The effect of increasing the memorability of category instances on estimates of category size

STEPHAN LEWANDOWSKY and PAUL W. SMITH
University of Toronto, Ontario, Canada

According to Tversky and Kahneman's (1973) availability heuristic, people sometimes make use of the ease with which instances are retrieved when they have to estimate proportions or frequencies. One implication of this availability heuristic is that any factor that affects memorability of instances from a category should also affect the estimated category size. In one of their experiments, Tversky and Kahneman found that, after being presented with a list of names, people judged the more famous names to be more frequent. Similarly, recall was found to be greater for the more famous names. Three experiments that used Tversky and Kahneman's paradigm are reported. Repeating nonfamous names resulted in their increased recallability and a corresponding increase in estimates of their frequency (Experiments 1 and 3). Making non-famous names more salient (Experiment 3) also had parallel effects on recallability and frequency estimates, indicating that different memory manipulations affected availability in a similar fashion. Furthermore, reliance on the heuristic was not changed as a function of prior knowledge (Experiment 2) or practice (Experiment 3).

Tversky and Kahneman (1973) suggested that people base their estimates of category size on "availability," that is, the ease with which instances of that particular category come to mind. Since its introduction, the heuristic has been invoked to explain many judgment biases, including the construction of stereotypes (Hamilton, 1981; Rothbart, Fulero, Jensen, Howard, & Birrell, 1978), egocentric biases in attributions of responsibility (Ross & Sicoly, 1979), and distortions in the estimations of the probability of lethal events (Lichtenstein, Slovic, Fischhoff, Layman, & Combs, (1978).

Despite such interest, there is great uncertainty about when and how the availability heuristic operates. To reduce the need for speculation, Hamilton (1981) has called for an analysis of what factors make information differentially available. Those studies that have actually looked at the memorability of information postulated to underlie the use of the heuristic present an unclear picture of how well memorability and category-size estimation match. In one of their original experiments, Tversky and Kahneman (1973) presented people with lists of names and asked them to judge whether there were more men's or women's names. They found that the sex for which the names were more famous was judged to be more frequent and also reflected better recall of names. Reyes, Thompson, and Bower (1980) found that the "vividness" of information led to greater recall of that information and to its correspondingly greater impact on a mock-jury decision on a defendant's guilt.

On the other hand, Beyth-Marom and Fischhoff (1977) asked for estimates of how many names of countries and kibbutzim began with various letters. As measures of availability, they used the number of instances recalled, the time taken to produce the first instance, and the number of instances produced in the first 5 sec. They found that only some of these measures of availability predicted frequency estimates, and then only some of the time. Alba, Chromiak, Hasher, and Attig (1980) varied the amount of time given to people to judge the number of words representing a given category from a list of implicitly categorized words. That there was no effect of time—time presumably used to generate category instances—suggested to Alba et al. that people were not relying on the availability of instances to make their judgments. It is thus not altogether clear that the heuristic actually operates when category size has to be estimated. In addition, even when a match between recall and judgment is found, other explanations could be advanced. Thus, Underwood, Zimmerman, and Freund (1971), for instance, found that recall of repeated words was better when the judged frequency of repetition was higher. They interpreted this relationship as being the product of a third factor: Rehearsal was postulated to lead to independent increases in judged frequency and recall. Although the frequency judged in their case was the number of repetitions rather than the number of category members, it demonstrates how, in a

situation analogous to that of Tversky and Kahneman (1973), a different causal interpretation can be made.

Much stronger support for the availability heuristic would be generated if manipulation of the memorability of category information resulted in corresponding changes in category-size estimation, regardless of the source of memorability changes. If mediation via recall is operating, then whatever affects the recall of instances should affect estimates of category size. Given that the evaluation of the availability heuristic is based on a comparison of two dependent measures, recall and category-size estimation, such source-independent effects would constitute the strongest evidence available.

EXPERIMENTS 1 AND 2

Using the famous-name paradigm from Tversky and Kahneman (1973), Experiment 1 looked at whether increasing the memorability of the less famous sex through repetition would increase the estimation of its proportion of the total number of names. Experiment 2 investigated the effect of prior knowledge on the task because, unlike in Experiment 1, in the original Tversky and Kahneman study subjects were not aware of their task prior to list presentation.

Method

Subjects. In Experiment 1, the subjects were 95 undergraduates at the University of Toronto. A different 45 undergraduates participated in Experiment 2. All subjects participated voluntarily in a classroom experiment.

