Availability Heuristic in Judgments of Set Size and Frequency of Occurrence

Melvin Mann, Jonathan Shefler, John Joudes, and Thomas E. Nelson

The availability heuristic has been widely cited as an important factor in judgment processes. However, the evidence that availability is important in judging category size is not as compelling. Moreover, several reports suggest that availability may be less of a factor in judgments of frequency and occurrence. Path analysis was used to examine the role of memory availability in judgments of category size and frequency of occurrence. In judgments of size, there was modest support for the availability heuristic, but in size judgments, the judgments were mildly influenced by the context of memory. By contrast, in accordance with earlier results, availability was not a significant factor when size judgments involved frequency of occurrence.

In 1973, Tversky and Kahneman published a classic article on the availability heuristic, a cognitive strategy thought to play a role in judgments of frequency and probability. They suggested that the availability heuristic was operative "when someone evaluates the frequency of classes or the probability of events by the ease with which relevant instances come to mind" (p. 207). Subsequent research generalized this conception, implicating the availability notion in a wide range of judgments, including memory for messages (e.g., Shefler & Ross, 1983), and semantic memory (see Shefler & Cott, 1984). For an important limitation on the availability heuristic based on the distinction between "on-line" and "memory-based" judgments, see Hanse & Park (1986).

Judgments are ubiquitous in everyday life, and they have been studied extensively in social psychology laboratories. This article examines the role of availability in two types of judgments: frequency-of-occurrence judgments and set-size judgments. We show that there is consistent support for the notion that availability is a significant determinant of set-size judgments; availability does not, however, seem to affect frequency-of-occurrence estimates.

Frequency-of-Occurrence Judgments

On returning home from a business trip, Mr. Executive may ask his wife how often their children had gotten into arguments during his absence. Note that whereas a frequency-of-occurrence estimate is invited here ("How often did it occur?") the events in question differ obviously from one another. As a consequence, the judges task may be treated in a sort of off-line or category-size judgment (see below), in which the judge considers the category of children and assigns an estimate for the number of children exemplar that occurred during his husband's absence.

In contrast with the case of Mr. Executive and his argumentative children, however, frequency estimates may be treated as repeated occurrences of an infrequent event. The repeated occurrence of a given word in a speech is a frequent event. In selecting new reading materials for elementary school, each day on an object that is subjective estimate is in the student's familiarity with problematic words (e.g., the word 'dog' on the basis of their prior reading assignments). Or, in a more political example, sentiment may estimate the number of times that Vice-President John Ford used the phrase "I don't recall" during the televised Watergate hearings. What role does availability play in judgments like these?

Hastie and Zacks (1984) reviewed an extensive set of experiments concerning frequency-of-occurrence judgments. Availability does not appear to be an important variable in such experiments. Frequency-of-occurrence judgments are, however, quite sensitive to actual event frequency. Hastie and Zacks concluded that frequency-of-occurrence information is stored through an "explicit or automatic encoding process" (p. 1372) that is "fundamental and inevitable" - a process that would presumably be independent of the availability heuristic.

Some theories have suggested that people may be equipped with cognitive systems that automatically code frequency of occurrence (Hastie & Zacks, 1984; Newell, 1973). Underwood, 1989, direct coding models to our cognitive accounts for some aspects of the frequency judgments. However (Huneman, Newell, & Underwood, 1982; Whinston & Underwood, 1982), other theories have been suggested more forms of indirect coding as in the choice models (e.g., Underwood, 1982; Newell, 1973). For discussions, frequency estimates are thought to be comprised under pairs and direct, computation is thought to be dependent on the number of memory traces or the strength of the traces associated with each event. The interplay between di-
As evidence for the automaticity of frequency judgments, Hastie and Zacks (1976) cited an important study by Howell (1973). In this experiment, respondents were shown a list of words presented between 10 and 100 times. After the list presentation, respondents were asked to judge the frequency with which each word had been shown. The results indicated that there was virtually no difference between the estimates of subjects who had been fully informed that the experiment was concerned with event frequency and those who had received instructions focusing on recall. In both groups the judged frequency of presentation was directly related to the actual presentation frequencies of the stimulus words. These results are consistent with the possibility that frequency information is encoded automatically, i.e., regardless of the respondent's conscious intentions; also see Navon, Benjamin, and Jones, (1986).

Howell's study (1973) is particularly instructive from the availability point of view. Subjects who expected the recall test remembered more of the words from the test list. If the availability of the individual study words served as an effective guide to frequency estimation, we might then anticipate that subjects who expected to be tested for recall (and who consequently remembered more of the study words) would produce higher frequency estimates, or more accurate frequency estimates, than subjects who had focused on event frequency. However, Howell's results did not follow this pattern, however. Indeed, they showed a modest overrecall of these predictions, in that subjects who had been given recall instructions produced slightly lower frequency estimates despite their superior recall of the words on the study list, and these estimates were somewhat less sensitive to the different event frequencies, especially when the event in question had been presented six or more times.

