COGNITIVE PSYCHOLOGY 5, 207–232 (1973)

Availability: A Heuristic for Judging Frequency and Probability[1,2]

AMOS TVERSKY AND DANIEL KAHNEMAN
The Hebrew University of Jerusalem and the Oregon Research Institute

This paper explores a judgmental heuristic in which a person evaluates the frequency of classes or the probability of events by availability, i.e., by the ease with which relevant instances come to mind. In general, availability is correlated with ecological frequency, but it is also affected by other factors. Consequently, the reliance on the availability heuristic leads to systematic biases. Such biases are demonstrated in the judged frequency of classes of words, of combinatorial outcomes, and of repeated events. The phenomenon of illusory correlation is explained as an availability bias. The effects of the availability of incidents and scenarios on subjective probability are discussed.

I. INTRODUCTION

Much recent research has been concerned with the validity and consistency of frequency and probability judgments. Little is known, however, about the psychological mechanisms by which people evaluate the frequency of classes or the likelihood of events.

We propose that when faced with the difficult task of judging probability or frequency, people employ a limited number of heuristics which reduce these judgments to simpler ones. Elsewhere we have analyzed in detail one such heuristic—representativeness. By this heuristic, an event is judged probable to the extent that it represents the essential features of its parent population or generating process. Evidence for representativeness was obtained in several studies. For example, a large majority of naive respondents believe that the sequence of coin tosses HTTHTH is more probable than either HHHHTH or HHHTTT, al-

[1] Address: Department of Psychology, Hebrew University of Jerusalem, Jerusalem, Israel.

[2] This work was supported by NSF grant GB-6782, by a grant from the Central Research Fund of the Hebrew University, by grant MH 12972 from the National Institute of Mental Health and Grants 5 S01 RR 05612-03 and RR 05612-04 from the National Institute of Health to the Oregon Research Institute.

We thank Maya Bar-Hillel, Ruth Beyth, Sundra Gregory, and Richard Kleinknecht for their help in the collection of the data, and Douglas Hintzman and Paul Slovic for their helpful comments on an earlier draft.

though all three sequences, of course, are equally likely. The sequence which is judged most probable best represents both the population proportion (½) and the randomness of the process (Kahneman & Tversky, 1972). Similarly, both naive and sophisticated subjects evaluate the likelihood that an individual will engage in an occupation by the degree to which he appears representative of the stereotype of that occupation (Kahneman & Tversky, 1973). Major biases of representativeness have also been found in the judgments of experienced psychologists concerning the statistics of research (Tversky & Kahneman, 1971).

When judging the probability of an event by representativeness, one compares the essential features of the event to those of the structure from which it originates. In this manner, one estimates probability by assessing similarity or connotative distance. Alternatively, one may estimate probability by assessing availability, or associative distance. Life-long experience has taught us that instances of large classes are recalled better and faster than instances of less frequent classes, that likely occurrences are easier to imagine than unlikely ones, and that associative connections are strengthened when two events frequently co-occur. Thus, a person could estimate the numerosity of a class, the likelihood of an event, or the frequency of co-occurrences by assessing the ease with which the relevant mental operation of retrieval, construction, or association can be carried out.

For example, one may assess the divorce rate in a given community by recalling divorces among one's acquaintances; one may evaluate the probability that a politician will lose an election by considering various ways in which he may lose support; and one may estimate the probability that a violent person will "see" beasts of prey in a Rorschach card by assessing the strength of association between violence and beasts of prey. In all these cases, the estimation of the frequency of a class or the probability of an event is mediated by an assessment of availability.[3] A person is said to employ the availability heuristic whenever he estimates frequency or probability by the ease with which instances or associations could be brought to mind. To assess availability it is not necessary to perform the actual operations of retrieval or construction. It suffices to assess the ease with which these operations could be performed, much as the difficulty of a puzzle or mathematical problem can be assessed without considering specific solutions.

That associative bonds are strengthened by repetition is perhaps the oldest law of memory known to man. The availability heuristic exploits

[3] The present use of the term "availability" does not coincide with some usages of this term in the verbal learning literature (see, e.g., Horowitz, Norman, & Day, 1966; Tulving & Pearlstone, 1966).

the inverse form of this law, that is, it uses strength of association as a basis for the judgment of frequency. In this theory, availability is a mediating variable, rather than a dependent variable as is typically the case in the study of memory. Availability is an ecologically valid clue for the judgment of frequency because, in general, frequent events are easier to recall or imagine than infrequent ones. However, availability is also affected by various factors which are unrelated to actual frequency. If the availability heuristic is applied, then such factors will affect the perceived frequency of classes and the subjective probability of events. Consequently, the use of the availability heuristic leads to systematic biases.

This paper explores the availability heuristic in a series of ten studies.[4] We first demonstrate that people can assess availability with reasonable speed and accuracy (Section II). Next, we show that the judged frequency of classes is biased by the availability of their instances for construction (Section III), and retrieval (Section IV). The experimental studies of this paper are concerned with judgments of frequencies, or of probabilities that can be readily reduced to relative frequencies. The effects of availability on the judged probabilities of essentially unique events (which cannot be reduced to relative frequencies) are discussed in the fifth and final section.

II. ASSESSMENTS OF AVAILABILITY

Study 1: Construction

The subjects ($N = 42$) were presented with a series of word-construction problems. Each problem consisted of a 3×3 matrix containing nine letters from which words of three letters or more were to be constructed. In the training phase of the study, six problems were presented to all subjects. For each problem, they were given 7 sec to estimate the number of words which they believed they could produce in 2 min. Following each estimate, they were given two minutes to write down (on numbered lines) as many words as they could construct from the letters in the matrix. Data from the training phase were discarded. In the test phase, the construction and estimation tasks were separated. Each subject estimated for eight problems the number of words which he believed he

[4] Approximately 1500 subjects participated in these studies. Unless otherwise specified, the studies were conducted in groups of 20–40 subjects. Subjects in Studies 1, 2, 3, 9 and 10 were recruited by advertisements in the student newspaper at the University of Oregon. Subjects in Study 8 were similarly recruited at Stanford University. Subjects in Studies 5, 6 and 7 were students in the 10th and 11 grades of several college-preparatory high schools in Israel.

could produce in 2 min. For eight other problems, he constructed words without prior estimation. Estimation and construction problems were alternated. Two parallel booklets were used, so that for each problem half the subjects estimated and half the subjects constructed words.

Results. The mean number of words produced varied from 1.3 (for XUZONLCJM) to 22.4 (for TAPCERHOB), with a grand mean of 11.9. The mean number estimated varied from 4.9 to 16.0 (for the same two problems), with a grand mean of 10.3. The product-moment correlation between estimation and production, over the sixteen problems, was 0.96.

