Journal of Personality and Social Psychology
1993, Vol. 65, No. 3, 448-457


0022-3514/93/$3.00

# Availability Heuristic in Judgments of Set Size and Frequency of Occurrence

Melvin Manis, Jonathan Shedler, John Jonides, and Thomas E. Nelson

The availability heuristic has been widely cited as an important factor in the judgment process. However, the evidence that availability is important in judging category size is not fully convincing. Moreover, several reports suggest that availability may not be a factor in judging frequency of occurrence. Path analysis was used in 3 experiments designed to assess the role of memorial availability in judgments of category size and frequency of occurrence. In judging set size, there was consistent support for the availability heuristic; that is, set size judgments were reliably influenced by the contents of memory. By contrast, in accordance with earlier results, availability was *not* a significant factor when Ss judged 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 "whenever someone evaluates the frequency of classes or the probability of events by the ease with which relevant instances come to mind" (p. 207). Subsequent researchers generalized this conception, invoking the availability notion to explain judgments in such diverse domains as legal decision making, attributions of causality, and the impact of vivid (vs. nonvivid) persuasive messages. (For reviews of the availability literature see Nisbett & Ross [1980] and Sherman & Corty [1984]; for an important limitation on the availability heuristic based on the distinction between "on-line" and "memory-based" judgments, see Hastie & Park [1986].)

Judgments are ubiquitous in everyday social life, and they have been studied extensively in social psychology laboratories. This article examines the role of availability in two types of judgment tasks: 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 feuding children had gotten into arguments during his absence. Note that whereas a frequency-of-occurrence estimate is invited here ("How often did x occur?"), the events in question differ noticeably from one another. As a consequence, the judge's task may be treated as a sort of set-size or category-size judgment (see below), in which the wife considers the category of children's arguments and estimates the number of relevant exemplars that occurred during her husband's absence.

In contrast with the case of Mr. Executive and his argumentative children, however, frequency estimates may involve the repeated occurrence of *unchanging events* (e.g., the repeated occurrence of a given word in a poem). Thus, in selecting new reading materials an elementary school teacher may rely on her subjective estimate as to the students' familiarity with problematic words (e.g., the word *bigot*), on the basis of their past reading assignments. Or, in a more political example, someone may estimate the number of times that Vice-Admiral John Poindexter used the phrase "I don't recall," during the televised Iran-contra hearings. What role does availability play in judgments like this?

Hasher 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 frequencies. Hasher and Zacks thus concluded that frequency-of-occurrence information is stored through an "implicit or automatic encoding process" (p. 1372) that is "fundamental" and "inevitable," a process that would presumably be independent of the availability heuristic.[1]

Melvin Manis and John Jonides, Department of Psychology, University of Michigan; Jonathan Shedler, Department of Psychological Services, Adelphi University; Thomas E. Nelson, Department of Political Science, Ohio State University.

This work was supported in part by grants to Melvin Manis from the Veterans Administration Office and from the Office of the Vice President for Research, University of Michigan. Experiment 1 was presented by Jonathan Shedler in partial fulfillment of the requirements for a doctoral dissertation at the University of Michigan.

Correspondence concerning this article should be addressed to Melvin Manis, Department of Psychology, University of Michigan, 580 Union Drive, Ann Arbor, Michigan 48109-1346.

[1] Some theorists have suggested that people may be equipped with cognitive counters that automatically code frequency of occurrence (Hasher & Zacks, 1984; Howell, 1973; Underwood, 1969). Direct coding models do not convincingly account for some aspects of the frequency-judgment literature, however (Hintzman, Nozawa, & Irmscher, 1982; Whitlow & Skaar, 1979). Other theorists have therefore rejected the possibility of a direct frequency code and have suggested some form of indirect coding in its place. In these models (see Hintzman, 1976; Howell, 1973, for discussions), frequency estimates are thought to be computed rather than stored directly. Computation is thought to depend on the number of memory traces or the strength of the traces associated with a given event. The controversy between di-

As evidence for the automaticity of frequency judgments Hasher and Zacks (1984) cited an important study by Howell (1973). In this experiment respondents were shown a list of words presented between 1 and 10 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 Naveh-Benjamin and Jonides, 1986).