A study by Zacks, Hastie, and Sanft (1982, Experiment 1) provides an additional challenge to the availability formulation when applied to frequency-of-occurrence judgments. In this experiment, subjects were presented with four successive study lists. One group of respondents was tested for recall of the individual items following each list presentation. These subjects showed enhanced recall for the first list in the next. Similarly, their recall of the fourth list was significantly better than the recall of a contrasting group who had made frequency-of-occurrence judgments for the individual words on the first three lists. Despite this recall difference, however, between subjects who had previously been tested for recall and those who had completed a series of frequency-of-occurrence tasks instead, the two groups did not differ in their performance on a frequency-discrimination task that focused on the individual items from the fourth study list. In brief, once again, judgments of event frequency were unaffected by group differences in availability, i.e., differences in the respondents' ability to recall the events in question.

In a somewhat different approach to this issue, Sheder, Jones, and Mannis (1985) presented individual study words with different frequencies. Following this presentation the respondents were asked to list all the words they could recall; finally, they were shown the original word list plus several foils and estimated the frequency of presentation for each item. It was estimated that an important role in frequency judgment the words that a respondent could recall would be expected to be re
current higher frequency-of-occurrence estimates than the words he or she could not recall. The results did not support this prediction, however. When the presentation frequency of the individual words was held constant statistically there was virtually no relationship between the availability of a given word in memory and its judged frequency of occurrence.

Set-Size Judgments

Social life sometimes requires spontaneous judgment of set size (category size). For example, a friend who is considering an invitation to join a club may ask how many of the club members are Jewish. The availability concept suggests that in judging such a question one may be significantly influenced by the number of Jewish club members one can recall. However, Aiba, Chromack, Hastie, and Attig (1980) proposed an alternative theory suggesting that availability (recall) may be unimportant in cases like this. These investigators hypothesized that people may routinely attend to generic categories as they encounter particular events or exemplars. Their theory suggests that if an individual is shown a study list that included the words violin, chair, cow, cow, rose, and tea, he or she will "obligatorily" keep track of the number of exemplars in the relevant higher order categories (e.g., the number of musical instruments, the number of furniture items, etc.). Hastie and Zacks (1986) considered this view and demonstrated that in
tion concerning both event frequency and category size is stored in memory by an implicit or automatic encoding pro
cess. In a similar vein, Watkins and LeCompte (1990) ques
tioned the importance of exemplar recall, i.e., availability in set-size estimates by showing that the accuracy of set-size judge
dments depends to a large extent on other, generic sources of information.

In contrast with the theories just cited, the availability hy
potheticals (Tversky & Kahneman, 1973) suggest that judgments of set size may not be automatic and might instead depend on the respondent's capacity to bring relevant exemplars or associ
tions to mind. Early demonstrations of this possibility were dramatic but not conclusive. For example, in Tversky and Kahn
eeman's (1973) widely cited "famous names" experiment sub
tspects heard a list of names, approximately half of which were of men and half were of women. For some subjects the male names were more famous than the female names. These sub
tspects recalled more male than female names from the study list (i.e., the male names were more available); moreover, other sub
tspects who had heard the same list strangely erroneously believed that the list included more male than female names. For sub
tspects in a contrasting condition, the female names were more famous; this list yielded superior recall of the female names and led a separate sample of respondents to conclude (erroneously) that it included more female than male names.

The famous names experiment demonstrates that memory (availability) and judgments of set size are both affected by a common variable (name). Demonstrations of this sort do not necessarily mean that memory has a causal impact on set-size
judgment, however, it could well be that the experimental manipulations of fame produced two independent effects: a main effect for availability bias that favored the recall of the more famous names in the study list and (b) a separate judgment bias, in which the category to which the exemplars belonged was thought to be larger than the category with the less famous exemplars. An account like this might also be applied to the results reported by Williams and Dewey (1986), who showed that recall and set-size estimates were both affected, perhaps independently, by the rate at which the exemplars were displayed and (b) the number of exemplars in each category.

In this article, we present the results of three experiments that mainly focus on the causal impact of availability (measured by recall on set-size judgments. Our results provide consistent support for the proposition that set-size judgments are reliably affected by the availability in memory of relevant exemplars. By contrast, frequency-of-occurrence judgments do not appear to be mediated by availability (recall).

Experiment 1

Method

Subjects. The subjects were 73 introductory psychology students (approximately half men and half women who participated for course credit.

Procedure. Two lists of names were prepared. Each list contained the names of 40 public figures (20 men and 20 women). For one list, the men were more famous than the women; for example, George Bush, Lee Iacocca, and Burt Reynolds vs. Laura Ashley, Jane Austen, and Franca Hugot. For the other list, the female names were more famous, Carole King, Margaret Thatcher, and Barbara Walters vs. Candice Bergen, Charles Beardsley, and Martha Stewart.

The procedure for selecting stimulus names was as follows. First, subjects drew from a large corpus of experimental subjects named lists of public figures. Next, a second group of subjects rated these names for fame. On the basis of these ratings, we selected 60 very famous male names, 60 less famous male names, 60 very famous female names, and 60 less famous female names. Twenty names from each of these groups were randomly chosen to serve as stimuli, whereas the remaining names were returned to use as foil in a subsequent recognition test.

