

The Seer-Sucker Theory: The Value of Experts in Forecasting

J. Scott Armstrong

People are willing to pay heavily for expert advice. Economists are consulted to tell us how the economy will change, stock analysts are paid large salaries to forecast the earnings of various companies, and political experts command large fees to tell our leaders what the future holds. The available evidence, however, implies that this money is poorly spent. But because few people pay attention to this evidence, I have come up with what I call the "seersucker theory": "No matter how much evidence exists that seers do not exist, suckers will pay for the existence of seers."

One would expect experts to have reliable information for predicting change and to be able to utilize the information effectively. However, expertise beyond a minimal level is of little value in forecasting change. This conclusion is both surprising and useful, and its implication is clear: Don't hire the best expert, hire the cheapest expert.

This is not to say that experts have no value they can contribute in many ways. One particularly useful role of the expert seems to be in assessing a current situation. And although estimates of current status play an important role in forecasting, I will deal only with the role of expertise in forecasting change.

Value of Experts: The Evidence

Many studies have been done on the value of expertise in a given subject area. Most evidence comes from the field of finance, but studies have also been done in psychology, economics, medicine, sports, and sociology. The relationship of accuracy to expertise in a particular field has been measured in various ways - education, experience, reputation, previous success, and self-identification. Expertise, above a very low level, and accuracy are unrelated (Exhibit 1) and accuracy may even drop after a certain level. This minimal expertise can be obtained quickly and easily.

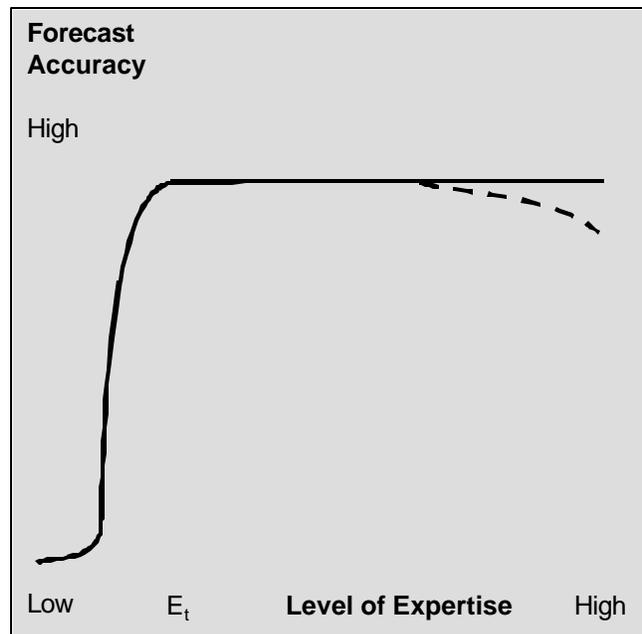
For example, in 1955, R. Taft surveyed 81 psychological studies that examined predictions made by experts and nonexperts. He concluded that nonpsychologists were more capable of making predictions about people's attitudes and behavior.

In one typical study done by Austin Grigg, Ph.D.'s, trainees in psychology, and naive subjects (undergraduates) each listened to 10-minute interviews with three clients and then predicted how each client would fill out three different personality tests. There was no difference in accuracy between the Ph.D.'s and the trainees, but both these groups did significantly better than the naive subjects. Thus, a small amount of expertise was useful.

More recently, Bernard Levy and E. Ulman asked professional mental health workers, students, and people with no mental health experience to distinguish 48 normal people from 48 psychiatric patients by looking at paintings the subjects had done. All the subjects' predictions were significantly better than chance, but here accuracy did not correlate with experience.

Exhibit 1

Some expertise seems to lead to a higher level of accuracy in forecasting change; beyond a minimal level, however, additional expertise does not improve accuracy – and there is even some evidence that it may decrease accuracy.



The performance of experts and novices in forecasting prices of stocks was first examined by Garfield Cox in 1930. He found no advantage for expertise. In 1933 Alfred Cowles examined 255 editorials by Hamilton, an editor of the *Wall Street Journal* who had gained a reputation for successful forecasting. During the period from 1902 to 1929, Hamilton forecasts 90 changes in the market: 45 were correct and 45 were incorrect. Cowles also found that a sample of 20 insurance companies did slightly worse in their investments than the market averages from 1928 to 1931; 16 financial services did slightly worse than the market average from 1928 to 1932; and forecasts in 24 financial publications were slightly worse than the market average over this same period. Other studies, some done as recently as the late 1970s, have reinforced these conclusions.

Roy Johnson and B.F. McNeal had 12 health care professionals – 5 staff psychologists, 6 social workers, and a physician - predict the length of hospital stay for 379 mental patients over an 18-month period. The scores ranged from 63 percent to 86 percent correct. The professionals with more experience in psychology were no more accurate.