Study 2: Retrieval

The design and procedure were identical to Study 1, except for the nature of the task. Here, each problem consisted of a category, e.g., *flowers* or *Russian novelists*, whose instances were to be recalled. The subjects ($N = 28$) were given 7 sec to estimate the number of instances they could retrieve in 2 min, or two minutes to actually retrieve the instances. As in Study 1, the production and estimation tasks were combined in the training phase and alternated in the test phase.

Results. The mean number of instances produced varied from 4.1 (city names beginning with F) to 23.7 (four-legged animals), with a grand mean of 11.7. The mean number estimated varied from 6.7 to 18.7 (for the same two categories), with a grand mean of 10.8. The product–moment correlation between production and estimation over the 16 categories was 0.93.

Discussion

In the above studies, the availability of instances could be measured by the total number of instances retrieved or constructed in any given problem.[5] The studies show that people can assess availability quickly and accurately. How are such assessments carried out? One plausible mechanism is suggested by the work of Bousfield and Sedgewick (1944), who showed that cumulative retrieval of instances is a negatively accelerated exponential function of time. The subject could, therefore, use the number of instances retrieved in a short period to estimate the number of instances that could be retrieved in a much longer period of time. Alternatively, the subject may assess availability without explicitly re-

[5] Word-construction problems can also be viewed as retrieval problems because the response-words are stored in memory. In the present paper we speak of retrieval when the subject recalls instances from a natural category, as in Studies 2 and 8. We speak of construction when the subject generates exemplars according to a specified rule, as in Studies 1 and 4.

trieving or constructing any instances at all. Hart (1967), for example, has shown that people can accurately assess their ability to recognize items that they cannot recall in a test of paired-associate memory.

III. AVAILABILITY FOR CONSTRUCTION

We turn now to a series of problems in which the subject is given a rule for the construction of instances and is asked to estimate their total (or relative) frequency. In these problems—as in most estimation problems—the subject cannot construct and enumerate all instances. Instead, we propose, he attempts to construct some instances and judges overall frequency by availability, that is, by an assessment of the ease with which instances could be brought to mind. As a consequence, classes whose instances are easy to construct or imagine will be perceived as more frequent than classes of the same size whose instances are less available. This prediction is tested in the judgment of word frequency, and in the estimation of several combinatorial expressions.

Study 3: Judgment of Word Frequency

Suppose you sample a word at random from an English text. Is it more likely that the word starts with a *K*, or that *K* is its third letter? According to our thesis, people answer such a question by comparing the availability of the two categories, i.e., by assessing the ease with which instances of the two categories come to mind. It is certainly easier to think of words that start with a *K* than of words where *K* is in the third position. If the judgment of frequency is mediated by assessed availability, then words that start with *K* should be judged more frequent. In fact, a typical text contains twice as many words in which *K* is in the third position than words that start with *K*.

According to the extensive word-count of Mayzner and Tresselt (1965), there are altogether eight consonants that appear more frequently in the third than in the first position. Of these, two consonants (*X* and *Z*) are relatively rare, and another (*D*) is more frequent in the third position only in three-letter words. The remaining five consonants (*K*,*L*,*N*,*R*,*V*) were selected for investigation.

The subjects were given the following instructions:

> "The frequency of appearance of letters in the English language was studied. A typical text was selected, and the relative frequency with which various letters of the alphabet appeared in the first and third positions in words was recorded. Words of less than three letters were excluded from the count.

> You will be given several letters of the alphabet, and you will be asked to judge whether these letters appear more often in the first or in the third position, and to estimate the ratio of the frequency with which they appear in these positions."

A typical problem read as follows:

"Consider the letter *R*.

Is R more likely to appear in
__ the first position?
__ the third position?
(check one)

My estimate for the ratio of these two values is ______: 1."

Subjects were instructed to estimate the ratio of the larger to the smaller class. For half the subjects, the ordering of the two positions in the question was reversed. In addition, three different orderings of the five letters were employed.

Results. Among the 152 subjects, 105 judged the first position to be more likely for a majority of the letters, and 47 judged the third position to be more likely for a majority of the letters. The bias favoring the first position is highly significant ($p < .001$, by sign test). Moreover, each of the five letters was judged by a majority of subjects to be more frequent in the first than in the third position. The median estimated ratio was 2:1 for each of the five letters. These results were obtained despite the fact that all letters were more frequent in the third position.

In other studies we found the same bias favoring the first position in a within-subject design where each subject judged a single letter, and in a between-subjects design, where the frequencies of letters in the first and in the third positions were evaluated by different subjects. We also observed that the introduction of payoffs for accuracy in the within-subject design had no effect whatsoever. Since the same general pattern of results was obtained in all these methods, only the findings obtained by the simplest procedure are reported here.

A similar result was reported by Phillips (1966) in a study of Bayesian inference. Six editors of a student publication estimated the probabilities that various bigrams, sampled from their own writings, were drawn from the beginning or from the end of words. An incidental effect observed in that study was that all the editors shared a common bias to favor the hypothesis that the bigrams had been drawn from the beginning of words. For example, the editors erroneously judged words beginning with *re* to be more frequent than words ending with *re*. The former, of course, are more available than the latter.

Study 4: Permutations

"Consider the two structures, A and B, which are displayed below.

(A)

```
x x x x x x x x
x x x x x x x x
x x x x x x x x
```

(B)

```
x x
x x
x x
x x
x x
x x
x x
x x
x x
```

A path in a structure is a line that connects an element in the top row to an element in the bottom row, and passes through one and only one element in each row.

In which of the two structures are there more paths?
How many paths do you think there are in each structure?"

Most readers will probably share with us the immediate impression that there are more paths in A than in B. Our subjects agreed: 46 of 54 respondents saw more paths in A than in B ($p < .001$, by sign test). The median estimates were 40 paths in A and 18 in B. In fact, the number of paths is the same in both structures, for $8^3 = 2^9 = 512$.

Why do people see more paths in A than in B? We suggest that this result reflects the differential availability of paths in the two structures. There are several factors that make the paths in A more available than those in B. First, the most immediately available paths are the columns of the structures. There are 8 columns in A and only 2 in B. Second, among the paths that cross columns, those of A are generally more distinctive and less confusable than those in B. Two paths in A share, on the average, about $\frac{1}{8}$ of their elements, whereas two paths in B share, on the average, half of their elements. Finally, the paths in A are shorter and hence easier to visualize than those in B.

Study 5: Combinations

Consider a group of ten people who have to form committees of r members, where r is some number between 2 and 8. How many different committees of r members can they form? The correct answer to this problem is given by the binomial coefficient $\binom{10}{r}$ which reaches a maximum of 252 for $r = 5$. Clearly, the number of committees of r members equals the number of committees of $10 - r$ members because any elected group of, say, two members defines a unique nonelected group of eight members.