Procedure. The subjects were told that they would hear a list of names and that they would be asked some questions afterward. There were no other instructions. The names were presented by telephone, in a single random order, at an approximate rate of one name every 2 s. Subjects assigned to Condition 1 heard the list with the most famous male names, and subjects assigned to Condition 2 heard the list with the most famous female names.

Following presentation of the name list, each subject responded to the following yes-no question: "How many men's names do you think you heard on the list?" and "How many women's names do you think you heard on the list?" In addition, each subject completed a memory test. To control for possible order effects, half of the subjects gave their category-availability estimates before completing the memory test, and half gave the memory test before completing the memory test. There were no significant effects associated with this variation in task order; hence, the two groups were combined in the analyses reported below.

The memory task was in two parts. First, subjects were asked to write all of the names they could remember; they were permitted as much time as they needed to do this. Next, they completed a recognition test that contained the 40 (last names and 80 foil in a random order. The subjects rated each name on a 6-point scale, where 6 indicated high confidence that the name was on the original list and 1 indicated high confidence that the name was a foil.

Results

Availability. The experimental manipulation of fame was intended to produce a bias in the relative availability in memory of male and female names. Because a general trend on measures of availability has not emerged from the literature, we estimated five-alternative availability measures (free recall. We used the recall protocols for number of male and female names recalled. We defined the availability bias as the proportion of men among the names that each subject recalled (N of males/(N of males + N of females)).

Spow Order. We assigned rank numbers to the names that each subject recalled so that the first recalled name received a value of 1, the second a value of 2, and so on. We then computed the mean rank for the recalled male names and the mean rank for the recalled female names in a lower mean equaled greater availability. Availability bias was defined as the difference between the two means (male names minus female names). (c) Recognition. The subjects rated each name for recognition using 6-point scale. We computed the mean recognition rating for male stimulus names and for female stimulus names (foils were ignored). Availability bias was defined as the difference between these means. (d) Recognition confidence. By coding the discrepancy between the subjects' recognition responses and the center of the recognition rating scale, we obtained a measure of the confidence that subjects placed in their ratings (scale ratings of 1 and 6 received a confidence score of 3, ratings of 2 and 5 received a confidence score of 2, and ratings of 3 and 4 received a confidence rating of 1). Availability bias was then defined as the difference between the mean confidence score for male and female stimulus names.

Results. Table 1 shows the availability bias as a function of sex, stimulus sex, and the subjects' judgments of the stimulus sex. In all, 18 conditions were included in the experiment. For each of the five availability measures (all ps < .001). The experiment manipulation of fame produced a significant availability bias, regardless of the availability measure we used. As Table 1 reveals, subjects in Condition 1 showed an availability bias in favor of male names, whereas subjects in Condition 2 showed an availability bias in favor of the female names. Separate t tests revealed that subjects in the two experimental conditions differed significantly on each of the five availability measures (all ps < .001).}

Judgment. For each subject, we defined judgment bias as the proportion of male names judged to be on the originally presented name list (estimated # of males/estimated number of

Table 1: Mean Availability Scores for Five Availability Measures in Experiment 1

<table>
<thead>
<tr>
<th>Condition</th>
<th>Path</th>
<th>Free recall</th>
<th>Spow order</th>
<th>Recognition</th>
<th>Confidence</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1</td>
<td>0.71</td>
<td>5.1</td>
<td>1.5</td>
<td>0.44</td>
<td>1.3</td>
<td></td>
</tr>
<tr>
<td>Condition 2</td>
<td>1.39</td>
<td>2.7</td>
<td>2.5</td>
<td>-0.36</td>
<td>-1.1</td>
<td></td>
</tr>
</tbody>
</table>
as expected, the experimental manipulation of fame produced a clear judgment bias, with subjects in Condition 1 overestimating the proportion of male names on the original list (M = 0.57) and subjects in Condition 2 underestimating the proportion of male names (M = 0.45). For our purposes M is the mean judgment bias across the subjects in each group. The difference between these judgment biases was highly significant, t(11) = 5.5, p < .001.

These results replicate the results originally reported by Tversky and Kahneman (1973). The experimental manipulation of fame produced a memory bias in the relative availability of male and female names, and it also produced a bias in our subjects’ set-size judgments. However, without further analysis we cannot conclude that these data support the causal sequence that is required by the availability hypothesis because the results reported in the preceding paragraphs are equally consistent with a non-availability model in which the recall and judgment effects are independent of one another (i.e., not causally related).

Evaluating the availability hypothesis. To test the independent effects of (a) our experimental variable (List 1 vs. List 2) and (b) availability bias on (c) our respondents’ set-size judgments, we turned to path analysis (Aiken, 1985), with the results shown in Figure 1.