Howell's study (1973) is particularly interesting 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. Howell's results did not follow this pattern, however. Indeed, they showed a modest reversal 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, Hasher, 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 from one list to 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-judgment 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, Shedler, Jonides, and Manis (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. If availability played an important role in frequency judgment, the words that a respondent could recall would be expected to receive 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 judgments of set size (or 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 conception suggests that in answering such a question one may be significantly influenced by the number of Jewish club members one can recall. However, Alba, Chromiak, Hasher, 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, for example, that if an individual is shown a study list that included the words *violin, chair, cow, oboe,* and *sofa,* 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.). Hasher and Zacks (1984) concurred in this view and hypothesized that information concerning both event frequency and category size is stored in memory by an implicit or automatic encoding process. In a similar vein, Watkins and LeCompte (1991) questioned the importance of exemplar recall (i.e., availability) in set-size estimates by showing that the accuracy of set-size judgments depends to a large extent on other, generic sources of information.

In contrast with the theorists just cited, the availability hypothesis (Tversky & Kahneman, 1973) suggests that judgments of set size may not be automatic and might instead depend on the respondent's capacity to bring relevant exemplars or associations to mind. Early demonstrations of this possibility were dramatic but not conclusive. For example, in Tversky and Kahneman's (1973) widely cited "famous names" experiment subjects 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 subjects recalled more male than female names from the study list (i.e., the male names were more available); moreover, other subjects who had heard the same study list erroneously believed that the list included more male than female names. For subjects in a contrasting condition, the female names were the 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 (fame). Demonstrations of this sort do not necessarily mean that memory has a causal impact on set-size

---

rect and indirect frequency codes has not been satisfactorily resolved; however, research continues to provide some support for the notion of a direct code for frequency estimates (Jonides & Jones, 1992).

judgments, however. It could well be that the experimental manipulation of fame produced two independent effects: (a) a memory (or availability) bias that favored the recall of the more famous names in the study list and (b) a separate judgmental bias, in which the category that included the more famous exemplars (e.g., the male names) 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 Durso (1986), who showed that recall and set-size estimates were both affected, perhaps independently, by (a) 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 (as 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.

*Stimuli.* 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 (e.g., George Bush, Lee Iaccoca, and Burt Reynolds vs. Laura Ashley, Jane Austen, and Francine Hughes); for the other list the female names were the more famous (e.g., Carol Burnett, Margaret Thatcher, and Barbara Walters vs. Claude Bolling, Charles Becham, and Michael York).

The procedure for selecting stimulus names was as follows: First, subjects drawn from the same population as the experimental subjects generated names 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 retained to use as foils 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 audiotape, in a single random order, at an approximate rate of 1 name every 2 s. Subjects assigned to Condition 1 heard the list with the more famous male names, and subjects assigned to Condition 2 heard the list with the more famous female names.

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

The memory test was in two parts. First, subjects were asked to write all of the names they could remember; they were permitted as much time to do this as they wished. Next, they completed a recognition test that contained the 40 list names and 80 foils 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 generally agreed on measure of availability has not emerged from the literature, we examined five alternative availability measures (a) *Free recall:* We scored the recall protocols for number of male and female names recalled. We defined the availability bias as the proportion of males among the names that each subject recalled (# of males/[# of males + # of females]). (b) *Spew 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 (a lower mean signified greater availability). Availability bias was defined as the difference between the two means (female names minus male names). (c) *Recognition:* The subjects rated each name for recognition using a 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 scores for male and female stimulus names. (e) $d'$: Finally, we scored the recognition ratings for hits, misses, false alarms, and correct rejections. These data were used to compute the $d'$ statistic of signal detection theory.

The experimental 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*

| Path | Free recall | Spew order | Recognition | Confidence | $d'$ |
|---|---|---|---|---|---|
| Condition 1 | 0.71 | 5.1 | 1.0 | 0.44 | 1.3 |
| Condition 2 | 0.39 | −2.7 | −0.7 | −0.36 | −1.1 |

males + estimated number females]). 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 proposes $M$ is the mean judgment bias across the subjects in each group. The difference between these judgment biases was highly significant; $t(71) = 5.3$, $p < .001$.

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 nonavailability 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 (Asher, 1983), 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 in this analysis 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.

Examining 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 .81), and availability in turn affected the respondents' judgment biases (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, *ns*).