William Avison and Gwynn Nettler examined predictions in nine public opinion polls from 1959 to 1971. Experts, as judged from the amount of schooling, were no better at forecasting change.

In three studies in which "expert forecasts" were more accurate, the gain was small:

- A small but statistically significant correlation for a sample of 26 experts in the social and natural sciences in forecasting 123 events in their fields was found by Kaplan, Skogstad, and Girshick in 1950.
- Similar findings were obtained in 1976 by Wise, who examined 1,556 predictions published in the United States between 1890 and 1940. For predictions related to social, technological, economic, and political changes, people with experience in the relevant field seemed to do slightly better than those outside the field.

- In a 1971 study by Robert Winkler, sportswriters did a little better than graduate students and faculty members in forecasting scores of collegiate and professional football games. The bookmakers' forecasts, in turn, were slightly better than those of the sportswriters.

Overall, the evidence suggests there is little benefit to expertise. And because improved accuracy shows up only in large samples, claims of accuracy by a single expert would seem to be of no practical value. Surprisingly, I could find no studies that showed an important advantage for expertise. This evidence does not include every area, however, and further studies may show that the seer-sucker theory cannot be generalized.

Another possibility is that researchers find it easier to publish evidence refuting than confirming the common notion that expertise is useful. However, in light of a 1977 study by Michael Mahoney, this possibility seems remote. Mahoney asked 75 reviewers to referee a paper. Two versions of the paper were presented to randomly selected subsamples of reviewers. The papers differed only in the results: one version had results favoring the common wisdom of the day and the other refuted it. A strong bias was found toward accepting the study that *agreed* with a commonly held hypothesis and rejecting the one that contradicted this hypothesis.

Is Accuracy Irrelevant?

Assume for a moment that the seer-sucker theory is true - that expertise is useless in forecasting change. Is there any rational explanation for why clients continue to purchase worthless information?

One explanation is that the client is not interested in accuracy, but only in avoiding responsibility. A client who calls in the best wizard available avoids blame if the forecasts are inaccurate. The evasion of responsibility is one possible explanation for why stock market investors continue to purchase expert advice in spite of overwhelming evidence that such advice is worthless.

The avoidance of responsibility is illustrated in a 1978 study by Joseph Cocozza and Henry Steadman. In New York, psychiatrists are asked to predict the dangerousness of mental patients - patients diagnosed as dangerous are then placed in involuntary confinement. Although numerous studies have shown that psychiatrists cannot predict who is dangerous, the expert's diagnosis was accepted by 87 percent of the courts in this study. Cocozza and Steadman suggest that their finding may illustrate a belief in magic - that secret knowledge of the specialist can control the unpredictable. The expert advice seems to relieve the court of further responsibility.

Cases involving risk and uncertainty seem most likely to lead to avoidance of responsibility. An example is provided in a study of long-range forecasts of bed requirements for six Michigan community hospitals. Clients were satisfied only when the forecasts matched their preconceptions. When differences arose, the hospital administrators followed their preconceptions anyway, ignoring the advice of experts. The preconceived forecasts exceeded the experts' forecasts for five of the six hospitals and led to decisions that resulted in a 50 percent oversupply of beds.

Ineffective Learning by Seers

"Expertise ... breeds an inability to accept new views." - Laski (1930)

The continued inclination for people to consult expert advice has been the subject of much study. In 1948 B.F. Skinner experimented with a pigeon in a cage. Food was given to the pigeon on a random time schedule. What happened? The bird concluded that a counterclockwise movement produced the food since it was doing that the first time food appeared. It repeated this behavior whenever it was hungry. This initial learning proved to be highly resistant to change, even though it had absolutely nothing to do with the appearance of food.

In 1958 Lloyd Strickland found that people do a good job of simulating pigeons. He had subjects act as managers of two subordinates, whom I will call Stan and Ned. The manager could see Stan's work and communicate with him easily. Communication with Ned was poor. However, both Stan and Ned produced the same amount and

quality of work. The manager trusted Ned but thought Stan required constant supervision - he had concluded that his management efforts were responsible for Stan's output.

In a more recent study, Daniel Kahneman and Amos Tversky discussed a flight school training program in which trainers adopted a recommendation from psychologists that they use only positive reinforcement - they praised successful work and said nothing otherwise. After a time the instructors concluded that positive reinforcement did not work; when they praised someone for successfully completing a series of complex maneuvers, the trainee would often do worse the next time. That happens because learning involves mistakes. A student cannot consistently perform well - an exceptionally good trial will usually be followed by a more average trial, and conversely for an exceptionally poor trial. The flight school trainers noticed this phenomenon and attributed it to their actions. As a result, they "learned" that what works is punishment for bad behavior (because the odds are that the next trial will be better). Rewards, they concluded, just lead to overconfidence of the learner.