To conclude that the judgments in Experiment 1 depend on the availability heuristic, it is necessary that the path from the experimental manipulation (fame) to the availability bias variable be substantial and that the path from the availability bias to judgment bias also be substantial. Alternatively, the path model allows for the possibility that our experimental manipulation might have directly influenced the respondents’ judgment bias, without any significant mediation from the relative availability of male versus female names. The variables that were entered into the path model are those that were described earlier: judgment bias was defined as the proportion of male names judged to be on the list; availability bias (operationalized as free recall) was defined as the proportion of male names among the recalled names. Finally, the experimental manipulation, fame, was represented by a dichotomous variable coded 1 for subjects assigned to Condition 1 and 0 for subjects assigned to Condition 2. To obtain the path coefficients two regressions were performed: (a) the regression of judgments on the experimental manipulation and on availability and (b) the regression of availability on the experimental manipulation. The path coefficients in Figure 1 are the resulting regression beta weights; they may be interpreted as the change in a particular (standardized) variable, given a standard unit of change in the variable leading to it.

Examine Figure 1, we see that the experimental manipulation (Name List 1 vs. Name List 2) had a strong effect on availability (as indicated by the path coefficient of .61). The direct effect of the experimental manipulation, by contrast, had virtually no impact on our respondents’ judgments (the path coefficient here was only .03, n.s.).

Discussion

These results provide strong support for the availability hypothesis. Moreover, this same pattern of results was obtained when path analyses were calculated, with availability operationalized as free recall, speed order, or recognition. The availability pattern was not obtained, however, when availability was operationalized in terms of recognition confidence or of the path coefficients we obtained using the different availability measures are presented in Table 2.

Of the inspection of the table suggests that free recall may be the most promising of our various availability measures in the sense that it maximizes the causal link between memory (availability) and judgment. With availability operationalized by free recall, Experiment 1 provides substantial evidence that our subjects’ set-size judgments were importantly influenced by the category members that were most available in memory.

Experiment 2

Experiment 2 was designed to provide additional evidence pertaining to the availability hypothesis and to assess the stability of the availability effect over a brief (1 min) retention interval. Rather than relying on the respondents’ familiarity with different individuals (famous vs. less famous) as a means of inducing an availability bias, biases in availability were manipulated by experimental means. The delay condition was introduced for largely exploratory purposes to examine the possibility that the availability heuristic might diminish in importance as a determinant of set-size judgments when the relevant exemplars were more fully consolidated into memory. As we show.

Table 2. Path Coefficients for Experiment 1

<table>
<thead>
<tr>
<th>Path</th>
<th>Free recall</th>
<th>Speed order</th>
<th>Recognition</th>
<th>Confidence</th>
<th>a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fame to availability</td>
<td>0.61**</td>
<td>0.76**</td>
<td>0.76**</td>
<td>0.62**</td>
<td>0.74**</td>
</tr>
<tr>
<td>Fame to judgment</td>
<td>0.61**</td>
<td>0.42**</td>
<td>0.35**</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>Availability to judgment</td>
<td>0.03</td>
<td>0.21</td>
<td>0.25</td>
<td>0.41**</td>
<td>0.38*</td>
</tr>
</tbody>
</table>

*p < .05. ** p < .01.
this conjecture proved to be incorrect, because the introduction of a delay period did not significantly affect the results that were observed.

Method

Six-hour subjects were recruited from the student body at the University of Michigan. They were paid $5 for participating in the study. Subjects were shown a series of 40 faces, names, etc., Georgia, Delaware, Sally, half of men and half of women. The group of each male was shown the cases whereas each female name was shown the cases. A given name was then presented twice in succession in a random list, the female names were shown the cases each, and the male names were shown the cases each. Each was exposed for 5-8 min intervals by means of a electronically controlled computer projector.

Following the study list, subjects in the no-delay condition were asked to estimate the number of different male names and the number of different female names that had been presented. The instructions emphasized our interest in the number of distinct names in each set. For example, if a particular name (like Bill) was presented more than once, it was to be counted as a single male instance. Finally, the respondents were presented with a recall test in Experiment 1, each subject was simply instructed to "write down all the names that you recall" and if you can remember.

Subjects assigned to the delay condition followed the same set of procedures except that they completed an additional classification task immediately after the name list had been presented. More accurately, when the name list was completed, delay subjects were instructed to do the following:

1. Think of the 50 States of the Union. Imagine the name(s) of each of the states. a separate index card. Try to separate the states/indexes into two classifications. Please write the two classifications that come to mind, whether it occurs obviously or not. Classification #1 --- Classification #2.

2. Repeat question 1. Try to think of as many different ways of classifying the states as you can.

Classification #2.

Subjects were instructed to provide up to two separate classifications for the 70 states, a task that took about 1.7 min.

Following the paper classification task, we asked the delay subjects for their subject estimates, and they attempted to recall the various names on the study list.