According to our analysis of intuitive estimation, however, committees of two members are more available than committees of eight. First, the simplest scheme for constructing committees is a partition of the group into disjoint subsets. Thus, one readily sees that there are as many as five disjoint committees of two members, but not even two disjoint committees of eight. Second, committees of eight members are much less distinct, because of their overlapping membership; any two committees of eight share at least six members. This analysis suggests that small committees are more available than large committees. By the availability hypothesis, therefore, the small committees should appear more numerous.

Four groups of subjects (total $N = 118$) estimated the number of possible committees of r members that can be formed from a set of ten people. The different groups, respectively, evaluated the following values of r: 2 and 6; 3 and 8; 4 and 7; 5.

Median estimates of the number of committees are shown in Fig. 1, with the correct values. As predicted, the judged numerosity of committees decreases with their size.

The following alternative formulation of the same problem was devised in order to test the generality of the findings:

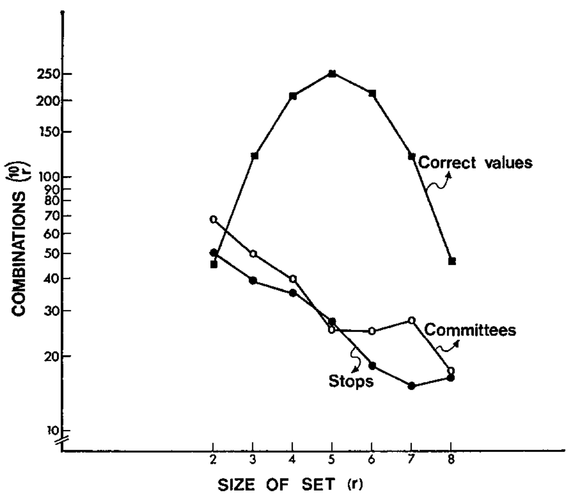

FIG. 1. Correct values and median judgments (on a logarithmic scale) for the Committees problem and for the Stops problem.

"In the drawing below, there are ten stations along a route between Start and Finish. Consider a bus that travels, stopping at exactly r stations along this route.

What is the number of different patterns of r stops that the bus can make?"

The number of different patterns of r stops is again given by $\binom{10}{r}$. Here too, of course, the number of patterns of two stops is the same as the number of patterns of eight stops, because for any pattern of stops there is a unique complementary pattern of non-stops. Yet, it appears as though one has more degrees of freedom in constructing patterns of two stops where "one has many stations to choose from" than in constructing patterns of eight stops where "one must stop at almost every station." Our previous analysis suggests that the former patterns are more available: more such patterns are seen at first glance, they are more distinctive, and they are easier to visualize.

Four new groups of subjects (total $N = 178$) answered this question, for $r = 2, \ldots, 8$, following the same design as above. Median estimates of the number of stops are shown in Fig. 1. As in the committee problem, the apparent number of combinations generally decreases with r, in accordance with the prediction from the availability hypothesis, and in marked contrast to the correct values. Further, the estimates of the number of combinations are very similar in the two problems. As in other combinatorial problems, there is marked underestimation of all correct values, with a single exception in the most available case, where $r = 2$.

The underestimation observed in Experiments 4 and 5 occurs, we suggest, because people estimate combinatorial values by extrapolating from an initial impression. What a person sees at a glance or in a few steps of computation gives him an inadequate idea of the explosive rate of growth of many combinatorial expressions. In such situations, extrapolating from an initial impression leads to pronounced underestimation. This is the case whether the basis for extrapolation is the initial availability of instances, as in the preceding two studies, or the output of an initial computation, as in the following study.

Study 6: Extrapolation

We asked subjects to estimate, within 5 sec, a numerical expression that was written on the blackboard. One group of subjects ($N = 87$) estimated the product $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$, while another group ($N = 114$) estimated the product $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times$

8. The median estimate for the descending sequence was 2,250. The median estimate for the ascending sequence was 512. The difference between the estimates is highly significant ($p < .001$, by median test). Both estimates fall very short of the correct answer, which is 40,320.

Both the underestimation of the correct value and the difference between the two estimates support the hypothesis that people estimate 8! by extrapolating from a partial computation. The factorial, like other combinatorial expressions, is characterized by an ever-increasing rate of growth. Consequently, a person who extrapolates from a partial computation will grossly underestimate factorials. Because the results of the first few steps of multiplication (performed from left to right) are larger in the descending sequence than in the ascending sequence, the former expression is judged larger than the latter. The evaluation of the descending sequence may proceed as follows: "8 times 7 is 56 times 6 is already above 300, so we are dealing with a reasonably large number." In evaluating the ascending sequence, on the other hand, one may reason: "1 times 2 is 2 times 3 is 6 times 4 is 24, and this expression is clearly not going very far. . . ."

Study 7: Binomial—Availability vs Representativeness

The final study of this section explores the role of availability in the evaluation of binomial distributions and illustrates how the formulation of a problem controls the choice of the heuristic that people adopt in intuitive estimation.

The subjects ($N = 73$) were presented with these instructions:

"Consider the following diagram:

X	X	O	X	X	X
X	X	X	X	O	X
X	O	X	X	X	X
X	X	X	O	X	X
X	X	X	X	X	O
O	X	X	X	X	X

A path in this diagram is any descending line which starts at the top row, ends at the bottom row, and passes through exactly one symbol (X or O) in each row.

What do you think is the percentage of paths which contain

6—X and no—O _____%
5—X and 1—O _____%
.
.
.
No—X and 6—O _____%

Note that these include all possible path-types and hence your estimates should add to 100%."

The actual distribution of path-type is binomial with $p = 5/6$ and $n = 6$. People, of course, can neither intuit the correct answers nor enumerate all relevant instances. Instead, we propose, they glance at the diagram and estimate the relative frequency of each path-type by the ease with which individual paths of this type could be constructed. Since, at every stage in the construction of a path (i.e., in each row of the diagram) there are many more X's than O's, it is easier to construct paths consisting of six X's than paths consisting of, say, five X's and one O, although the latter are, in fact, more numerous. Accordingly, we predicted that subjects would erroneously judge paths of 6 X's and no O to be the most numerous.

