Direct measures of availability and judgments of category frequency

RUTH BEYTH-MAROM
The Hebrew University, Jerusalem, Israel

and

BARUCH FISCHHOFF
Decision Research, 1201 Oak Street, Eugene, Oregon 97401

Tversky and Kahneman (1973) have proposed that people judge the numerosity of categories by the availability of category instances. The present study tests this hypothesis using several direct measures of availability for lexicographically defined categories (e.g., "How many countries are there beginning with the letter___ ?"). Although these availability measures were successful in predicting some estimates, including several judgmental errors, the role of availability in frequency estimation appeared to be quite complex. A number of possible confounding factors are discussed.

Tversky and Kahneman (1973) have proposed that people evaluate "the frequency of classes or the probability of events by availability, i.e., by the ease with which relevant instances come to mind" (p. 207). In a series of nine studies they showed that although availability is correlated with ecological frequency, it is also affected by other factors and can, therefore, lead to systematic biases.

In only two of their nine studies, however, was availability measured. Experiments 1 and 2 gave subjects 7 sec to estimate the total number of instances they would be able to produce from a particular category and then had them actually produce instances for a period of 2 min. Subjects' short-term estimates proved to be good predictors of their long-term production. In the remaining studies, plausible arguments were advanced for why particular manipulations should affect availability, which should, in turn, produce biased estimates. Although the predicted biases were observed, availability was not directly measured.

The present investigation explores the use of a number of direct measures of availability in hopes of attaining a better understanding of how the availability heuristic works.

METHOD

Thirty-six first-year Hebrew-speaking psychology majors participated in the present study as partial fulfillment of course requirements. The entire experiment was conducted in Hebrew. Each subject was tested individually in a soundproof room while seated at a table beside the experimenter. A cassette tape recorder rested on the table, with the attached microphone 20-30 cm away from the subject, and recorded subjects' responses. The following instructions were read by the experimenter, who throughout the experiment remained impassive:

"In the present experiment we are interested in people's ability to retrieve items from their memory. In some of the questions you will be asked to retrieve and list various items; the more you retrieve, the better.

"In the remainder of the questions you will be asked to give a numerical answer to questions whose precise answer you probably do not know. Therefore, we will ask you to give the numerical answer which seems to you most likely to be correct."

In the first part of the experiment, each subject performed two retrieval tasks, one dealing with names of countries and one with names of kibbutzim (communal settlements in Israel). The order of the two tasks was randomized over subjects. Instructions for those tasks were as follows:

"There are some 140 countries in the world [240 kibbutzim in Israel]; we will concentrate on countries [kibbutzim] whose names begin with one particular letter. We will give you one letter and you will be asked to list and simultaneously say aloud all the countries [kibbutzim] which you recall whose names begin with that letter. When you feel that you can remember no more, tell me. Is everything clear?

"Try to list all the names of the countries [kibbutzim] whose names begin with the letter _____ . START."

Subjects generally devoted about 2 or 3 min to each retrieval task. After giving up, they were asked whether they had remembered all of the possible instances. After replying "No" (as all subjects did), each was asked, "Of the 140 or so countries [240 or so kibbutzim], how many do you think begin with this same letter?"

Three initial letters for countries and three for kibbutzim were used. Letters were chosen so that the true number of members in the categories varied fairly widely. The letters used and the true number of members were, for countries, "Aleph" (19), "Kuf" (12), and "Yud" (5); and for kibbutzim, "Mem" (31), "Beit" (15), and "Hei" (8). One-third of the subjects (12) produced instances and such "aided" estimates for each of the categories. Following the two retrieval and aided (after production) estimation tasks, the recorder was turned off and subjects were instructed in the second, "unaided estimation" task.

"I will now allow you only 5 sec to estimate the number of countries or kibbutzim whose names begin with a certain letter. I will state the letter. Five seconds later, I will point to you indicating that your time is up and you will give me your estimate."

