonetheless deserve mention. First, the statistical approach demands that empirical evidence, rather than authority, be the deciding factor in determining the predict-
ive accuracy of any device. Hence, the statistical approach is egalitarian: It trusts no one and takes little on faith. In fact, Armstrong's (1978) notion of a "seersucker theory of prediction" seems to capture the attitude of the statistical approach to all undocumented claims of expertise. The theory has only one axiom: For every seer there is a sucker. A second issue concerns inconsistency in judgment due to fatigue, boredom, memory and attentional limitations, and so on. Such inconsistency is not generally useful. Indeed, if someone has a valid rule that is applied inconsistently, predictive accuracy will suffer. However, clinical judgment can be improved by techniques such as "bootstrapping," in which a model of the clinical thought processes predicts more accurately than the person from whom the model was developed (Goldberg, 1970). Such models have been developed in many fields, and the results are encouraging (see Camerer, 1981, for a review and theoretical discussion).

A DECISION ANALYSIS

I have tried to make a case for both the clinical and statistical approaches, and the question naturally arises: Which is better? Such a naive question deserves an answer like: It depends. This section considers some of the factors upon which it depends.

To begin, consider Figure 1, which shows a decision matrix with choices as rows and states of the world as columns. For the sake of simplicity, only two choices and states are shown. First, consider the choice alternatives: One can decide that a phenomenon of interest is systematic and thus capable of being predicted, or one can decide that the phenomenon is random and not predictable. Now consider the possible states of nature. The phenomenon is systematic in the first column, whereas it is random in the second. The intersection of rows and

<table>
<thead>
<tr>
<th>STATES OF NATURE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SYSTEMATIC</strong></td>
</tr>
<tr>
<td>SYST. HIT</td>
</tr>
<tr>
<td><strong>RANDOM</strong></td>
</tr>
<tr>
<td>MYTH, MAGIC, ILLUSIONS OF CONTROL</td>
</tr>
<tr>
<td>LOST OPPORTUNITIES, ILLUSIONS OF NO CONTROL</td>
</tr>
<tr>
<td>HIT</td>
</tr>
</tbody>
</table>

FIG. 1 Decision matrix for comparing the clinical and statistical approaches.
columns results in four possible outcomes: the “hits,” shown in the diagonal, and the errors, shown in the off-diagonal. Note that there are two kinds of errors. If one decides that a phenomenon is systematic when it is random, the error that result is manifest in myths, magic, superstitions, and illusions of control (Langer, 1975). This error is most likely to characterize the clinical approach, which seeks causal explanations for all behavior. Moreover, there are numerous examples of this type of error, which have been discussed in the behavioral decision theory literature (for a review, see Einhorn & Hogarth, 1981; Nisbett & Ross, 1980).

Let us now consider the other error, which is more likely to characterize the statistical approach. In this case, one decides that a phenomenon is random when it is systematic. This error results in lost opportunities and illusions of the lack of control. For example, consider the state of knowledge of the movement of heavenly bodies after Copernicus but before Kepler. The Copernican revolution put the sun at the center of the solar system with the planets revolving in circular orbits. This model of planetary motion gives reasonably accurate predictions. However, we know that the orbits are not circular; they are elliptical, and errors in prediction occurred. If probabilism were around during the time of Copernicus, one might have explained planetary motion as consisting of circular orbits plus a random error term. Such a probabilistic model would explain most of the variance, but it would represent a lost opportunity to understand more clearly the true nature of the phenomenon. Of course, successes in seeking to explain all the variance of behavior are dramatic. However, dramatic failures also exist. Recall Einstein’s famous statement that, “God does not play dice with the world.” His unsuccessful attempts to disprove quantum theory illustrate the difficulty of abandoning the goal of perfect predictability.

What conclusions can be drawn from the foregoing analysis? First, the choice between the clinical and statistical approach in any given situation will depend on:

1. One’s beliefs regarding the probabilities of the states of nature. I have considered only two states, but there are many states representing various levels of systematicity and error. Hence, one’s prior probabilities over the various states will greatly affect the choice of strategy.

2. The relative costs of the two types of errors. For example, to what degree is superstition an appropriate price to pay for not missing an opportunity to predict more accurately?

3. The relative payoffs for the hits/correct choices. Hence, the choice between the clinical and statistical approaches can be seen as a special case of decision making under uncertainty; each has its associated risks and potential benefits. At the least, this conceptualization demonstrates why the controversy will never be resolved. Researchers will “place their bets” differently, whether the field be personality theory or particle physics.
CONCLUSION

The clinical versus statistical controversy represents a basic conflict about the predictability of behavior. The evidence is clear and convincing that the statistical approach does a better job of forecasting, but the clinical approach is not without its virtues. Indeed, I tend to think of the clinical approach as a high risk strategy; that is, the chance of being able to predict all the variance of behavior (or even a substantial amount) is very low, but the payoff is correspondingly high. On the other hand, the acceptance of error to make less error is likely to be a safer and more accurate strategy over a wide range of practical situations. Thus, the statistical approach leads to better performance on the average. In my view, this is a compelling argument for its use.

ACKNOWLEDGMENTS

This research was supported by a contract from the Office of Naval Research.

REFERENCES


Hillel J. Einhorn
Center for Decision Research
Graduate School of Business
University of Chicago
1101 East 58th Street
Chicago, IL 60637