Median estimates of the relative frequency of all path-types are presented in Fig. 2a, along with the correct binomial values. The results confirm the hypothesis. Of the 73 subjects, 54 erroneously judged that there are more paths consisting of six X's and no O than paths consisting

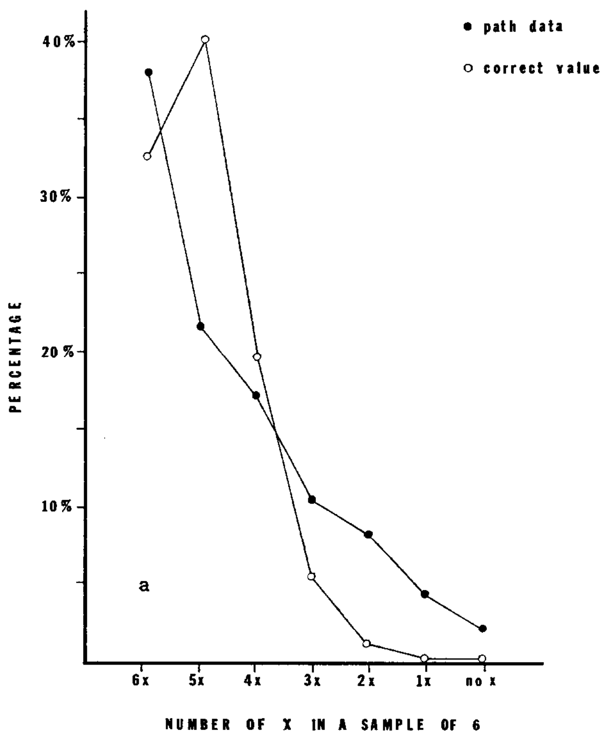

FIG. 2a. Correct values and median judgments: Path problem.

of five X's and one O, and only 13 regarded the latter as more numerous than the former ($p < .001$, by sign test). The monotonicity of the subjective distribution of path-types is apparently a general phenomenon. We have obtained the same result with different values of p (4/5 and 5/6) and n (5, 6 and 10), and different representations of the population proportions (e.g., four X's and one O or eight X's and two O's in each row of the path diagram).

To investigate further the robustness of this effect, the following additional test was conducted. Fifty combinatorially naive undergraduates from Stanford University were presented with the path problem. Here, the subjects were not asked to estimate relative frequency but merely to judge "whether there are more paths containing six X's and no O, or more paths containing five X's and one O." The subjects were run individually, and they were promised a $1 bonus for a correct judgment. The significant majority of subjects (38 of 50, $p < .001$, by sign test) again selected the former outcome as more frequent. Erroneous intuitions, apparently, are not easily rectified by the introduction of monetary payoffs.

We have proposed that when the binomial distribution is represented as a path diagram, people judge the relative frequency of the various outcomes by assessing the availability of individual paths of each type. This mode of evaluation is suggested by the sequential character of the definition of a path and by the pictorial representation of the problem. Consider next an alternative formulation of the same problem.

> "Six players participate in a card game. On each round of the game, each player receives a single card drawn blindly from a well-shuffled deck. In the deck, 5/6 of the cards are marked X and the remaining 1/6 are marked O. In many rounds of the game, what is the percentage of rounds in which
>
> 6 players receive X and no player receives O ____%
> 5 players receive X and 1 player receives O ____%
> .
> .
> .
> No player receives X and 6 players receive O ____%
>
> Note that these include all the possible outcomes and hence your estimates should add to 100%."

This card problem is formally identical to the path problem, but it is intended to elicit a different mode of evaluation. In the path problem, individual instances were emphasized by the display, and the population proportion (i.e., the proportion of X's in each row) was not made explicit. In the card problem, on the other hand, the population proportion is explicitly stated and no mention is made of individual instances. Consequently, we hypothesize that the outcomes in the card problem will be evaluated by the degree to which they are representative of the composition of the deck rather than by the availability of individual instances. In the card problem, the outcome "five X's and one O" is the most representative, because it matches the population proportion (see Kahneman & Tversky, 1972). Hence, by the representativeness heuristic, this outcome should be judged more frequent than the outcome "six X's and no O," contrary to the observed pattern of judgments in the path problem. The judgments of 71 of 82 subjects who answered the card problem conformed to this prediction. In the path problem, only 13 of 73 subjects had judged these outcomes in the same way; the difference between the two versions is highly significant ($p < .001$, by a χ^2 test).

Median estimates for the card problem are presented in Fig. 2b. The contrast between Figs. 2a and 2b supports the hypothesis that different representations of the same problem elicit different heuristics. Specifically,

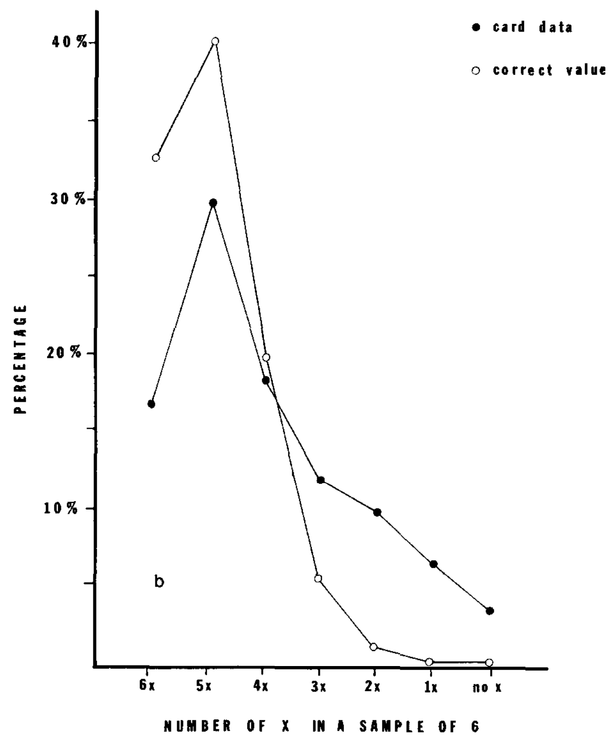

FIG. 2b. Correct values and median judgments: Card problem.

the frequency of a class is likely to be judged by availability if the individual instances are emphasized and by representativeness if generic features are made salient.

IV. AVAILABILITY FOR RETRIEVAL

In this section we discuss several studies in which the subject is first exposed to a message (e.g., a list of names) and is later asked to judge the frequency of items of a given type that were included in the message. As in the problems studied in the previous section, the subject cannot recall and count all instances. Instead, we propose, he attempts to recall some instances and judges overall frequency by availability, i.e., by the ease with which instances come to mind. As a consequence, classes whose instances are readily recalled will be judged more numerous than classes of the same size whose instances are less available. This prediction is first tested in a study of the judged frequency of categories. Next, we review previous evidence of availability effects on the judged frequency of repetitions. Finally, the role of the availability heuristic in judgments of the frequency of co-occurrences is discussed.

