identify all the potential variables. We then must provide subjective probability distributions for both the parameters that govern the generation of these variables and also for the parameters that link these variables to the dependent variable under study. For many problems, unfortunately, the identification and assessment problems jointly constitute the most significant costs of dealing with other variables. The observation costs are trivial in comparison, and once we bear the former costs, we almost certainly want to observe and process the complete data set.

I am proposing, therefore, that we behave only *as if* we were formally solving this decision problem. We identify through economic theory and/or introspection certain variables that are potentially important. This is, essentially, the first phase in the formal decision problem. The left-out variables are not, however, formally identified. Instead, we summarize their influence in a contamination parameter $\boldsymbol{\beta}^c$, the prior on which essentially determines the extent to which we are committed to "believe" the regression result.

Just as if we were solving the formal decision problem, we may decide to observe other variables because of either low R^2, peculiar residuals, or peculiar coefficient estimates. At the very least, the probability distribution over the new parameters must imply the original on $\boldsymbol{\beta}^c$.

This constraint prejudices the coefficients on the new variables to zero; that is, by leaving the variable out of the equation to begin with, thereby expressing interest in a "false" model, we have revealed that we think the variable will not significantly distort inferences on the other parameters. This is the case if the regression coefficients are negligible or if the added variables are orthogonal to the original set (or a combination).

This analysis can obviously be carried on to additional stages. At each stage the constraints on the new variables become more severe. Incidentally, the order in which variables are added to the equation influences the interpretation of the evidence. For example, two researchers may end up with the same set of explanatory variables. If these variables have been added to the equations in different orders, then the researchers have revealed different priors and must also make different interpretations of the data evidence.

CHAPTER **10**

SYSTEMATIC JUDGMENTAL ERRORS

10.1 "Explaining Your Results" as Access-Biased Memory 307
10.2 Biases in Personal Probabilities 315
10.3 Social Learning Processes 319

A theme of this book is that judgment is the critical input into the analysis of nonexperimental data. Systematic errors in the formation of judgment may lead to significant systematic errors in the interpretation of evidence. The elimination of systematic judgmental errors is thus highly desirable. As a first step in that direction, we may identify in this chapter what seem to be the more consequential systematic errors.

10.1 "Explaining Your Results" as Access-Biased Memory

QUESTION. What do the following quotations have in common?

The stock market reacted today to the favorable news released by the Commerce Department that our fourth-quarter trade surplus established a new record.

Casey Stengel demonstrated again his lack of managerial talent by replacing pitcher Whitey Ford by a wild Ryne Duren, who proceeded to walk in the winning run.

The negative estimated effect of the price of butter on the consumption of wheat is fully consistent with the fact that bread and butter are jointly consumed.

Answer. (a) All three statements are "explanations" of certain events. In the terminology of probability, where A and B are events, an explanation of an event

A is another event B such that $P(B)$ is large and $P(A|B)$ is significantly greater than $P(A)$.[1] (b) All four explanations were offered after the events occurred. More formally, the statement "$P(A|B)$ is close to one" is made only after both A and B are known to occur.

Suspicion is cast upon *ex post* explanations by the popular English phrases "20-20 hindsight" and "Monday-morning quarterbacking." These phrases allude to the tendency to think in retrospect that events were perfectly predictable, whereas, in fact, the events could not have been foreseen. When probabilities are computed incorrectly, inferential errors are certain to occur. In commenting on the "silly certainty of hindsight," Fischhoff (1975) observes that "if we believe, because of creeping determinism, that the past holds few surprises for us, then we fail to realize that we have anything to learn from it.... A surprise-free past is prologue to a surprise-full future."

It may be useful to construct explicit statistical models to illustrate the inferential errors. Actually, one has already been constructed. The data-instigated models discussed in the previous chapter tend to over-explain the data, and the inferences implied by such models should be discounted. In contraposition to the Fischhoff quotation, the potential error is excessive learning from the given experiment, not insufficient learning.

Another model of hindsight—"access-biased" memory—is now discussed. According to this model a great wealth of prior information is available to aid in the interpretation of current observations. Unfortunately, prior information is not accessed costlessly. As a result, the information that is remembered may bear only a fuzzy relationship to the actual past. The recalled information may even be a version of the past that is distorted to suit the present purposes. If the present purposes are explaining some event A, it is possible that events B, favorable to the outcome A, will be remembered; whereas events C, which are also stored in memory but are unfavorable to A, will be forgotten. It is then asserted that A was inevitable since B is true, and the fact that C is also true is ignored. This we call access-biased memory.