### *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, spew order, or recognition. The availability pattern was not obtained, however, when availability was operationalized in terms of recognition confidence or $d'$. The path coefficients we obtained using the different availability measures are presented in Table 2.

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 (5–7 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,

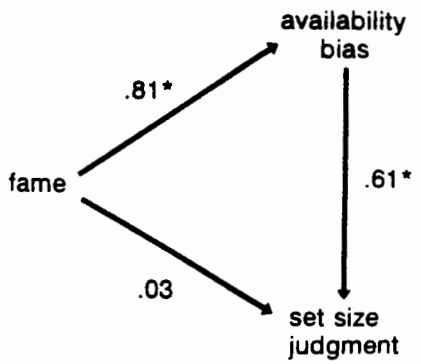


*Figure 1.* Path analytic results from Experiment 1. (*$p < .01$.)

Table 2
*Path Coefficient for Experiment 1*

| Path | Free recall | Spew order | Recognition | Confidence | $d'$ |
|---|---|---|---|---|---|
| Fame to availability | .81** | .76** | .76** | .62** | .74** |
| Fame to judgment | .03 | .21 | .25 | .41** | .38* |
| Availability to judgment | .61** | .41** | .35* | .21 | .20 |

* $p < .05$. ** $p < .01$.

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

### Method

Sixty-four 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 first names (e.g., George, Deborah, Sally, etc.), half of men and half of women. For one group of subjects each male name was shown five times, whereas each female name was shown twice; a given name was never presented twice in succession. In a contrasting list, the female names were shown five times each, and the male names were shown twice. Each name was exposed for a 5-s interval by means of an electronically controlled carousel 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 nonetheless to be counted as a single male exemplar. Finally, the respondents were presented with a recall test (as in Experiment 1); each respondent was simply instructed to "write down all the names (both male and female) that you can remember."

Subjects assigned to the delay condition followed this same set of procedures, except they completed an irrelevant classification task immediately after the name list had been presented. More concretely, 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 of each of the states . . . on a separate index card. Try to separate the states (index cards) into two categories. Please write the first classification that comes to mind, whether it seems obvious or not.
> Classification #1 _____ vs. _____.
>
> (2) Repeat question 1. Try to think of as many different ways of classifying (grouping) the states as you can.
> Classification #2 _____ vs. _____.

Subjects were instructed to provide up to seven classifications for the 50 states, a task that took about 5–7 min.

Following the state classification task, we asked the delay subjects for their set-size 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 bias was our measure of availability, as in Experiment 1 (i.e., proportion 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 2, where the different entries 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; $\beta = .75$ and .85, 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; $\beta = .49$ and .55.[2] Figure 2 also indicates that the direct (unmediated) effects of the different study lists were less critical as determinants of the respondents' judgment biases ($\beta = .23$ 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 seemingly enhanced importance of the direct path under delay conditions is of doubtful replicability.[3]

## Experiment 3

The preceding experiments present a consistent pattern: Set-size judgments were substantially affected by biases in the respondents' recall protocols, as suggested by the availability formulation (Tversky & Kahneman, 1973). By contrast, the literature on frequency-of-occurrence estimates has consistently failed to support the availability conception (see 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 women and 21 men) 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 afterward. They then saw a total of 60 first names, 30 of them of women and 30 of men. The male and female names were randomly divided into three sets, each containing 10 names. For subjects assigned to List 1, the different male names were shown either once, three times, or five times, whereas the female names were shown five times, seven times, or nine times. For subjects assigned to List 2, the male names were shown either five times, seven times, or nine times, whereas the female names were shown once, three

---

[2] For reasons that are presently obscure, Shedler (1987, Experiment 2) reported a nonsignificant path linking availability and set-size judgments in an experiment that is closely related to the present Experiment 2 ($\beta = .14$). This is, however, an aberrant result in our experience because it is contrary to the pattern that we repeatedly observed in the present work and in other experiments we have conducted on this issue.

[3] Readers may be surprised by the fact that a path coefficient of .49 (in Figure 2a) was significant at only the .05 level, whereas a path coefficient of .41 (in Figure 2b) was significant at $p < .01$. These results derived from differences between the standard errors of the coefficients in the two analyses; that is, the standard errors of the path coefficients in Figure 2a were larger than the standard errors for Figure 2b.