Materials. The subjects saw two lists of names in Experiment 1. List 1 consisted of 20 famous women's names and 15 nonfamous mens' names, with the men's names being presented three times each. List 2 comprised 40 famous men's names and 30 nonfamous women's names, with no repetitions. The actual percentage of famous names in both lists was thus 57%. In Experiment 2, the subjects saw only List 2 (no repetitions). Items in both lists were randomly ordered. Famous names were drawn from movie stars, politicians, and other well-known personalities. Nonfamous names were drawn from the telephone directory.

Design and Procedure. In Experiment 1, each list was presented visually at 2 sec per item on overhead transparencies. The instructions were to study each list. After having viewed each list, the subjects judged the proportions of names of each sex. It was emphasized that the proportional judgment was to be of unique names in the list, that is, was not to include repetitions. The subjects were asked to recall men's names and then women's names, and then to score themselves. In Experiment 2, two sets of instructions were distributed randomly to members of the class. The twenty-six subjects in Group A were instructed to study the list and were told that there would be further instructions later. The nineteen subjects in Group B were instructed to study the list and were told that, afterwards, they would estimate the proportion of each sex in the list. Presentation parameters were the same as in Experiment 1.

Results and Discussion

Experiment 1. For List 2 (control), the mean judged percentage of men's names (the famous sex) was 64%. For List 1, including repeated nonfamous names, the mean judged percentage of women's names (the famous

sex) was reduced to 56%, which came very close to the true value of 57%. The difference between these mean judgments was highly reliable [t(93) = 4.50, p < .0001]. Consequently, it is implied that the judgment for the control list (no repetitions) was reliably different from the true value. Thus, the same effect of famousness found in the original Tversky and Kahneman (1973) experiment was obtained here.

The mean recall of famous names was 44% for List 1 and 40% for List 2. The mean recall of nonfamous names was 28% for List 1 and 7% for List 2. A recall ratio was calculated to represent recall performance in a form most comparable to estimations of the proportion of names of the famous sex (cf. Reyes et al., 1980). This score consists of the ratio of the number of famous names recalled to the total number of names recalled. The recall ratio was .611 for List 1 (with repetitions) and .851 for the control list [t(93) = 15.8, p < .0001]. Thus, the decrease in the recall ratio resulting from repetition of nonfamous names nicely paralleled the decrease in the judged proportion of names of the famous sex. Increasing memorability led to a corresponding increase in the estimated size of the category whose instances were made more memorable.

Experiment 2. Group A produced a mean estimate of 66% men's names (the famous sex). This was reliably different from the actual percentage [t(25) = 3.807, p < .05]. For the fully informed group (Group B), that percentage was 63%, which was also significantly different from the true value [t(18) = 2.25, p < .05]. Group A's estimate did not differ from Group B's [t(43) = .88]. In fact, only 2.2% of the variance in these results was accounted for by the difference in instructions between groups. It was thus concluded that the prior knowledge of the task had no effect upon proportional estimates, which ruled out strategies such as the subject's counting the items when he or she was aware of the task to follow.

EXPERIMENT 3

This experiment was designed to look at the effects of two different manipulations of memorability on proportional judgments, and to attempt to therefore expand the generality of the results obtained thus far. In addition, memorability was systematically varied over a wide range.

Method

Subjects. Thirty-six subjects participated voluntarily and were paid $4.00 each for a 1-h session.

Apparatus. The subjects were tested individually in an Industrial Acoustics Company sound-attenuating chamber. The entire experiment was controlled by an IBM PC microprocessor that displayed lists on a BMC KG 12C monitor and that scored responses directly from the keyboard.

Materials. There were six lists of names, each list consisting of 60 items, 30 drawn from each sex. For half the lists, the men's names were famous, and for the other half, the women's names were famous. The names were drawn from the same sources as in

Experiment 1. There were always an equal number of different men's and women's names, although, as described below, the total number of tokens for each sex sometimes differed.

Design. There were two memorability manipulations, repetition and salience. Nonfamous items were repeated three times for the medium- and five times for the high-repetition condition. In two no-manipulation control lists, no repetition took place, and each list consisted of 30 men's and 30 women's names. This was the composition of lists for the medium- and high-salience conditions as well; but for these conditions, presentation time for each item was increased and visual salience was enhanced.

The names were presented with variable interitem intervals to discourage counting. They were always presented for 1.55 sec, except in the medium- and high-salience conditions. In the medium-salience condition, the names of the nonfamous sex appeared for 3.75 sec. In the high-salience condition, the names of the nonfamous sex appeared for 5 sec, and background and foreground color was alternated every .62 sec. In these two lists, no repetition of names took place.