A formal model of access-biased memory consists of the following elements:[2]

a. An uncertain parameter p, $0 \leqslant p \leqslant 1$.
b. A beta prior distribution with parameters r and n.

[1]Good (1975) proposes as a measure of the degree of explanation the number $\log P(A|B) - \log P(A) + \gamma \log P(B)$ where $0 \leqslant \gamma \leqslant 1$. The first two terms measure the increase in probability, and the last measures the plausibility of B.

[2]This material is from Leamer (1975).

c. T independent random variables $X_1, \ldots, X_T$ each distributed binomially with parameters N and p.
d. An observation consisting of the random variable S_k where S_k is the number of the random variables X_i that assume the value k. $(S_0 + S_1 + \cdots + S_N = T)$.

The parameters of the prior, r and n, are considered to be the outcome of a current binomial experiment. The random variables $X_1, \ldots, X_T$ are considered to be the outcomes of T previous independent sampling experiments. These are assumed to be stored in memory that consists of $N+2$ counters: one that counts the total number of experiments and $N+1$ that count the number of experiments involving exactly k successes, $k = 0, 1, \ldots, N$. Memory is accessed by observing the total number of experiments T and also *one* of the other counters, S_k.

This is intended primarily to model the phenomenon that may be called "explaining your results." It is common practice first to estimate parameters from sample information and then to give reasons why these estimates are correct. The sample information is thereby supported by nonsample information in a way qualitatively in agreement with formal Bayesian analysis, and the "posterior distribution" that informally pools the quantitative sample information with the qualitative nonsample information is implicitly suggested to be more concentrated than the likelihood function alone.

Examples of this sort of thing are abundant; one suffices to illustrate the phenomenon. In a study of the effects of ability on earnings and schooling, Griliches (1974) obtains the "wrong" sign for a coefficient and reports, "This is an unexpected and strange result, which leads us to reexamine our model. Before we do that however, it is worth noting that the results may not be all that foolish (The human facility for rationalization is boundless)..." He then proceeds to explain why this wrong sign might, in fact, be correct.

The essential feature of this example is that Griliches' access to qualitative nonsample information is *selective*. We may assume, as does Griliches, that any result can be explained. But reasons why a coefficient can assume a "wrong" sign are constructed (remembered) only given the data signal of an estimated coefficient with the wrong sign. The signal of the wrong sign triggers not a general reaccess of memory but a selective one, aimed at remembering previous (qualitative) "experiments" that favor the "wrong sign."

The analogous situation in our formal model is that instead of remembering a random selection of the previous binomial experiments, only those experiments that led to exactly k successes are remembered. Each of these experiments necessarily favors the value $p = k/N$, just as reasons why a sign may be "wrong" necessarily support the hypothesis of wrong signs. But as we shall see, this is information that should be discounted, depending especially on T, the total number of previous experiments. It may turn out that the remembered previous experiments that apparently support the value $p = k/N$, in fact should be taken as evidence against this value.

Attention should also be drawn to the fact that Griliches' access to at least some of the qualitative nonsample information occurs *after* a current data set is observed. If this access to memory were random, we may justifiably ask why it did not occur *before* the data were observed. The only apparent answer is no answer at all—an accident of no importance occurred. The hypothesis of selective access to memory, on the other hand, has as a consequence the fact that it is optimal to search memory after the data are observed. We may, therefore, conclude that when we observe memory access after data analysis we have evidence in favor of the hypothesis of selective access relative to the hypothesis of random access.

The selective access model can itself be explained in terms of the library or computer storage models of memory used by psychologists (Howe, 1970). Experiments rather than sufficient statistics are stored in memory, because the computation of a sufficient statistic requires retrieval, computation, and storage operations each time a new experiment is observed. Storage is conceptually limited only by retrieval costs, and the decision to store experiments rather than sufficient statistics can be interpreted as the economic decision to save computation and some retrieval costs at the expense of greater retrieval costs when the information is actually to be used. In other words, there is a configuration of retrieval costs, computation costs, and information use patterns that discourages the computation and storage of sufficient statistics.