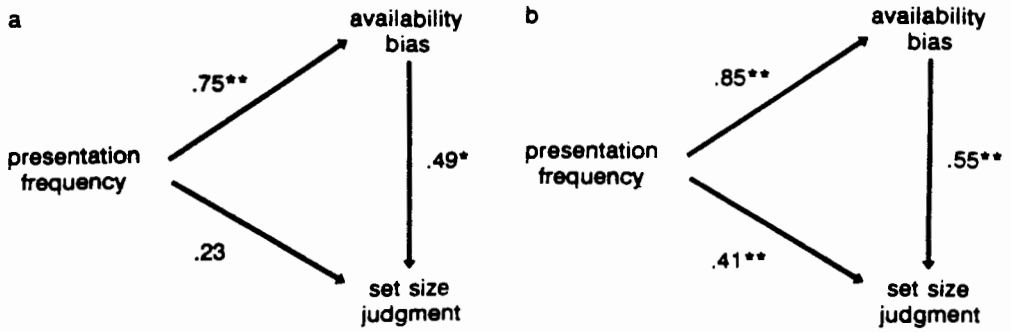


*Figure 2.* Path analytic results from Experiment 2. (Panel A derives from the no-delay condition; Panel B derives from the delay condition. $^{*}p < .05$. $^{**}p < .01$.)

times, or five 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 consecutive exposures. Each name was presented for a 4-s duration; there was a 1-s (blank screen) interval between the presentations of successive stimuli.

After the presentation of the study list 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]/all names recalled [male plus female]). Biases in the respondents' set-size judgments were similarly defined (i.e. [the number of male names in subjects' estimates]/[the number of male names plus female names in subjects' estimates]). 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 rated frequency of presentation for the different male names on the study list]/[average rated 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, among 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, then, 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, note that there was no reliable evidence of a direct study list effect, because the set-size judgments were independent of the condition (study list) to which a subject had been assigned ($\beta = .07$).

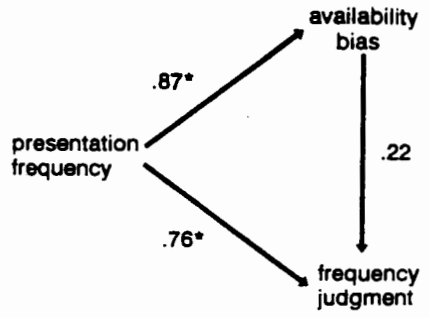


*Figure 3.* Path analytic results from Experiment 3: Frequency-of-occurrence estimates. ($^{*}p < .01$.)

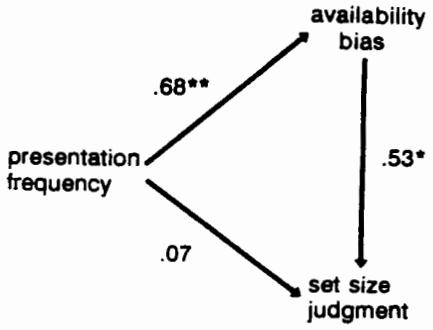


*Figure 4.* Path analytic results from Experiment 3: Set-size estimates. ($^{*}p < .05$. $^{**}p < .01$.)

These results are consistent with the pattern of results observed in Experiments 1 and 2; they 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 (as 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 or automatic encoding process (Hasher & Zacks, 1984; Shedler 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 (a) theories that posit a direct, internal counting mechanism that is independent of other memory processes (Hasher & Zacks, 1984; Underwood, 1983) and (b) theories that hypothesize a derived (or computed) basis for frequency estimates on the basis of the number and the strength of the relevant memory traces (Hintzman, 1976; Howell, 1973).

The set-size task yielded a rather different pattern of results. All three of our experiments showed that memorial 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 Tversky and Kahneman's (1973) availability formulation.

In reflecting on these results it is interesting to note that the first names we used as experimental stimuli are very closely related to their superordinate gender categories; the names were selected because they unequivocally signal the gender of the named individual. Despite the seeming automaticity of these gender inferences (e.g., the routine, virtually unthinking inference that the name 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 worthy 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*

Alba et al. (1980) have suggested that on exposure to different events or exemplars, people routinely abstract "occurrence-rate information for the higher-order" (p. 370), superordinate categories that are implied. This theory would suggest that presentation of such names as Mary, Beth, Sally, and so forth, would automatically affect subjects' estimates as to the number of different female names on the study list, regardless of subjects' ability to recall the originally presented names.