Results and Discussion

The data from Experiment 2 were subjected to two path analyses separating the results obtained in the no-delay and the delay conditions. Recall was our measure of availability, as in Experiment 1 (i.e., proportions of male names among the names recalled); judgment bias was also defined as in Experiment 1 (i.e., the estimated proportion of male names in the study list). The results are shown in Figure 1, where the different curves are regression beta weights. Both path analyses are consistent with the results of Experiment 1, showing clear support for the availability hypothesis. Differences in the structure of the study list (presentation frequency) had a strong and significant impact on the availability bias, as measured by the number of male names that were recalled divided by all names that were recalled; β = .75 and .83, for the no-delay and delay respondents, respectively. Moreover, individual differences in the availability bias were substantially related to biases in the respondents' judgment bias; β = .49 and .57. Figure 2 also indicates that the delay-immediated effects of the different study lists were less critical as determinants of the respondents' judgment biases (β = .23 and .41). Although the direct effect deriving from presentation frequency reached an acceptable level of statistical significance (p < .01) in the delay condition, these results were not significantly different from the results observed in the no-delay condition (p > .40). In brief, the additional increased importance of the direct path under delay conditions is of dubious replicability.

Experiment 3

The preceding experiments present a consistent pattern: Session judgments were substantially affected by biases in the respondents' recall protocols, as suggested by the availability for- mulation (Tversky & Kahneman, 1973). But contrary, the literature on frequency-of-occurrence estimates has consistently failed to support (to availability conception the above). Experiment 3 was designed to provide further evidence concerning the role of availability in these two domains. In particular, we wanted to make certain that the contrasting patterns observed in these domains did not derive from some simple methodological artifact (e.g., the use of names vs. common nouns as stimuli). Experiment 3 used a common study list to examine the role of availability in judgments of (a) presentation frequency and (b) set size. We expected to find substantial support for the availability hypothesis when our respondents were judging set size but not when they were asked for frequency-of-occurrence estimates.

Method

Subjects: Sixty-three students at the University of Michigan (43 men and 21 women) participated in Experiment 3, they were paid $5 each.

Procedure: Subjects were seated in front of a personal computer. They were told that they would see a series of names and would be asked some questions afterwards. They were shown a total of 660 first names; 30 of them were men's and 30 of them were women's. The male and female names were randomly divided into three sets, each containing 33 names. For subjects assigned to Set A, the different male names were shown either twice, three times, or five times, whereas the female names were shown three, five times, or seven times, whereas the female names were shown three, three times.

1 For reviews that are presently underway, Shadish, 1987; Experiment 2) demonstrates a significant relationship between availability and settling judgments is an extension of what is currently related to the present Experiment 2 (p = .14). Thus, however, as expected this up our experience because it is contrary to the suggestion that we repeatedly observed in the current work and in other experiments that we have conducted on this issue.

2 Reader may be surprised by the fact that a path coefficient of .49 in Figure 1 is significantly less than the .57 found for Experiment 2 (path coefficient of .41). Figure 28 was significant p < .05. These results derived from differences between the standard errors of the coefficients obtained in this study and those of Experiment 2. The standard errors of the partial coefficients in Figure 1 are larger than the standard errors for Figure 28.
times, or few times. Within each list, the different names were presented in one of two random sequences, with the restriction that a given name (e.g., Mary) never appeared in two successive exposures. Each name was presented for a 4-s duration; there was a 1-s blank interval between the presentations of successive stimuli.

After the presentation of the study lists, subjects attempted to recall as many of the names as they could. Half of the subjects then made set-size estimates indicating the number of different male and female names that had been presented on the study list, whereas the other half estimated the presentation frequencies for the individual names.

As in our earlier experiments, biases in availability were quantified by a ratio measure: [male names recalled/male names recalled (male plus female)]. Biases in the respondents' set-size judgment were similarly defined (i.e., the number of male names in subjects' estimates/[the number of male names plus female names in subjects' estimates]). To estimate the frequency of presentation, judgments were also quantified by a ratio measure to indicate the perceived presentation frequency for the male versus the female names (i.e., average recall frequency of presentation for the different male names on the study list/[average recall frequency of presentation for the female names on the study list]).

Results

The data from Experiment 3 were subjected to path analysis. Figure 3 presents the results from those respondents who had made frequency-of-occurrence judgments. Figure 4 presents the results for those who had made set-size judgments.

Frequency estimates. Figure 3 indicates that the contrasting study lists substantially affected the respondents' recall protocols (i.e., whether they recalled more male or more female names from the study list). Not surprisingly, subjects whose study list included a relatively high presentation frequency for the male names produced a higher proportion of male names in recall, than those for whom the male names were shown less frequently ($\beta = .87$). It is important to note, however, that performance on the recall test was not a significant determinant of subjects' frequency estimates. That is, setting the subjects assigned to a given study list, biases in recall did not reliably predict whether the respondents believed that the male or the female names had been presented more frequently ($\beta = .22$). Biases in the respondents' frequency estimates were, however, directly affected by subjects' list assignment (i.e., by whether the male or the female names had been presented more frequently). $\beta = .76$. In brief, although the frequency-of-occurrence estimates for the male versus the female names could not be reliably predicted from the respondents' recall biases, they were predictable from the exposure frequencies the subjects had experienced.