Each subject was exposed to each condition, and across each block of six subjects, every condition appeared once at each position in the series. Across the six blocks of six subjects each, each individual list occurred once at each position in every condition.

Procedure. The subjects were instructed that, after having viewed the list, they would be asked to recall the men's names and then the women's names, and then to judge the proportion of each sex in the list. For half the subjects, proportional judgment preceded recall, and for the other half, this order was reversed. They were told that there was great variation in the proportions and the lengths of the lists (although each list held 60 tokens and an equal number of types of each sex). It was again emphasized that the proportional judgment was of different names and not of total numbers of names including repetitions. Because famousness was balanced for sex, always recalling males first meant that the names of the famous sex were recalled first half the time. The same was true of the proportional judgments. After presentation of each list, the subjects counted backwards by 3s for 10 sec as a distractor task before proceeding to the test phase.

Scoring. Recalled names and judgments were typed by the subject on the keyboard. Correct recalls were those judged to have both the first and last names correct. Spelling was not important. The response had to be a homophone of the target name. Only one phoneme wrong in either the first or last name was allowed if the recall was to be counted. The scorer was blind to the conditions being scored.

Results and Discussion

Order of presentation of conditions had no effect (F < 1.0) on either the proportional judgment or the recall ratio. Analyses were then performed separately for each type of memory manipulation.

Repetition manipulation. The recall scores for names of the famous sex were 32% for the no-manipulation level and 42% for the medium- and 43% for the high-repetition conditions, respectively. For the names of the nonfamous sex, the corresponding scores were 6%, 24%, and 41%, respectively. When combined into recall ratios as described above, the values were .85, .65, and .52 for the no-manipulation, medium-repetition, and high-repetition conditions, respectively. The effect of the repetition condition on the recall ratio was highly reliable [F(2,70) = 46.1, p < .0001]. This significant difference supplies a check on the effectiveness of the manipulation and a score against which the proportional estimates can be compared (see Figure 1).
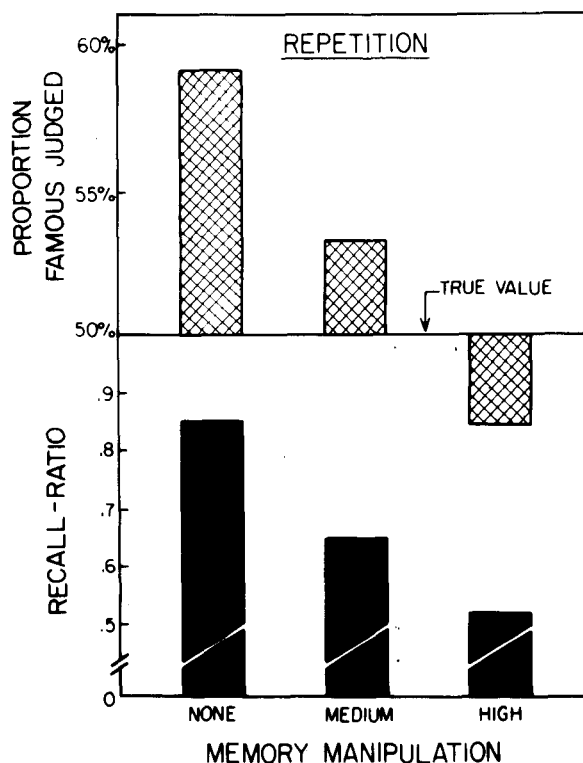

Figure 1. Mean estimated proportions of names of the famous sex and ratio of famous people recalled to total names recalled for Experiment 3, repetition conditions.

A significant difference in the proportional estimates for famous names [F(2,70) = 6.55, p < .01] was found across levels of memorability. Thus, the strong parallel indicated in Figure 1 was supported by the analysis of variance.

It is worth emphasizing how robust the familiarity effect on estimation was. Only by repeating the nonfamous names five times was it possible to drive the estimate of the size of that category above the estimate of the size of the famous-sex category (53% to 47%).

Salience manipulation. The effect of the salience manipulation on estimation was reliable [F(2,70) = 3.27, p < .05]. The actual recall for the famous sex was 32% for the no-manipulation condition and 34% for both the medium- and the high-salience conditions. For the nonfamous sex, the recall was 6% for the no-manipulation, 7% for the medium-salience, and 9% for the high-salience conditions, respectively. The recall ratios were .86, .83, and .79 for the no-manipulation, the medium-salience, and the high-salience conditions, respectively. This effect was less dramatic than that for the repetition manipulation, but was still reliable [F(2,70) = 4.12, p < .03]. Figure 2 shows memory performance and proportional judgments, this time for the salience manipulations.