Study 8: Fame, Frequency, and Recall

The subjects were presented with a recorded list consisting of names of known personalities of both sexes. After listening to the list, some subjects judged whether it contained more names of men or of women, others attempted to recall the names in the list. Some of the names in the list were very famous (e.g., Richard Nixon, Elizabeth Taylor), others were less famous (e.g., William Fulbright, Lana Turner). Famous names are generally easier to recall. Hence, if frequency judgments are mediated by assessed availability, then a class consisting of famous names should be judged more numerous than a comparable class consisting of less famous names.

Four lists of names were prepared, two lists of entertainers and two lists of other public figures. Each list included 39 names recorded at a rate of one name every 2 sec. Two of the lists (one of public figures and one of entertainers) included 19 names of famous women and 20 names of less famous men. The two other lists consisted of 19 names of famous men and 20 names of less famous women. Hence, fame and frequency were inversely related in all lists. The first names of all personalities always permitted an unambiguous identification of sex.

The subjects were instructed to listen attentively to a recorded message. Each of the four lists was presented to two groups. After listening to the recording, subjects in one group were asked to write down as many names as they could recall from the list. The subjects in the other group were asked to judge whether the list contained more names of men or of women.

Results. (a) Recall. On the average, subjects recalled 12.3 of the 19 famous names and 8.4 of the 20 less famous names. Of the 86 subjects in the four recall groups, 57 recalled more famous than nonfamous names, and only 13 recalled fewer famous than less famous names ($p < .001$, by sign test).

(b) Frequency. Among the 99 subjects who compared the frequency of men and women in the lists, 80 erroneously judged the class consisting of the more famous names to be more frequent ($p < .001$, by sign test).

Frequency of Repetitions

The preceding study supported the notion that people judge the frequency of a class by assessed availability, i.e., by the ease with which the relevant instances come to mind. In that study, subjects judged the frequency of classes which consisted of distinct instances, e.g., female entertainers or male politicians. Most research on judged frequency, in contrast, has been concerned with the frequency of repetitions, e.g., the number of times that a particular word was repeated in a list.

When the number of repetitions is relatively small, people may attempt to estimate the frequency of repetitions by recalling specific occurrences. There is evidence (see, e.g., Hintzman & Block, 1971) that subjects retain some information about the specific occurrences of repeated items. There are situations, however, in which occurrences cannot be retrieved, e.g., when the total number of items is large, when their distinctiveness is low, or when the retention interval is long. In these situations, subjects may resort to a different method for judging frequency.

When an item is repeated several times in a list, the association between the item and the list is strengthened. Thus, a subject could use the strength of this association as a clue to the frequency of the item. Hence, one could judge the frequency of repetitions either by assessing the availability of specific occurrences or by a more global assessment of the strength of the item-list association. As a consequence, factors which either enhance the recallability of specific occurrences or strengthen the association between item and list should increase the apparent frequency of the item. This analysis of frequency judgments is closely related to the theoretical treatments proposed by Hintzman and Block (1971) and by Anderson and Bower (1972). A somewhat different analysis has been offered by Underwood (1969a).

The general notion that factors which affect availability have a corresponding effect on the apparent frequency of repetitions has been supported in several studies. For example, the occurrences of an item are more likely to be stored and recalled as distinct units when they are widely spaced. Indeed, Underwood (1969b) showed that items are judged more frequent under conditions of distributed rather than massed practice, and Hintzman (1969) showed that the apparent frequency of an item increases with the spacing between its repetitions in the list. Another factor which enhances the memorability of repetitions is vocal rehearsal. Correspondingly, Hopkins, Boylan, and Lincoln (1972) showed that items that were pronounced were perceived as more frequent than items that were read silently.

According to the present analysis, the judgment of frequency is often mediated by an assessment of item-list associations. In many situations, however, the items to which the list is most strongly associated are also the items that are most likely to be retrieved when the subject attempts to recall the list. Hence, the recallability of items from a list provides an indirect measure of the strength of the association from these items to the list. As a consequence, there should be a positive correlation between the recallability of items and their apparent frequency. Indeed, the studies of Leicht (1968) and Underwood, Zimmerman, and Freund (1971) showed that, at any level of actual frequency, items that were better recalled were judged more frequent.

In concluding the discussion of the apparent frequency of repetition, it is important to emphasize that the availability heuristic is not the only method by which frequency of repetition can be estimated. In some contexts, people may have access to a "frequency counter" (see Underwood, 1969a). In other contexts, when the number of repetitions is large (see, e.g., Howell, 1970), frequency judgments may be mediated by an assessment of rate of occurrence, or inferred from a schema of the relevant structure. For example, in estimating the number of trials in which the red light came on rather than the blue or the green, in a 1000-trial probability-learning experiment, the subject probably infers the estimate from his schema of the statistical structure of the sequence. Frequency estimates obtained from studies of binary and multiple probability learning show that, in general, people are quite accurate in judging relative frequencies of events (see Vlek, 1970, for a review). To the extent that availability plays a role in these judgments, it is probably by affecting the schema to which the subject refers in estimating frequency.

Frequency of Co-occurrence

Some recent research has been concerned with judgment of the frequency with which pairs of items have occurred together. The strategies employed to estimate the frequency of a single item can also be employed to estimate the frequency of an item-pair. In addition, the repetition of a pair strengthens the association between its members. The subject may, therefore, use the strength of the association between the members of a pair as a clue to its frequency.

An interesting bias in the judgment of the frequency of co-occurrence has been reported by Chapman (1967) and Chapman and Chapman (1967, 1969). In the initial study, Chapman used two sets of words, and constructed a list in which each word in the first set was paired with each word in the second set. All pairs were visually presented an equal number of times. The subjects were told in advance that they would be required to report how often each word was paired with each other word. In spite of this warning, they made consistent errors in their subsequent judgments of frequency. The frequency of the co-occurrence of related words was overestimated, creating an *illusory correlation* between such words. For example, *lion–tiger* was incorrectly judged to have been shown more often than *lion-eggs*, and *bacon-eggs* was judged more frequent than *bacon–tiger*. A similar illusory correlation was found between unusually long words. For example, *blossoms-notebook* was erroneously judged to have been shown more often than *boat–notebook*. Chapman attributed this result to the distinctiveness of the long words.