Set-size judgments. Figure 4 presents the results obtained from the subjects who made set-size estimates. In contrast with the frequency-of-occurrence estimates (see Figure 3), the set-size judgments were substantially affected by biases in the free recall protocols ($\beta = .53$), which in turn were dependent on the contents of the contrasting study lists ($\beta = .68$). Finally, noise that there was no reliable evidence of a direct study list effect because the set-size judgments were independent of the condition study list in which a subject had been assigned ($\beta = .07$).

---

**Figure 3:** Path analytic results from Experiment 3. (Panel A derives from the set-size condition. Panel B derives from the delay condition. *P < .05. *P* < .01)

**Figure 4:** Path analytic results from Experiment 3: Frequency-of-occurrence estimates. *P* < .01

**Figure 5:** Path analytic results from Experiment 3: Set-size estimates. *P* = .03. *P* < .01
These results are consistent with the patterns of results observed in Experiment 1 and may provide substantial support for the availability heuristic as a determinant of set-size estimates. By contrast, the direct path linking exposure frequency to set-size judgments was substantially weaker and far from significant.

Discussion

Estimating frequency of occurrence. The results of Experiment 3 were consistent with earlier work in showing that judgments of presentation frequency are largely independent of recall. The relative availability in recall of male versus female names was thus not a significant factor in our respondents' judgments concerning the presentation frequency of the individual male versus female names on the study list.

Estimating set size. Estimates of set size were strongly and positively affected by the names most available in memory. The direct unmediated effect of presentation frequency was weak and unreliable, though positive in our earlier studies. These results replicate the pattern observed in Experiments 1 and 2.

General Discussion

Experiment 3 provides added support for the idea that frequency-of-occurrence judgments derive from an implicit automatic counting process (Hastie & Zacks, 1984; Shede1, et al., 1985), a process that does not appear to rely on the availability heuristic. These data do not, however, allow us to choose between the theories that posit a direct, internal-counting mechanism that is independent of other memory processes (Hastie & Zacks, 1984; Underwood, 1983) and the theories that propose a size-determined (or computed) basis for frequency estimates on the basis of the number and the strength of the relevant memory traces (Krugman, 1976; Howell, 1973).

The size cue yielded a rather different pattern of results. All three of our experiments showed that memorability availability was significantly related to subjects' category-size judgments. That is, in estimating the number of male and female names on the study list, our respondents were consistently affected by the ratio of male to female names that came to mind in the recall test. These results are consistent with the Krugman and Kohneman's (1973) availability formalism.

In reflection on these results it is interesting to note that the first names we used as experimental stimuli are very closely related to the superordinate gender category; the names were selected because they unequivocally signal the gender of the naming individual. Despite the near identity of these gender identifiers (e.g., the routine, virtually unthinking inference that the same Paul refers to a man), the present results indicate that the respondents' set-size judgments were significantly influenced by the individual names they could recall. We find this result of note because it seems plausible to assume that availability effects might be even stronger in experiments with exemplars that were less closely related to their superordinates.

Set Size and Availability: Other Accounts

Alb et al. (1980) have suggested that across different events of estimation, people maintain distinct "occur@-rate information for the higher-order" (p. 320) superordinate category that is not implied. This theory would suggest that presentation of all names at Mars, Sears, and So forth, would automatically affect subjective estimates to the number of different female names on the study list, regardless of subject's ability to recall the originally presented names.

In support of this account, Alb et al. (1980) reported two experiments showing that their respondents' set-size estimates were reliably affected by the contents of the study list they had seen (e.g., the number of exemplars on the study list that were associated with different conceptual categories). These results were quite stable across a variety of presentation formats and setting circumstances, leading the authors to conclude that they probably reflected the operation of an automatic encoding process.

Experiment 1 in the Alb et al. article (1980) perhaps the most important for our present purposes. In this study, respondents made set-size estimates within a single format: some subjects were given lists as 3's per decision, under the assumption that the necessity for 3-decision judgments would make it unlikely that the subjects were "using the category name...to generate and count instances of the category that had occurred in the experimental context" (p. 323). Because subjects assigned to the 3-decision condition exhibited essentially the same pattern of results as subjects who were permitted substantially longer decision times, Alb et al. (p. 323) concluded that

Some internal representations of the category name were being tagged with frequency information during the presentation of exemplars (memory aids), and that this representation was available to us as a basis for estimating the size of the superordinate category.

Results like this may plausibly be interpreted as a challenge to the availability conception, but they are not decisive. Whereas fast pacing may make it difficult for a subject to conduct an exhaustive memory search for previously presented exemplars, an abbreviated memory search would surely be possible, even within Alb et al.'s (1980) shortest (5-decision interval). Recent research suggests a related possibility. The availability of a category can perhaps be meaningfully indexed by the ease with which its exemplars are retrieved (Schwarz & Bless, 1987), for example. In the Alb et al. experiment, for example, if the study list had included many exemplars of the category musical instruments, even respondents who were forced to make rapid category-size estimates would probably have found a simple master to generate recall or of more of the previously presented exemplars; they might thus use their experienced ease of retrieval as a cue that the category in question (musical instruments) was probably a large one.