In support of this account Alba 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., by 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 testing circumstances, leading the authors to conclude that they probably reflected the operation of an automatic encoding process.

Experiment 1 in the Alba et al. article (1980) is perhaps the most important for our present purposes. In this study respondents made set-size decisions within a paced format; some subjects were given as little as 2 s per decision, under the assumption that the necessity for fast decisions 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. 373). Because subjects assigned to the 2-s decision condition yielded essentially the same pattern of results as subjects who were permitted substantially longer decision times, Alba et al. (p. 373) concluded that

> Some internal representation of the category name had been tagged with frequency information *during the presentation of exemplars* (emphasis added), and that this representation was available to the subject 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 Alba et al.'s (1980) shortest (2 s) 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 et al., 1991). In the Alba 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 it a simple matter to generate (recall) one or more of the previously presented exemplars; they might then use their experienced ease of retrieval as a cue that the category in question (musical instruments) was probably a large one.

Shedler and Manis (1986, Experiment 2) have reported other results that seem inconsistent with the view that availability affects set-size judgments. These investigators demonstrated that vividly presented information was (a) unusually memorable (available) and also (b) very influential as a determinant of their respondents' set-size judgments. In contrast with the results obtained in the present experiments, however, path 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 list (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 students' 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 exemplar 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 (unmediated) path linking set-size judgments to such experimental manipulations as stimulus familiarity and frequency of presentation. Although only one of our set-size experiments produced statistically significant results with respect to this 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.[4]

Additional evidence concerning the less-than-perfect association between availability and set-size judgments may be found in an intriguing series of experiments by Watkins and LeCompte (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 LeCompte 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 *generic 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 judgment often depends to 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 superordinate 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 an empirical regularity: 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., name) were largely independent of the memorial availability of that event.

### *Implications for Social Cognition*

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

[4] 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 exemplars we can produce in a memory search) may be of substantial importance when we informally attempt to answer such questions as How many Blacks were enrolled in the Medical School?, How many women?, and so forth.

Set-size estimates are not only involved in everyday discourse, they have also been introduced into the social psychology laboratory. In Hamilton's (1981) work on illusory correlation, for example, respondents were initially presented with a series of written passages describing behaviors that had been performed by the members of two groups: Group A and Group B. Following this presentation, subjects were asked for set-size estimates as to the number of desirable versus undesirable behaviors performed by the members of the two groups. In one study, in a result that roughly parallels the availability findings reported here, Hamilton, Dugan, and Trolier (1985, Experiment 2 and see Table 5) 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 a rather different experimental context, Zarate (1990) creatively used set-size estimates to explore selective information processing. After viewing an extended series of photographs, subjects in this study estimated the number of Black, White, male, 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 respondents are asked for set-size estimates, frequency-of-occurrence judgments do not appear to be mediated by the availability heuristic; these judgments are seemingly controlled by a more "primitive" mechanism, a mechanism that is largely unaffected by intention or by conscious recall (cf. 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 stimulus familiarity as a possible mediator of the exposure effect (Moreland & Zajonc, 1977, 1979; Wilson, 1979).[5] In one experiment, for example, Moreland and Zajonc (1977, Experiment 2) showed 10 Japanese ideographs to their subjects; each ideograph was exposed either 0, 1, 3, 9, or 27 times. Respondents were then shown the 10 ideographs that had previously been presented plus 10 others they had not seen before; they rated the subjective familiarity of each stimulus on a 7-point scale. An unobtrusive measure of affect was also obtained for each of the Japanese ideographs by having the respondents guess its meaning in English (choices were made from a set of 20 adjectives that varied in favorableness; Anderson, 1968).

The results indicated that frequency of presentation had a significant impact on both subjective familiarity and on affect (favorability). That is, the frequently 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 exposure effect; that is, increased frequency of exposure was associated with a gain in positive affect even after the effect of subjective recognition (familiarity) was held constant statistically. Similarly, in an analysis that was restricted to the stimuli that had been rated as unfamiliar, Moreland and Zajonc (1977) found a significant correlation between the number of times a stimulus had been shown and the affective rating it received.