The other feature of our model—selective retrieval—has been the implicit study of numerous psychologists under the headings of secondary organization and associationism (Howe, 1970, p. 60). Events are clustered or categorized for later retrieval on the basis of some contentful categorization. For example, events contiguous in time may be accessed in blocks by retrieval questions such as "what happened after that?" It is here assumed that the relevant previous experiments were not conducted seriatim and that categorization by time would not aid retrieval. Instead, experiments are categorized by their meaning or implications for inference: those that favor one hypothesis are stored in one file, those that favor another are stored in another file.

To summarize, we are suggesting the two-part hypothesis: (1) events, not sufficient statistics or their qualitative equivalent, are stored in memory; (2) events are categorized in memory for later retrieval depending on their implications for inference. The intent of this section is not to test this

hypothesis but only to explore its implications. We have, however, already offered some qualitative evidence—the Griliches quotation. We may add to this the author's (and perhaps the readers') casual reading of numerous papers in econometrics and casual observation of economics seminars too numerous and too dull to recount in detail here.

The principal implication of this model is that the information attained from memory ought to be discounted. When it is not, and when a researcher retrieves from memory only those experiments similar in content to his current experiment, he greatly understates his uncertainty and places excessive faith in the validity of the current evidence.

This leads to the second question addressed in this section: given this form of memory, which of the categories is it optimal to access? Which is better: to retrieve experiments that tend to support the current experiment or ones that tend to cast doubt on it? In terms of the formal model, given some loss function for estimating p, which value of k is optimal? For example, a value of $k = N(r/n)$ necessarily accesses experiments that favor the same values of p as the current experiment, that is, $p = r/n$. It turns out that with squared error loss it is better in the sense of minimizing Bayes risk to access experiments that slightly contradict the current experiment.

Parenthetically, it may be observed that this model applies not only to personal memory but also to social information processes. The complaint that journals publish only extreme results is quite common. Newspapers publish "bad" news. My friends transmit to me only their most titillating stories. Casual observation thus suggests that information is categorized by its implications for inference and that social information transmission tends to emphasize the extremes. It is easy to construct a model of information transfer that makes this desirable, if it is understood by the participants that the information is selectively transmitted. Failure to understand the selective nature of the transmission results in erroneous inferences—the conclusion that the newspaper's man-bites-dog story accurately portrays the average relationship between men and dogs.

The structure of the formal model is the following:

1. An infinite population having a proportion, p, of its elements that possess a given attribute.
2. An "experience-free" prior for p in the beta family, which for convenience we take to be the diffuse prior

$$f_0(p) \propto p^{-1}(1-p)^{-1}. \qquad (10.1)$$

3. A binomial sample consisting of r successes in n independent trials with likelihood function

$$f(r,n|p) \propto p^r(1-p)^{n-r}. \qquad (10.2)$$

4. A set of T experiences stored in memory, each of which consists of a binomial sample of size N with r_t successes, $t=1,\ldots,T$.
5. A rule for accessing memory, which we take to be the following: select an integer k and memory reports the number of experiences S_k for which $r_t=k$.

It is assumed for simplicity that the cost function for accessing memory allows for one choice of k essentially for free but disallows any further interrogation of memory.[3]

In principle, memory could be interrogated before the sample values (r,n) are observed. The strategy of letting k be a function of r and n includes as a special case k independent of r and n. It cannot, therefore, increase expected loss, and does, in fact, decrease it, except in unusual circumstances. We therefore search memory after observing r and n. Combining the experience-free prior with the likelihood function to form a prememory distribution, we obtain

$$f(p|r,n)\propto p^{r-1}(1-p)^{n-r-1}\propto f_\beta(p|r,n) \tag{10.3}$$

where f_β indicates a beta distribution

$$f_\beta(p|r,n)=B^{-1}(r,n-r)p^{r-1}(1-p)^{n-r-1}, \qquad 0\leqslant p\leqslant 1, 0<r<n$$

where

$$B(r,n-r)=\frac{(r-1)!(n-r-1)!}{(n-1)!}.$$

A postmemory distribution is formed by multiplying the prememory distribution (10.3) times the likelihood function of p depending on s, the memory output. This function can be derived by some straightforward probability manipulations. The number of successes in each of the experiences is assumed to be binomially distributed with parameter p

$$\pi(k,p)=P(r_t=k|p)=\binom{N}{k}p^k(1-p)^{N-k}.$$

Conditional on p each experience is independent and contains k successes with probability $\pi(k,p)$. The number of experiences with k successes is, therefore, binomially distributed with parameter π

$$P(S_k=s|p,k)=\binom{T}{s}\pi^s(1-\pi)^{T-s} \tag{10.4}$$

$$\propto\left[p^{ks}(1-p)^{(N-k)s}\right]\left[1-\binom{N}{k}p^k(1-p)^{N-k}\right]^{T-s}$$