Shede1 and Mathews (1985; Experiment 2) have reported other results that seem consistent with the view that availability affects set-size judgments. These investigators demonstrated that vividly prepared information was as unusually memorable as real information and different from their respondents' set-size estimates. In contrast with the results obtained in the present experiments, however, push analysis indicated that there was no significant relationship between availability and the respondents' set-size estimates. How can we explain these results in light of the present findings?
One possibility focuses on the unusual procedure that Shedler and Manis (1986) used to assess memorial availability (Experiment 2). Subjects in this study were provided with a series of male and female names and told the university affiliations of these individuals, i.e., whether they were students at Princeton or Stanford University. The information about some students was made especially vivid. That is, as some of the names were presented, the respondents simultaneously saw photographs that purportedly showed what these individuals looked like. To assess what subjects could remember from the study just as a measure of availability, each name was then re-presented, accompanied by the appropriate photograph in those cases where photographs had originally been shown. The respondents were to indicate whether the student in question was enrolled at Stanford or at Princeton. The results indicated that memory of the targets' university affiliations was significantly enhanced by the vividness manipulation (photographs). However, performance on the memory test was not related to the respondents' set-size judgments regarding such matters as the number of men on the study list who were enrolled at Stanford, the number of women enrolled at Stanford, the number of men at Princeton, and so forth.

The availability measure devised by Shedler and Manis (1986, Experiment 2) was rather different from the free-recall procedure that was used in the present studies. Most important, by presenting each subjects' name as a means of assessing memory the respondents were required to indicate the university affiliations of many people who would not otherwise have been considered (recalled); hence, the availability measure in the Shedler-Manis experiment probably included a good deal beyond the exemplar information that might have spontaneously come to mind when subjects made their set-size estimates.

Another aspect of the Shedler-Manis experiment (Shedler & Manis, 1986) that may be critical is that these investigators manipulated vividness to produce systematic differences in availability. The present experiments, on the other hand, focused on familiarity and frequency of presentation. The rather different data patterns that were observed in these studies may thus derive from the different independent variables that were involved. The Shedler-Manis results suggest that for reasons that are presently unclear, vividness manipulations may affect set-size estimates through an automatic mechanism that is largely independent of what comes to mind (i.e., independent of the availability heuristic). By contrast, in experiments that involve such independent variables as stimulus familiarity or frequency of presentation, the exemplars that are available in memory appear to be substantially more important.

Overall, these experiments support the view that a single recall (availability) often has a significant impact on set-size estimates. This does not mean that other determinants are of negligible importance, however. Indeed, the present experiments also suggest that there is probably a modest but consistent, direct (immediated) path linking set-size judgments with experimental manipulations as stimuli familiarity and frequency of presentation. Although only one of our set-size relationships produced statistically significant results with regression to the direct path (see Figure 2b), the consistent appearance of positive beta weights in all four of the relevant path analyses (see Figures 1, 2, and 4) suggests that a weak causal process is probably operating here, a process that is independent of the availability heuristic.8

Additional evidence concerning the less-than-perfect association between availability and set-size judgments may be found in an intriguing aspect of experiments by Watkins and Le-Compte (1991). These investigators noted that the accuracy of their respondents' category-size estimates could not be plausibly accounted for by the exemplars that they had recalled. Watkins and Le-Compte suggested, as a consequence, that their judges' set-size estimates must have depended on other information as well and speculated that this additional information was probably derived from general memory; the respondents' recall of a set of items as a set, as distinct from their episodic recall of individual exemplars, i.e., the information that was available.

Set-Size versus Frequency-of-Occurrence Judgments

The theory underlying the availability heuristic appeals to a plausible, common sense notion—the idea that judgments often depend on some extent on consciousness, on the things people have in mind. The present studies provide consistent support for this approach in the estimation of category size: by contrast, the availability heuristic does not appear to play an important role when subjects estimate frequency of occurrence.

Why should this be? Why should availability affect set-size judgments but not frequency estimates?

1. We normally regard the contents of memory as constituting a good approximation to the things that we have experienced. Under this assumption, the remembered exemplars from a certain category e.g., the individual female names that are recalled from a study list provide a good cue for estimating the overall size of the transpositive category (female names). After all, because the size of a category is, by definition, equal to the number of relevant exemplars, the cases that come to mind provide a plausible judgment guide, if our memory is reasonably accurate.

2. Now consider the frequency-estimation task. The availability of a given name and its prior frequency of presentation are not related as a matter of definition but instead depend on the experimental design: the fact that the repeated occurrence of a given stimulus event e.g., the repeated presentation of a name—is directly related to the probability that this event will be available (recallable) at some later point in time. As a consequence, other things being equal, a name (or other event) that we can recall is likely to have been seen more frequently than a name (or event) that we cannot recall. The present results are consistent with earlier research in suggesting that this principle is largely ignored when respondents are asked to estimate frequency of occurrence, because the occurrence estimates associated with a given event e.g., names were largely independent of the memorial availability of that event.