This evidence led Moreland and Zajonc (1977) to conclude that subjective familiarity was not necessary for the generation of exposure effects, which seemed to occur almost automatically. Our frequency-of-occurrence results imply an additional form of automatic (nonconscious) processing: An ideograph that seemed familiar in the studies by Moreland 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 set-size judgments and frequency-of-occurrence judgments warrants further examination. The studies described herein suggest that these judgment variables are differentially sensitive to the contents of consciousness (availability). Future research might profitably examine the possibility that they differ in other ways as well. For example, if set-size judgments are more dependent on recall, they might perhaps be more easily disrupted by cognitive overload, by distraction, or by the introduction of negative (vs. positive) exemplars.

[5] The authors thank Eugene Burnstein for suggesting the relationships between our work on frequency of occurrence judgments and the Moreland–Zajonc research on familiarity.

## References

Alba, J. W., Chromiak, W., Hasher, L., & Attig, M. S. (1980). Automatic coding of category size information. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 370–378.

Anderson, N. R. (1968). Likeableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology, 9,* 270–272.

Asher, H. B. (1983). *Causal modeling* (2nd ed.). Beverly Hills, CA: Sage.

Hamilton, D. L. (1981). Illusory correlation as a basis for stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 115–144). Hillsdale, NJ: Erlbaum.

Hamilton, D. L., Dugan, P. M., & Trolier, T. K. (1985). The formation of stereotypic beliefs: Further evidence for distinctiveness-based illusory correlations. *Journal of Personality and Social Psychology, 48,* 5–17.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 12,* 1372–1388.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93,* 258–268.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 10, pp. 47–91). San Diego, CA: Academic Press.

Hintzman, D. L., Nozawa, G., & Irmscher, M. (1982). Frequency as a nonpropositional attribute of memory. *Journal of Verbal Learning and Verbal Behavior, 21,* 127–141.

Howell, W. C. (1973). Storage of events and event frequency: A comparison of two paradigms in memory. *Journal of Experimental Psychology, 98,* 260–263.

Jonides, J., & Jones, C. (1992). Direct coding for frequency of occurrence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 368–378.

Moreland, R. L., & Zajonc, R. B. (1977). Is stimulus recognition a necessary condition for the occurrence of exposure effects? *Journal of Personality and Social Psychology, 35*, 191–199.

Moreland, R. L., & Zajonc, R. B. (1979). Exposure effects may not depend on stimulus recognition. *Journal of Personality and Social Psychology, 37*, 1085–1089.

Naveh-Benjamin, M., & Jonides, J. (1986). On the automaticity of frequency coding: Effects of competing task load, encoding strategy, and intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 378–386.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*, 195–202.

Shedler, J. K. (1987). *Availability: Plausible but questionable*. Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Shedler, J. K., Jonides, J., & Manis, M. (1985). *Availability: Plausible but questionable*. Paper presented at the 26th annual meeting of the Psychonomic Society, Boston.

Shedler, J. K., & Manis, M. (1986). Can the availability heuristic explain vividness effects? *Journal of Personality and Social Psychology, 51*, 26–36.

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (pp. 189–286). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207–232.

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

Underwood, B. J. (1983). *Attributes of memory*. Glenview, IL: Scott, Foresman.

Watkins, M. J., & LeCompte, D. C. (1991). Inadequacy of recall as a basis for frequency knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 1161–1176.

Whitlow, J. W., Jr., & Skaar, E. (1979). The role of numerosity in judgments of overall frequency. *Journal of Experimental Psychology: Human Learning and Memory, 5*, 409–421.

Williams, K. W., & Durso, F. T. (1986). Judging category frequency: Automaticity or availability? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 387–396.

Wilson, W. R. (1979). Feeling more than we can know: Exposure effects without learning. *Journal of Personality and Social Psychology, 37*, 811–821.

Zacks, R. T., Hasher, L., & Sanft, H. (1982). Automatic encoding of events frequency: Further findings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 106–116.

Zarate, M. A. (1990). *Cultural normality and social perception*. Unpublished doctoral dissertation. Purdue University, West Lafayette, IN.

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 audiotaped if selected for taping. Any speaker or participant who does not wish his or her presentation to be audiotaped must notify the person submitting the program either at the time the invitation is extended or before the December 3 deadline for proposal submission.