In these studies, subjects first are assuming that their own actions control the situation. (This effect shows up even in studies of gamblers.) Second, they are looking for evidence to support their hypotheses; that is, they are looking for confirming evidence and avoiding disconfirming evidence. This latter phenomenon can occur even in the absence of any notion of control or emotional involvement, as shown in the following experiment.

P.C. Wason presented subjects with a three number sequence: 2,4,6. The subjects were told that this sequence had been generated by a rule that the experimenter had in his head. The subjects were then asked to learn the rule by generating additional three-number sequences (e.g., 8,10,12). After each try, the experimenter told the subject if the new sequence agreed with the rule. The subject could generate as many three-number sequences as she wished; when she felt confident of the rule, she wrote it down.

The correct rule was "three numbers in increasing order of magnitude," that is, $a < b < c$. Despite its simplicity, only about 25 percent of the people tested discovered the correct rule. Subjects usually selected a hypothesis (e.g., "add two to each successive number") and looked only for evidence to confirm this hypothesis - they did not attempt to refute it. In other words, most people refuse even to entertain the possibility that they are wrong!

The story gets worse. Subjects who wrote the wrong rule were allowed to try again - to generate additional sets of numbers to obtain more evidence. About half these subjects continued to search for confirmation for the rule they had been told was wrong.

It is not clear whether subjects failed to accept disconfirming evidence because they were unable or because they were unwilling. When asked how they would find out whether their hypothesis was wrong, however, few recognized the need to look for disconfirming evidence by generating a sequence of numbers inconsistent with their hypothesis.

Is Wason correct that people avoid disconfirming evidence? If you believe "yes," I can present confirming evidence to make you happy. If you believe "no," I can give a prima facie argument and you may get upset at me. So I leave it to you to seek disconfirming evidence.

The evidence just cited, however, implies why one might expect expertise to reduce accuracy (as suggested by the dotted line in the chart on page 20). The greater one's feeling of expertise, the less likely that disconfirming evidence will be used. Loren Chapman and J.P. Chapman studied this issue by asking 32 subjects with high expertise to examine data from homosexual and heterosexual subjects. The information was contrived so that there were no relationships among variables that previous literature had found to be irrelevant. Nevertheless, the practicing clinicians saw the relationships that they expected to see (which, incidentally, were the same invalid relationships expected by a group of nonexperts), and they had great difficulty seeing valid relationships even when these were dramatic. In a related study, George Strickler found that although people with high expertise rejected valid disconfirming evidence, subjects with much less expertise improved their accuracy by using disconfirming evidence.

Salvaging the Expert

The seer-sucker theory implies that clients will continue to depend upon experts. It is important, then, to consider whether experts can improve their ability to forecast change. The prospects are not good; evidence reviewed by Nisbett and Wilson shows that experts are often unaware of how they make judgments and predictions. For example, a 1964 study by E.C. Webster showed that decisions in employment interviews are typically made in the first 30 seconds of the interview. Moreover, the reasons for the decision are not usually understood by the interviewer.

Still, there is hope. Detailed instructions for improving judgmental forecasting are provided by Hillel Einhorn and Robin Hogarth, and additional suggestions are given in my book, *Long-Range Forecasting*. Probably the key is to make an active search for disconfirming evidence. Without this search, disconfirming evidence is often ignored, misinterpreted, or misused.

The advice to seek disconfirming evidence is not new - it is the principle behind "objective" scientific experiments. Unfortunately, it is not often used even by scientists, and training does not seem to help. In a study using Wason's 2-4-6 problem, Mahoney and DeMonbreun found that the aversion to disconfirming evidence is just as prevalent among physical scientists as it is among psychologists.

I have recommended an old solution for the problem of finding disconfirming evidence. The method of multiple hypotheses, first suggested by T.C. Chamberlin in 1890, can be used to change one's role from advocate of a particular belief to arbiter

among various competing viewpoints. When using multiple hypotheses, disconfirming evidence for one hypothesis could be confirming evidence for another hypothesis.

Although experts are poor at forecasting change, this does not mean that judgmental forecasting is useless. However, since all available evidence suggests that expertise beyond an easily achieved minimum is of little value in forecasting change, the most obvious advice is to hire inexpensive experts. Also, look for unbiased experts - those who are not actually involved in the situation. Finally, there is safety in numbers. Robin Hogarth has suggested using at least three independent experts and preferably six to ten!