In subsequent studies, Chapman and Chapman (1967, 1969) investigated the significant implications of the phenomenon of illusory correlation to impression formation and clinical judgment. They presented naive judges with clinical test material and with clinical diagnoses for several hypothetical patients. Later, the judges evaluated the frequency of co-occurrence of various symptoms and diagnoses in the data to which they had been exposed. Illusory correlation was again observed. The judges markedly overestimated the co-occurrence of pairs that were judged to be natural associates by an independent group of subjects. For example, "suspiciousness" had been rated as calling to mind "eyes" more than any other part of the body. Correspondingly, the judges greatly overestimated the frequency of the co-occurrence of suspiciousness with peculiar drawing of the eyes in the Draw-a-Person test. An ominous finding in the Chapmans' study was that naive judges erroneously "discovered" much of the common but unvalidated clinical lore concerning the interpretation of the Draw-a-Person and the Rorschach tests. Furthermore, the illusory correlation effect was extremely resistant to contradictory data. It persisted even when the actual correlation between the associates was negative. Finally, the illusory correlation effect prevented the judges from detecting correlations that were in fact present in the test material (see also Golding & Rorer, 1972).

Availability provides a natural explanation for illusory correlation.

We propose that an assessment of the associative bond between two items is one of the processes that mediate the judged frequency of their co-occurrence. The association between two items is strengthened whenever they co-occur. Thus, when a person finds that the association between items is strong, he is likely to conclude that they have been frequently paired in his recent experience. However, repetition is not the only factor that affects associative strength. Factors other than repetition which strengthen the association between the members of a pair will, therefore, increase the apparent frequency of that pair. According to this account, illusory correlation is due to the differential strength of associative bonds. The strength of these bonds may reflect prior association between the items or other factors, such as pair-distinctiveness, which facilitate the formation of an association during learning. Thus, the various sources of illusory correlation can all be explained by the operation of a single mechanism—the assessment of availability or associative strength. The proposed account of the judgment of the frequency of co-occurrences is tested in the last two studies.

Study 9: Illusory Correlation in Word Pairs

This study essentially replicates Chapman's (1967) original result and establishes the relation between judgments of the frequency of pairs and cued recall, i.e., the recall of the second word of the pair, called response, given the first, called stimulus.

A set of twenty pairs of words was constructed. Ten of the pairs consisted of highly related (HR) words, the other ten consisted of unrelated (UR) words. In five of the HR pairs, stimulus and response were natural associates: *knife–fork, hand–foot, lion–tiger, table–chair, winter–summer.* (The first three pairs were taken from Chapman's list.) In five other pairs, stimulus and response were phonetically similar: *gown–clown, cake–fake, blade–blame, flight–fleet, spoon–spanner.* The ten UR pairs were obtained by replacing the stimulus word in each of the above ten pairs, respectively, by the words: *head, lamp, house, paper, dish, bread, box, pencil, book, phone.* Thus, the entire set of pairs was constructed so that each response word appeared with two stimulus words, one which was highly related to it and one which was not. A message which included these word-pairs was recorded on tape at a rate of one pair every 5 sec. Ten of the twenty pairs were repeated three times in the message and the other ten pairs were repeated twice. Pairs that shared the same response word (e.g., *knife–fork, head–fork*) were repeated the same number of times. The order of the pairs was randomized. To minimize the effects of primacy and recency, the same two filler pairs were recorded both at the beginning and at the end of the message.

All subjects ($N = 98$) were instructed to listen attentively to the message. Following the recording, one group of 30 subjects was asked for cued recall: each subject was given a list of all twenty stimulus words (in one of four random orders) and was asked to write the corresponding response words. A second group of 68 subjects was asked for frequency judgments: each subject was given a list of all twenty pairs (again, in one of four random orders) and was asked to judge whether each of the pairs had appeared twice or three times in the message.

Results. (a) Cued recall. For each subject, the number of response words correctly recalled was counted, separately for the HR and the UR pairs under each of the two repetition levels (i.e., 2 and 3). Table 1a presents the mean probability of recall for each of the four conditions. A 2×2 analysis of variance showed that subjects recalled significantly more words from the HR pairs than from the UR pairs ($t = 9.4$, 29 *df*, $p < .001$), and that they recalled significantly more words from the pairs that had been repeated more often ($t = 2.44$, 29 *df*, $p < .05$). The interaction between the two factors was not significant.

(b) Judged frequency. Table 1b presents the mean judged frequency of the HR and the UR pairs for the two levels of actual frequency. A 2×2 analysis of variance showed that the HR pairs were judged more frequent than the UR pairs ($t = 4.62$, 67 *df*, $p < .001$), although they were, in fact, equally frequent. The effect of actual frequency was also significant ($t = 7.71$, 67 *df*, $p < .001$). The interaction between the two factors was not.

Further analyses showed that the differences between HR and UR pairs, in both cued recall and judged frequency, were significant separately for the natural associates and for the phonetically similar pairs.

Study 10: Illusory Correlation in Personality Traits

Chapman's original study, as well as Study 9, employed a correlational design where each response was paired with more than one stimulus. Ac-

TABLE 1
Mean Probability of Recall and Mean Judged Frequency

		(a) Cued recall				(b) Judged frequency	
		Relatedness				Relatedness	
		Low	High			Low	High
Actual frequency	3	.41	.85	Actual frequency	3	2.45	2.63
	2	.31	.77		2	2.26	2.42

cording to the present analysis, however, the illusory correlation effect is due to differences among item pairs in the strength of the associative bond between their members. Consequently, the same effect should also occur in a noncorrelational design, where each response is paired with a single stimulus, and vice versa. The present study tests this prediction. In addition, it shows that people can assess the availability of associates, i.e., the degree to which the response word is made available by the stimulus word.

A set of sixteen pairs of personality traits was constructed. Eight of the pairs—the highly related pairs—consisted of traits which tend to be associated with each other. The other eight pairs—the unrelated pairs—consisted of traits which are not generally associated with each other. The highly related (HR) pairs were: *kind–honest, poised–relaxed, passive–withdrawn, alert–witty, selfish–greedy, meek–silent, brutal–nasty, healthy–active*. The unrelated (UR) pairs were: *nervous–gentle, lucky–discreet, eager–careful, clever–prudent, humble–messy, nice–anxious, casual–thrifty, clumsy–mature*. In a pilot study designed to validate the classification of the pairs, 36 subjects assessed, for each pair, the probability that a person who has the first trait of that pair also has the second (e.g., the probability that an alert person is witty). The average estimated probabilities for each of the HR pairs exceeded the average estimates for all the UR pairs.

A message which included all pairs was recorded on tape at a rate of one pair every 5 sec. Two HR and two UR pairs appeared in the list at each of four levels of frequency, from a single occurrence to four occurrences. The order of the pairs in the message was randomized and five filler pairs were recorded at the beginning and the end of the message.