[3] An alternative reasonable assumption is that extreme events are the most memorable, i.e., accessed at least cost.

which, given s, is the likelihood function of p. Note that the first term in the brackets in this expression is the likelihood function that we would usually use to characterize the information in our s-remembered experiences, each of which consists of a sample of size N with k successes. The second term is a discount factor that should apply to this information because of the way it was accessed.

If we let the discount factor be

$$d=d(N,k,T,s,p)=\left[1-\binom{N}{k}p^k(1-p)^{N-k}\right]^{T-s},$$

the postmemory distribution of p formed by multiplying the prememory distribution (10.3) times the memory likelihood function (10.4) is

$$f(p|r,n,k,T,s,N)\propto f_\beta(p|r+sk,n+Ns)\cdot d \tag{10.5}$$

which is the usual beta distribution times the discount factor d. The discount factor can be written as

$$d=\left[1-\binom{N}{k}B(k+1,N-k+1)f_\beta(p|k+1,N+2)\right]^{T-s}$$

$$=\left[1-(N+1)f_\beta(p|k+1,N+2)\right]^{T-s}$$

which assumes a minimum value at $p=k/N$, thereby discounting the memory evidence that would otherwise necessarily favor the value of $p=k/N$.

To illustrate the effect of the discount factor, let us consider the case when $k=0$, that is, when memory is accessed by the question "are there any previous experiences that involved no successes?" The factor then becomes

$$d=\left[1-(1-p)^N\right]^{T-s}$$

which assumes a minimum of zero at $p=0$ and a maximum of one at $p=1$. This has the effect of pushing the posterior distribution away from $p=0$, depending positively on the number of forgotten experiences, $T-s$. Of course, any experiences that are remembered necessarily favor $p=0$. The net effect on the distribution thus depends on both the remembered experiences and the forgotten experiences. There is an informal lesson to be drawn from this. Old men with many experiences must tell more stories in support of their theories if they expect to generate the same amount of believability as a young man. Or to put it differently, the wisdom of age is greatly exaggerated if memory failures are ignored.

We now turn to the problem of optimal memory interrogation. Consider a researcher who has current experimental support for some proposition x against an alternative y. If he searches memory for further evidence in favor of x, he expects to find it but "hopes" he doesn't, since the absence of previous similar results will informatively cast doubt on the proposition while the presence of previous similar results is unsurprising and relatively uninformative. Similarly, if he searches memory for evidence in favor of the alternative y, he expects not to find it but "hopes" that he does. In both cases he anticipates obtaining relatively uninformative information. This symmetry makes ambiguous the decision whether to search memory for experiments in favor of x or in favor of y, an ambiguity that can be resolved only in the context of specific problems.

For our problem, optimal choice of k, the memory-accessed value, is not obvious but depends on a conceptually straightforward preposterior analysis. Let us take the variance of p as the measure of uncertainty and seek to find the value of k that minimizes the expected posterior variance.

The expected posterior moments can be written as

$$E(Ep|s) = Ep = \frac{r}{n}$$

$$E(V(p|s)) = E(E(p^2|s)) - E\left\{[E(p|s)]^2\right\}$$

$$= Ep^2 - E\left\{[E(p|s)]^2\right\}$$

$$= \frac{r(r+1)}{n(n+1)} - E\left\{[E(p|s)]^2\right\}.$$

That is, the expected mean is just the prior mean (r/n), and the expected variance depends on k only through the factor $E\{[E(p|s)]^2\}$. These can be computed by appropriately defining the joint distributions of p and s, as in Leamer (1975).