The conditions under which the seer-sucker theory holds are not well known - it may or may not apply to all areas of forecasting. However, in view of the evidence, it seems wise to put the burden of proof upon the experts to show that their expertise in a given area is valuable.

Further Reading

Armstrong, J. Scott, *Long-Range Forecasting: From Crystal Ball to Computer*. New York: Wiley- Interscience, 1978.

Avison, William R. and Nettler, Gwynn, "World views and crystal balls," *Futures*, 8 (1976): 11-21.

Chamberlin, T.C., "The method of multiple working hypotheses." *Science*, 148 (1965): 754-759. Reprinted from *Science*, 1980.

Chapman, Loren J. and Chapman, J.P., "Illusory correlation as an obstacle to the use of valid psychodiagnostic signs," *Journal of Abnormal Psychology*, 74 (1969): 271-280.

Cocozza, Joseph J. and Steadman, Henry J., "Prediction in psychiatry: An example of misplaced confidence in experts," *Social Problems*, 25 (1978): 265-276.

Cowles, Alfred, "Can stock market forecasters forecast?" *Econometrica*, 1 (1933): 309-324.

- Cowles, Alfred, "Stock market forecasting," *Econometrica*, 12 (1944): 206-214.
- Cox, Garfield V., *An Appraisal of American Business Forecasts*. Chicago: University of Chicago Press, 1930.
- Einhorn, Hillel J. and Hogarth, Robin M., "Confidence in judgment: Persistence of the illusion of validity," *Psychological Review*, 85 (1978): 394-416.
- Griffith, J.R. and Wellman, B.T., "Forecasting bed needs and recommending facilities plans for community hospitals," *Medical Care*, 17 (1979): 293-303.
- Grigg, Austin E., "Experience of clinicians and speech characteristics and statements of clients as variables in clinical judgment," *Journal of Consulting Psychology*, 22 (1958): 315-319.
- Hartsough, W. Ross, "Illusory correlation and medicated association: A finding," *Canadian Journal of Behavioral Science*, 7 (1975): 151-154.
- Hogarth, Robin M., "A note on aggregating opinions," *Organizational Behavior and Human Performance*, 21 (1978): 121-129.
- Johnston, Roy and McNeal, B.F., "Statistical vs. clinical prediction: Length of neuropsychiatric hospital stay," *Journal of Abnormal Psychology*, 72 (1967): 335-340.
- Kahneman, Daniel and Tversky, A., "On the psychology of prediction," *Psychological Review*, 80 (1973): 237-251.
- Kaplan A., Skogstad A.L. and Girshick, M.A., "The prediction of social and technological events," *Public Opinion Quarterly*, 14 (Spring 1950): 93-110.
- Langer, Ellen J. and Roth, Jane, "Heads I win, tails it's chance: The illusion of control as a function of the sequence of outcomes in a purely chance task," *Journal of Personality and Social Psychology*, 32 (1975): 951-955.
- Laski, Harold, "Limitations of the expert," *Chemical Technology*, 4 (April 1974): 198-202. Originally from *Harper's*, 1930.
- Levy, Bernard I. and Ulman, E., "Judging psychopathology from paintings," *Journal of Abnormal Psychology*, 72 (1967): 182-187.
- Mahoney, Michael J., "Publication prejudices: An experimental study of confirmatory bias in the peer review system," *Cognitive Therapy and Research*, 1 (1977): 161-175.
- Mahoney, Michael J., and DeMonbreun, B.G., "Psychology of the scientist: An analysis of problem-solving bias," *Cognitive Therapy and Research*, 1 (1977): 229-238.
- Nisbett, Richard E., and Wilson, T.D., "Telling more than we can know: Verbal reports on mental processes." *Psychological Review* 84 (1977): 231-259.
- Skinner, B.F., "Superstition in the pigeon," *Journal of Experimental Psychology*, 38 (1948): 168-172.
- Strickland, Lloyd H., "Surveillance and trust," *Journal of Personality*, 26 (1958): 200-215.
- Strickler, George, "Actuarial, naive clinical, and sophisticated clinical prediction of pathology from figure drawings," *Journal of Consulting Psychology*, 31 (1967): 492-494.
- Taft, R., "The ability to judge people," *Psychological Bulletin*, 52 (1955): 1-28.

Wason, P.C., "On the failure to eliminate hypotheses in a conceptual task," *Quarterly Journal of Experimental Psychology* 12 (1960): 129-140.

Webster, E.C., *Decision Making in the Employment Interview*. Montreal: Eagle, 1964.

Winkler, Robert L., "Probabilistic prediction: Some experimental results," *Journal of the American Statistical Association*, 66 (1971): 675-685.

Wise, George, "The accuracy of technological forecasts, 1890-1940," *Futures*, 8 (1976): 411-419.