All subjects were told to listen attentively to a recorded message. Following the recording, subjects were assigned one of three different tasks. The subjects in the *recall* group ($N = 62$) were given a list consisting of all 16 stimulus-traits and were asked to recall the response member of each pair. The subjects in the *assessed-recall* group ($N = 68$) were presented with the 16 trait-pairs and were asked to indicate, on a seven-point scale, the likelihood that they would have been able to recall each response-trait if they had been given the stimulus-trait, immediately after hearing the list. The subjects in the *judged-frequency* group ($N = 73$) were given a list of all the 16 trait-pairs and were asked to judge how often each pair appeared in the message. Four lists with different orders were employed for each of the three tasks.

Results. (a) Recall. The number of items that were correctly recalled by each subject was recorded separately for the HR and the UR pairs. On the average, subjects correctly completed 41% of the HR pairs, and only 19% of the UR pairs. The difference is highly significant ($t = 9.27$, 61 df, $p < .001$).

(b) Assessed recall. The mean rating of assessed recall was computed for each of the trait-pairs. The product–moment correlation, over the 16 pairs, between mean assessed recall and the proportion of correct responses in the recall group was 0.84. Apparently, people can assess the recallability of associates with reasonable accuracy.

(c) Judged frequency. Figure 3 shows mean judged frequency as a function of actual frequency, separately for the HR and the UR pairs. The difference between the two curves is highly significant ($t = 3.85$, 72 df, $p < .001$).

Although judgments of frequency were generally accurate, a slight but highly systematic bias favoring related pairs was present. The results support the proposed account of judgment of frequency in terms of the availability of associations, and demonstrate the presence of "illusory correlation" in a non-correlational design.

V. RETRIEVAL OF OCCURRENCES AND CONSTRUCTION OF SCENARIOS

In all the empirical studies that were discussed in this paper, there existed an objective procedure for enumerating instances (e.g., words that begin with *K* or paths in a diagram), and hence each of the problems had an objectively correct answer. This is not the case in many real-

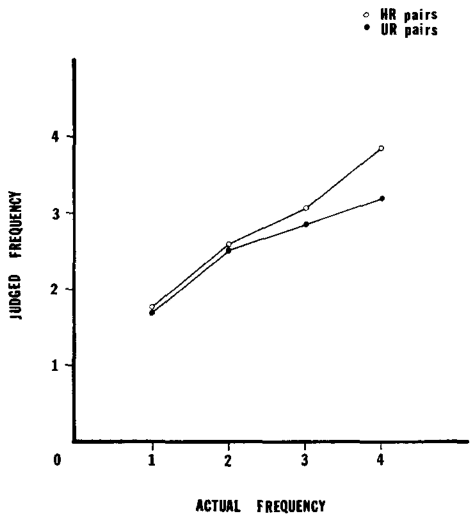

FIG. 3. Average judged frequency as a function of actual frequency for highly-related (HR) and unrelated (UR) trait-pairs.

life situations where probabilities are judged. Each occurrence of an economic recession, a successful medical operation, or a divorce, is essentially unique, and its probability cannot be evaluated by a simple tally of instances. Nevertheless, the availability heuristic may be applied to evaluate the likelihood of such events.

In judging the likelihood that a particular couple will be divorced, for example, one may scan one's memory for similar couples which this question brings to mind. Divorce will appear probable if divorces are prevalent among the instances that are retrieved in this manner. Alternatively, one may evaluate likelihood by attempting to construct stories, or scenarios, that lead to a divorce. The plausibility of such scenarios, or the ease with which they come to mind, can provide a basis for the judgment of likelihood. In the present section, we discuss the role of availability in such judgments, speculate about expected sources of bias, and sketch some directions that further inquiry might follow.

We illustrate availability biases by considering an imaginary clinical situation.[6] A clinician who has heard a patient complain that he is tired of life, and wonders whether that patient is likely to commit suicide may well recall similar patients he has known. Sometimes only one relevant instance comes to mind, perhaps because it is most memorable. Here, subjective probability may depend primarily on the similarity between that instance and the case under consideration. If the two are very similar, then one expects that what has happened in the past will recur. When several instances come to mind, they are probably weighted by the degree to which they are similar, in essential features, to the problem at hand.

How are relevant instances selected? In scanning his past experience does the clinician recall patients who resemble the present case, patients who attempted suicide, or patients who resemble the present case *and* attempted suicide? From an actuarial point of view, of course, the relevant class is that of patients who are similar, in some respects, to the present case, and the relevant statistic is the frequency of attempted suicide in this class.

Memory search may follow other rules. Since attempted suicide is a dramatic and salient event, suicidal patients are likely to be more memorable and easier to recall than depressive patients who did not attempt suicide. As a consequence, the clinician may recall suicidal patients he has encountered and judge the likelihood of an attempted suicide by the degree of resemblance between these cases and the present patient. This

[6] This example was chosen because of its availability. We know of no reason to believe that intuitive predictions of stockbrokers, sportscasters, political analysts or research psychologists are less susceptible to biases.

approach leads to serious biases. The clinician who notes that nearly all suicidal patients he can think of were severely depressed may conclude that a patient is likely to commit suicide if he shows signs of severe depression. Alternatively, the clinician may conclude that suicide is unlikely if "this patient does not look like any suicide case I have met." Such reasoning ignores the fact that only a minority of depressed patients attempt suicide and the possibility that the present patient may be quite unlike any that the therapist has ever encountered.

Finally, a clinician might think only of patients who were both depressed and suicidal. He would then evaluate the likelihood of suicide by the ease with which such cases come to mind or by the degree to which the present patient is representative of this class. This reasoning, too, is subject to a serious flaw. The fact that there are many depressed patients who attempted suicide does not say much about the probability that a depressed patient will attempt suicide, yet this mode of evaluation is not uncommon. Several studies (Jenkins & Ward, 1963; Smedslund, 1963; Ward & Jenkins, 1965) showed that contingency between two binary variables such as a symptom and a disease is judged by the frequency with which they co-occur, with little or no regard for cases where either the symptom or the disease was not present.

Some events are perceived as so unique that past history does not seem relevant to the evaluation of their likelihood. In thinking of such events we often construct *scenarios*, i.e., stories that lead from the present situation to the target event. The plausibility of the scenarios that come to mind, or the difficulty of producing them, then serve as a clue to the likelihood of the event. If no reasonable scenario comes to mind, the event is deemed impossible or highly unlikely. If many scenarios come to mind, or if the one scenario that is constructed is particularly compelling, the event in question appears probable.