The expected posterior variance as a function of k is illustrated for one case in Figure 10.1. It is assumed that the current sample size n is equal to 10, that this sample involves five successes, that there are $T=5$ previous experiments stored in memory, and that each of these five previous experiments involve $N=10$ trials. Note that although $N=10$, the least desirable ($EV(ps)$ maximum) memory accessed value is $k=5$. Past experiences that involve k successes favor the value $p=k/N$. Thus the least desirable way to interrogate memory is to ask if there are any previous experiences that favor the value $p=k/N=.5$, the very value that is most favored by the current sample, $r/n=.5$. It is almost as undesirable to search memory for extreme experiences, $k=0$ or $k=10$. The best thing to do is to ask for experiences that slightly contradict the current sample in the sense of favoring the values $p=.8$ or $p=.2$.

This resolves the ambiguity referred to previously. Since the value r/n is currently the most favored value of p, the researcher regards it to be highly

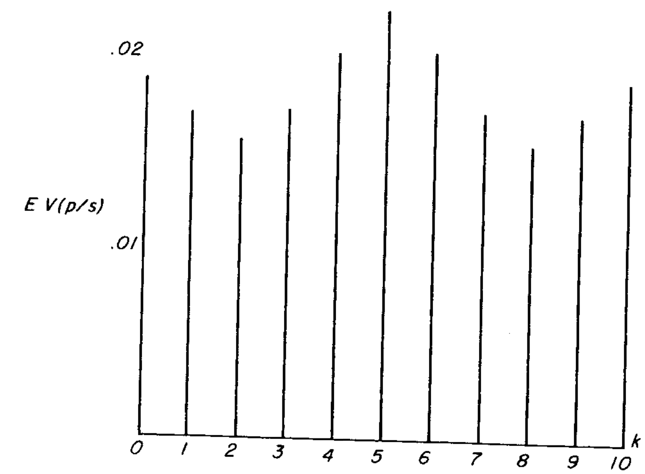

Fig. 10.1 Expected posterior variance; $n=10, r=5, N=10, T=5$.

likely that there will be previous experiments that favor this same value. Counterbalancing this influence is the fact that these experiments when remembered are not particularly informative. Slightly contradictory experiments, however, are much more informative and also are reasonably likely to be in memory. Greatly contradictory experiments, although especially informative, are highly unlikely to be in memory and therefore should not be searched out.

With the *caveat* that our inferences do not necessarily generalize to other situations, especially those free of memory failures, we conclude with the following homilies:

a. A young man with one good story may merit your attention more than an old man with several.
b. Spend your time thinking of reasons why you are slightly wrong. Waste not your time thinking why you are greatly wrong—you are likely to go home with an empty basket. Least of all think of reasons why you are right—though you are likely to fill your basket, the fruits you bring home will contain little knowledge.

10.2 Biases in Personal Probabilities

We have discussed in the previous section an error that is potentially made when prior probabilities are formed after the data are observed. This is the error of failing appropriately to discount prior information that otherwise necessarily is supportive of the current sample. There are other interesting "errors" that are often made and that have important implications for our study of inference with nonexperimental data. Many such errors are discussed by Tversky and Kahneman (1974) and by Slovic (1972), and Hogarth (1975). In this section we briefly discuss some of the more interesting of these phenomena.

Memory Failures

According to the model of the previous section, events are memorable because they have implications for inference, or to put it differently, storage is arranged in anticipation of a well-defined inferential problem. As an illustration of a similar memory error without an anticipated inferential problem, glance over the following list of names without reading further the textual materials:

Gerald Ford
Betty Jackson
Maria Muldaur
Ralph Nader
Ronald Reagan
Barbara Wilson
William Shakespeare
Katherine Burgoyne
Al Jolson
Deborah Hirsch

Now cover up the list of names and answer the question, "Are there more men or more women on this list?"

Tversky and Kahneman (1974) report that in such experiments there is a tendency to overstate the proportion of men, when the listed men are relatively more famous (as above), and to overstate the proportion of women, when the listed women are relatively more famous. The important point of the example is that the more memorable events are most likely used to form conscious opinions. The resulting opinions may significantly distort the past when the ease of memory is related in some way to the evidential content of the event. Somewhat the same point has been made in our discussion of access-biased memory: the current instance of an event makes more memorable similar past instances.

OPTIMISM/PESSIMISM

Another error in forming opinions is the confusion of a utility function with prior information. Optimists are people who think desirable events will likely happen; pessimists think they won't. Although it does not

involve a logical contradiction to condition opinions on utilities, most thoughtful observers would argue against such a practice. The existence of optimists and pessimists has a firm foundation in folklore and has been more scientifically established by Slovic (1966).