Implications for Social Cognition

These experiments provide consistent evidence that availability is measured by free recall plays a significant role in set-size

8 The p value of these combined results was less than .07, using a one-tailed sign test.
estimation. This suggests that the contents of consciousness (i.e., the experiences we can produce in memory) may be of substantial importance when we are making inferences about our own experiences. For example, when we ask people how many black cars they saw in the Medical School, how many women and so forth.

Several additional points are worth noting. First, people are not usually very good at everyday discourse. They have also been introduced into the social psychology laboratory. In Hamilton's (1981) work on illusion correlation, for example, respondents were usually presented with a series of written passages describing behaviors that had been described by the members of four groups. Group A and Group B. Follow-up interviews with subjects asked for set-size estimates as to the number of desireable versus undesirable behaviors performed by the members of the two groups. In one study a result that closely paralleled the availability findings reported here. Hamilton, Dugas, and Tietier (1985, Experiment 2) showed that a stereotype index derived from these set-size estimates was systematically related to respondents' recall of the information they had previously seen. In another different experiment context. Zarko (1972) repeatedly used set-size estimates to examine the effects of retrieval processing. After viewing an extended series of photographs subjects in this study had to name the number of black, white, man, and female targets they had seen. The results indicated that pictures of black men were generally encoded as blacks rather than as men, presumably because their color was most distinctive. In contrast with the results that are observed when responses are set for in-examine estimates. frequency of occurrence judgments do not appear to be mediated by the availability heuristic; these judgments are seemingly controlled by a more "verbatim" mechanism, a mechanism that is largely unaffected by intention or by mnemonic recall of Hasher & Zacks (1984). Frequency of occurrence judgments per se have not been prominent in social psychology. Closely related variables have, however, been examined in a number of experiments that focus on subjective similarity in a possible mediator of the effects of availability. (Miller & Zajonc, 1977; 1979; Watson, 1979) In one experiment, for example. Millard and Zajonc (1977, Experiment 2) showed 10 Japanese ideographs to their subjects; each ideograph was assigned either 0, 1, 2, 3, or 4 times. Respondents were then told the 10 ideographs that had previously been presented plus 10 others that had not been seen. They rated the subjective similarity of each stimulus on a 9-point scale. An undebatable measure of affect was also obtained for each of the Japanese ideographs by having the respondents draw a picture that resembled the ideograph. English ideographs were then rated from a set of 20 adjectives that varied in floridness; Anderson, 1983).

The results indicated that frequency of presentation had a significant impact on both subjective familiarity and on affect (floridity). That is, the frequency presented ideographs were thought to be more familiar, and they were also associated with more favorable adjectives. Multiple regression indicated, however, that familiarity was not necessary for the appearance of the subjective effect; that is, increased frequency of exposure was associated with a bias in positive affect even after the effect of subjective recognition (familiarity) was held constant quantitatively. Similarly, an analysis that was restricted to the stimuli that had been rated as unfamiliar (Millard and Zajonc, 1977) found a significant correlation between the number of times a stimulus had been shown and the effective rating it received. This evidence led Millard and Zajonc (1977) to conclude that subjective familiarity was not necessary for the generation of exposure effects, which tended to occur almost automatically. Our frequency-of-occurrence results imply an additional form of automatic invariance processing: an ideograph that seemed familiar in the studies by Millard and Zajonc (1977), i.e., an ideograph that appeared to have been presented several times) need not have been spontaneously available in recall memory.

Final Comments

We believe that the distinction between in-examine judgments and frequency of occurrence judgments may assist in clarifying the relationship among these variables in the context of both social and psychological judgment. Our results indicate that people's judgments of frequency of occurrence may be influenced by both automatic and deliberate retrieval processes. Further research might explore the possibility that the same people differ in other ways as well. For example, if in-examine judgments are more dependent on recall, they might perhaps be more easily disrupted by cognitive overload, by distraction, or by the introduction of negative cues as positive counterparts.

The authors thank Eugene Berkowitz for suggesting the relationship between memory and in-examine judgments and the Norand-Zajonc research on familiarity.

References


Hattie, J. & Burd, R. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or not. Psychological Review, 93, 338-344.


Received November 18, 1991
Revision received February 15, 1993
Accepted March 10, 1993

1994 APA Convention “Call for Programs”

The “Call for Programs” for the 1994 APA Annual convention appears in the September issue of the APA Monitor. The 1994 convention will be held in Los Angeles, California, from August 12 through August 16. The deadline for submission of program and presentation proposals is December 3, 1993. Additional copies of the “Call” are available from the APA Convention Office, effective in September. As a reminder, agreement to participate in the APA convention is now presumed to convey permission for the presentation to be audio taped if selected for soing. Any speaker or participant who does not wish her or his presentation to be audio taped must notify the person submitting the program either at the time the invitation is extended or before the December 3 deadline for proposal submission.