Many of the events whose likelihood people wish to evaluate depend on several interrelated factors. Yet it is exceedingly difficult for the human mind to apprehend sequences of variations of several interacting factors. We suggest that in evaluating the probability of complex events only the simplest and most available scenarios are likely to be considered. In particular, people will tend to produce scenarios in which many factors do not vary at all, only the most obvious variations take place, and interacting changes are rare. Because of the simplified nature of imagined scenarios, the outcomes of computer simulations of interacting processes are often counter-intuitive (Forrester, 1971). The tendency to consider only relatively simple scenarios may have particularly salient effects in situations of conflict. There, one's own moods and plans are more available to one than those of the opponent. It is not easy to adopt the opponent's view of the chessboard or of the battlefield, which may be why the mediocre player discovers so many new possibilities when he switches sides in a game. Consequently, the player may tend to regard his opponent's strategy as relatively constant and independent of his own moves. These considerations suggest that a player is susceptible to the *fallacy of initiative*—a tendency to attribute less initiative and less imagination to the opponent than to himself. This hypothesis is consistent with a finding of attribution-research (Jones & Nisbett, 1971) that people tend to view their own behavior as reflecting the changing demands of their environment and others' behavior as trait-dominated.

The production of a compelling scenario is likely to constrain future thinking. There is much evidence showing that, once an uncertain situation has been perceived or interpreted in a particular fashion, it is quite difficult to view it in any other way (see, e.g., Bruner & Potter, 1964). Thus, the generation of a specific scenario may inhibit the emergence of other scenarios, particularly those that lead to different outcomes.

Images of the future are shaped by the experience of the past. In his monograph *Hazard and choice perception in flood plain management*, Kates (1962) writes:

> "A major limitation to human ability to use improved flood hazard information is a basic reliance on experience. Men on flood plains appear to be very much prisoners of their experience . . . Recently experienced floods appear to set an upper bound to the size of loss with which managers believe they ought to be concerned [p. 140]."

Kates attributes much of the difficulty in achieving more efficient flood control to the inability of individuals to imagine floods unlike any that have occurred.

Perhaps the most obvious demonstration of availability in real life is the impact of the fortuitous availability of incidents or scenarios. Many readers must have experienced the temporary rise in the subjective probability of an accident after seeing a car overturned by the side of the road. Similarly, many must have noticed an increase in the subjective probability that an accident or malfunction will start a thermonuclear war after seeing a movie in which such an occurrence was vividly portrayed. Continued preoccupation with an outcome may increase its availability, and hence its perceived likelihood. People are preoccupied with highly desirable outcomes, such as winning the sweepstakes, or with highly undesirable outcomes, such as an airplane crash. Consequently, availability provides a mechanism by which occurrences of extreme utility (or disutility) may appear more likely than they actually are.

A Final Remark

Most important decisions men make are governed by beliefs concerning the likelihood of unique events. The "true" probabilities of such events are elusive, since they cannot be assessed objectively. The subjective probabilities that are assigned to unique events by knowledgeable and consistent people have been accepted as all that can be said about the likelihood of such events.

Although the "true" probability of a unique event is unknowable, the reliance on heuristics such as availability or representativeness, biases subjective probabilities in knowable ways. A psychological analysis of the heuristics that a person uses in judging the probability of an event may tell us whether his judgment is likely to be too high or too low. We believe that such analyses could be used to reduce the prevalence of errors in human judgment under uncertainty.

REFERENCES

ANDERSON, J. R., & BOWER, G. H. Recognition and retrieval processes in free recall. *Psychological Review*, 1972, **79**, 97–132.

BOUSFIELD, W. A., & SEDGEWICK, C. H. An analysis of sequences of restricted associative responses. *Journal of General Psychology*, 1944, **30**, 149–165.

BRUNER, J. S., & POTTER, M. C. Interference in visual recognition. *Science*, 1969, **144**, 424–425.

CHAPMAN, L. J. Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 151–155.

CHAPMAN, L. J., & CHAPMAN, J. P. Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 1967, **73**, 193–204.

CHAPMAN, L. J., & CHAPMAN, J. P. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 1969, **74**, 271–280.

FORRESTER, J. W. *World dynamics*. Cambridge, Mass.: Wright-Allen, 1971.

GOLDING, S. L., & RORER, L. G. "Illusory correlation and the learning of clinical judgment." *Journal of Abnormal Psychology*, 1972, **80**, 249–260.

HART, J. T. Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 689–691.

HINTZMAN, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, **80**, 139–145.

HINTZMAN, D. L., & BLOCK, R. A. Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 1971, **88**, 297–306.

HOPKINS, R. H., BOYLAN, R. J., & LINCOLN, G. L. Pronunciation and apparent frequency. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 105–113.

HOROWITZ, L. M., NORMAN, S. A., & DAY, R. S. Availability and associative symmetry. *Psychological Review*, 1966, **73**, 1–15.

HOWELL, W. C. Intuitive "counting" and "tagging" in memory. *Journal of Experimental Psychology*, 1970, **85**, 210–215.

JENKINS, H. M., & WARD, W. C. Judgment of contingency between responses and outcomes. *Psychological Monographs*, 1965, **79**, (1, Whole No. 594).

JONES, E. E., & NISBETT, R. E. The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner. *Attribution: Perceiving the causes of behavior.* General Learning Press, 1971.

KAHNEMAN, D., & TVERSKY, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology,* 1972, **3**, 430–454.

KAHNEMAN, D., & TVERSKY, A. On the psychology of prediction. *Psychological Review,* 1973, in press.

KATES, R. W. Hazard and choice perception in flood plain management. Department of Geography Research Paper No. 78, University of Chicago, 1962.

LEICHT, K. L. Recall and judged frequency of implicitly occurring words. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 918–923.

MAYZNER, M. S., & TRESSELT, M. E. Tables of single-letter and bigram frequency counts for various word-length and letter-position combinations. *Psychonomic Monograph Supplements*, 1965, **1**(**2**), 13–32.

PHILLIPS, L. D. Some components of probabilistic inference. Technical Report No. 1, Human Performance Center, University of Michigan, 1966.

SMEDSLUND, J. Note on learning, contingency, and clinical experience. *Scandinavian Journal of Psychology,* 1966, **7**, 265–266.

TULVING, E., & PEARLSTONE, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior,* 1966, **5**, 381–391.

TVERSKY, A., & KAHNEMAN, D. Belief in the law of small numbers. *Psychological Bulletin,* 1971, **76**, 105–110.

UNDERWOOD, B. J. Attributes of memory. *Psychological Review,* 1969, **76**, 559–573. (a)

UNDERWOOD, B. J. Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 83–94. (b)

UNDERWOOD, B. J., ZIMMERMAN, J., & FREUND, J. S. Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, 1971, **87**, 149–162.

VLEK, C. A. J. Multiple probability learning: Associating events with their probabilities of occurrence. *Acta Psychologica*, 1970, **33**, 207–232.

WARD, W. C., & JENKINS, H. M. The display of information and the judgment of contingency. *Canadian Journal of Psychology*, 1965, **19**, 231–241.

(Accepted January 2, 1973)