The confusion between prior information and utility functions leads, in the language of this book, to a confusion between interpretive searches, which require prior information, and simplification searches, which require utility functions. As an example, consider the (constant elasticity of substitution) function

$$y(x_1,x_2)=\alpha\left[x_1^{-\beta}+(\gamma x_2)^{-\beta}\right]^{-\delta/\beta}. \tag{10.6}$$

In the limit, as β goes to zero, this function goes to

$$y^*(x_1,x_2)=\alpha x_1^{\gamma_1}x_2^{\gamma_2}, \tag{10.7}$$

which, conveniently, is linear in the logarithms. A common practice in economics is to test the restriction $\beta=0$, hoping that the restriction is accepted. We may ask if this is an interpretive search or a simplification search. My own opinion is that the restriction derives originally from a utility function—it is ever so much neater to work with the second function than the first. With the passage of time, however, what was first a simplification search has now become—inappropriately, I think—an interpretive search. The many studies that failed to reject the restriction $\beta=0$ generate a feeling among economists that the restriction is an hypothesis that is favored by prior information. Although these studies do, indeed, contain information about the parameter β, the fact that many of them fail to reject $\beta=0$ is only remotely connected to the accumulated probability in a region around $\beta=0$. It is important, as we have argued, to distinguish the output of a simplification search from the interpretation of the data.

EGOCENTRISM

In formulating opinions, about other human beings especially, you are necessarily forced to use yourself and your experiences as a norm. In so doing there is a clear tendency to regard yourself to be overly representative, that is, to fail to realize appropriately the diversity of humanity. An experiment I have done in my own classes is to have students estimate the weight of a relatively heavy person and a relatively light person. Not only do light people consistently underestimate the weight of the heavy subject, moreover, they fail to allow for their increased uncertainty in their choice of confidence intervals. Not only do light people have no concept of what it means to be heavy, but also they fail to realize that they have no such knowledge.

The importance of this phenomenon in social research could be overstated as follows: judgment is critically important in the analysis of nonexperimental data, and the diversity of judgments of academics is limited by the sameness of their lives. Thus the interpretation of evidence by academics is a class phenomenon: relative unanimity within the class and potentially sharp disagreements with other classes.

LAW OF SMALL NUMBERS

Tversky and Kahneman (1971a) in a study of psychologists observe that even people with some formal training in statistical inference have a deep belief in the "law of small numbers," according to which a sample is necessarily representative of the population as a whole. As an example of the law, the probability of a tail after five heads in a row is quite a bit larger than one-half because a tail is "due" or, if you like, because the sequence of six coin flips is strongly thought to be representative of the universe of coin flips consisting of 50% heads and 50% tails. A more important example is the unwarranted belief in the estimates generated by a relatively small sample, as demonstrated, first, by the researcher's willingness to go to great lengths to "explain" the sample result and, second, by the researcher's surprise at the extent to which estimates can change as evidence accumulates.

OVERCONFIDENCE

Heretofore we have been concerned largely with the location of prior distributions. A few things may also be said about the dispersions of these distributions. One phenomenon is overconfidence. In a study of students at the Harvard Business School, Alpert and Raiffa (1969) found that 426 out of 1000 98%-confidence intervals failed to capture the true value of the item being estimated. You might have expected approximately 20 misses, and the fact that as many as 426 intervals were "wrong" has to be considered strong evidence of overconfidence. However, the extent to which this can be considered to be a general phenomenon is subject to considerable doubt. In particular, it may be very sensitive to the method of eliciting the interval (Hogarth, 1975, provides many references).

CONSERVATISM

The counterpart of overconfidence with respect to prior information is undervaluation of current information (Edwards, 1968). To give a simple example, a coin is flipped to decide from which of two urns to select a ball. The first urn contains 80% red balls and the second contains 20%. If a red ball is selected, the conditional probability that it came from the first urn is "objectively" .8. Subjects consistently underestimate this probability, that

is, they fail to revise their prior (.5) as much as would be suggested by Bayes' rule. There is a considerable literature in psychology, referenced by Hogarth (1975), that explores the circumstances in which this phenomenon occurs.