Although it would be desirable to show the interdependence of these dependent variables (judgment and recall ratio) within each cell, most of these correlations were nonsignificant, presumably due to the restricted
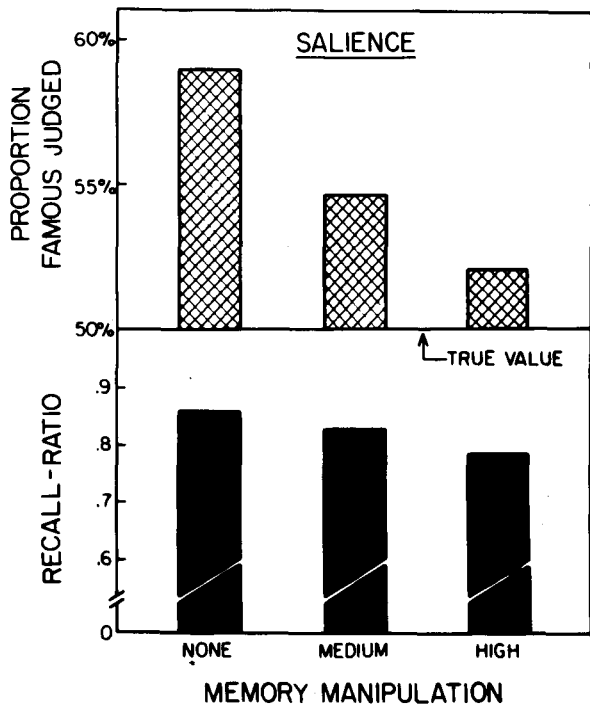
Figure 2. Mean estimated proportions of names of the famous sex and ratio of famous people recalled to total names recalled for Experiment 3, salience conditions.

range of both measures for a given condition. The overall correlation across conditions and all subjects, however, was highly reliable [$r(216) = .24$, $p < .0005$]. The magnitude of this correlation is comparable to that obtained by Reyes et al. (1980).

CONCLUSION

The mechanism underlying the availability heuristic in the present paradigm has been shown to be extremely robust. Neither prior knowledge of the task nor practice across six lists affected the subjects' behavior. Experiment 3 showed that, regardless of the source of memorability, judgments were paralleled by recall performance. This is the strongest evidence so far generated in support of the position that recallability determines category-size judgments.

The argument is still correlational, but seems to be tenable, especially because we demonstrated a dependence between recall and judgment for two ways of increasing memorability. Presumably, the present

methods of incrementing memory performance were highly unlikely to have any effects other than improving recall performance. This may not be true for other studies that have used less direct manipulations of memorability—for example, "vividness" (Reyes et al., 1980).

Furthermore, the likelihood that some third variable is causing the judgment behavior in the present study is substantially reduced, because it would have to be affected independently by both memory manipulations. Similarly, it becomes quite implausible that two different manipulations affected recall and judgment behavior independently, because that would again require that both our variables had the same nonmemorial consequences.

The most parsimonious conclusion is that memorability, regardless of its source, mediates the proportional judgments in this paradigm.

REFERENCES

ALBA, J. W., CHROMIAK, W., HASHER, L., & ATTIG, M. S. Automatic encoding of category information. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, **6**, 370-378.

BEYTH-MAROM, R., & FISCHHOFF, B. Direct measures of availability and judgments of category frequency. *Bulletin of the Psychonomic Society*, 1977, **9**, 236-238.

HAMILTON, D. Illusory correlation as a basis for stereotyping. In D. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior*. Hillsdale, N.J: Erlbaum, 1981.

LICHTENSTEIN, S., SLOVIC, P., FISCHHOFF, B., LAYMAN, M., & COMBS, B. Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, **4**, 551-578.

REYES, R. M., THOMPSON, W. C., & BOWER, G. H. Judgmental biases resulting from differing availabilities of arguments. *Journal of Personality and Social Psychology*, 1980, **39**, 2-12.

ROSS, M., & SICOLY, F. Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, 1979, **37**, 322-336.

ROTHBART, M., FULERO, S., JENSEN, C., HOWARD, J., & BIRRELL, P. From individual to group impressions: Availability heuristics in stereotype formation. *Journal of Experimental Social Psychology*, 1978, **14**, 237-255.

TVERSKY, A., & KAHNEMAN, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, **5**, 207-232.

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