TRIVIALITY

Another error that is made is the confusion of the mass of information with the value of the information. A researcher sometimes attempts to convince his reader of the validity of his study by inundating the reader with great masses of information. Although it is obvious that the "informational weight" of a study is not equal to its physical weight, there is a tendency to unduly correlate the two.

For example, Oskamp (1965) provided psychological clinicians with written descriptions of the personalities of some subjects and then asked the clinicians to answer 25 questions about the subjects. He found that although the ability of the clinicians to answer the questions correctly was little affected by increases in the length of the written description, the confidence ascribed to the answers rose dramatically.

METHODOLOGICAL PATTERNING

Without comment, I would like to describe (with some literary license) Skinner's (1948) experiment of randomly induced behavior. Hungry birds fed at *random* intervals are observed to adopt the peculiar behavior of odd head movements, hopping from side to side, and the like. The apparent explanation is that on receipt of the seed, the bird hypothesizes that the seed is a reward for the most recently antecedent trick. If the bird happened to have twitched his head just before the seed arrived, the bird naturally tries twitching his head again. The increased frequency of head twitching makes more likely the event that a twitch will precede a seed, and eventually the bird is twitching frequently with seeds always following twitching. The belief in this relationship is so strong that even after the seed stops arriving altogether the bird may twitch as many as 10,000 times. Slovic (1972) sees parallels between pigeons and stockbrokers. I certainly would not want to draw parallels between pigeons and social scientists.

10.3 Social Learning Processes

The opinions and decisions of an individual are only partly dependent on his own observations. The advice and comments of other individuals is another highly important source of information and decision rules. It is embarrassing, on almost the last page of a book dealing with learning, to make the observation that the personal learning heretofore discussed constitutes only a small part of the learning process. The space allocated to social-learning processes reflects my knowledge, not my assessment of their importance.

IMPROPER POOLING OF INFORMATION ACROSS INDIVIDUALS

Two or more individuals who have opinions based on the same information should not change their opinions when confronted by the (necessarily coincident?) opinions of others. In practice, people are unduly affected by the opinions of others, to the point where unanimity is sometimes confused with certainty.

An experiment, first performed by Asch (1952), demonstrates the remarkable tendency of conformity of opinion in groups. Seven individuals are asked to identify which of three lines is the same length as some standard line. The first six individuals are, in fact, collaborators of the experimenter, and they deliberately and uniformly select (aloud) a line that is not the right one. The subject, who is unaware of the collusion, is then forced to choose between either giving what he feels is the right answer or conforming with the group. Faced with this decision, in 33.2% of the cases the subjects conformed to the group and gave the incorrect response. This contrasts with subjects not exposed to group pressure, who answered incorrectly only 7.4% of the time.

It may be argued, of course, that this is not improper pooling of information. Subjects may legitimately be influenced by the information revealed by others. What the experiment fails to distinguish is conformity due to information transfer from conformity merely to please the other members of the group. If a correct answer were sufficiently well rewarded, would the subject continue to conform? Yes, if he gets information from the group. No, if he is merely trying to please his colleagues.

ADVOCACY ABILITY

Under the heading of triviality we observed the confusion of the physical mass of information with its real content. Other features of the reporting style are also likely to greatly influence our willingness to believe results and arguments. Especially when the data evidence is relatively weak, it is possible to end up with a uniform professional opinion about some empirical issue, merely because one scholar has unusually fine advocacy abilities. We may even think that our judgment in his favor is heavily influenced by observed phenomena.

In a study designed to score advocates to eliminate ability bias, Warner (1975) had two teams of advocates each write pro and con briefs about a proposed expressway in Toronto. One pro and one con case was mailed to 1360 randomly selected electors. Warner found that the proportion of

electors who favored the expressway after having received the briefs varied from .53 to .85, depending on which pro and which con case they received. The empirical facts are the same in all cases. The variability of the proportion in favor of the expressway is partly attributable to sampling error but mostly to advocacy ability.

CONSENSUS PRECEDES CERTAINTY

Divergent opinions when confronted with observations tend to converge to each other as well as to the "truth." Dickey and Fischer (1975) observe that for a broad class of sampling contexts, the posterior mean converges to the true value at the rate of $T^{-1/2}$, but the dispersion of posterior means across individuals goes to zero at the rate T^{-1} (T is the sample size).

OTHER

A consensus model is given by DeGroot (1974). See also Pruitt (1971) and Stone (1961).