Each subject made unaided estimates for the two country

and two kibbutz letters for which he or she did not produce instances. Thus, 12 subjects made unaided estimates and the remaining 24 made aided estimates for each letter.

RESULTS

The following information was derived from each subject's responses: (1) unaided (5-sec) estimates of category size, (2) the time between the word "START" and production of the first instance, (3) the number of instances produced in the first 5 sec (the period allotted for the unaided estimation task), (4) the number of instances produced overall, and (5) postproduction (aided) estimates of category size. Summaries of these data appear in Table 1.

Countries

Comparing the top row of the table and row a reveals a substantial bias in unaided estimates of the relative sizes of the categories. In particular, the "Yud" category was judged to be considerably larger than "Kuf" ($z = 2.72$, Mann-Whitney U test), even though it is actually less than half the size. This reversal is, however, predicted by either measure of short-term availability (rows b and c). Subjects took half as long to produce the first instance of a "Yud" country and produced twice as many instances in the first 5 sec. The availability explanation of the bias in estimates seems even more plausible when the instances included in the two categories are considered. The five "Yud" instances (Israel, Jordan, Japan, Greece, Yugoslavia—all written in Hebrew with the same first letter) are much more familiar than the 12 "Kuf" instances (Cambodia, Cameroun, Canada, Columbia, Congo-Brazzaville, Congo-Kinshasa, Costa Rica, Cyprus, Cuba, Kenya, Korea, Qatar). However, although "Aleph" instances are less available than "Yud," by both short-term measures (rows b and c), "Aleph" is judged to be a more numerous category (row a). Clearly, availability is not the only determinant of judged numerosity.

Table 1
Production and Estimation Data Country and Kibbutz Categories

	Countries			Kibbutzim		
Letter	Aleph	Kuf	Yud	Mem	Beit	Hei
Actual number in category	19	12	5	31	15	8
a. Median unaided (5-sec) estimate of category size	15	7	11.7	29	23	10
b. Median time to produce first instance (in seconds)	1.9	2.9	1.5	4.5	5.7	5.8
c. Mean number of instances produced in first 5 sec	1.5	1.1	2.0	.5	.4	.6
d. Median number of instances produced	9.5	3.3	5	3.8	4	3
e. Median aided (after production) estimate	16	10	6.5	15	20	10

After exhausting their memories for category instances, subjects produced considerably more accurate estimates of category size (row e vs row a). This superior performance suggests that at least some of the bias induced by relying on the availability heuristic can be eliminated by using that heuristic properly, that is, by working hard to see how many instances are in memory. Note, however, that the order of the final estimates (row e) is not the order of the number of instances produced (row d). Again factors other than availability (at least as measured here) are at work.

Kibbutzim

Subjects' unaided estimates (row a) were generally quite good. Differences in category availability (rows b and c) were, however, so small here (none being statistically significant) that it is hard to imagine them serving as the basis of the discrimination shown in the estimates. In contrast with the country task, exhaustive listing of instances produced poorer estimates of category size (row e), suggesting that hard work alone is not sufficient for accurate estimation. Although the order of these final estimates was the same as that of the number of instances produced (row d), production differences were quite small.

DISCUSSION

The results indicate that where instances were easily available, with the country task, unaided estimates of category size were uncorrelated with actual size, but were correlated with two direct measures of availability: time to produce first instance and number of instances in the first 5 sec. In this situation, exhaustively listing instances in memory leads to improved estimation. The fact that aided estimates of country category size do not correlate completely with total production (rows d and e) indicates that there must be other factors influencing frequency judgments.

For the kibbutz task, instances were not easily available; half our subjects were unable to produce any examples at all within 5 sec. Here, short-term measures of availability (rows b and c) did not correlate with unaided estimates. The correlation between unaided estimates and actual category size in this situation must be taken as further evidence for the influence of factors other than ease of instance production on frequency estimates.

Further research is needed to identify these factors. We suspect that they will prove to be highly situation specific. Tversky and Kahnèman (1973) suggested that direct knowledge of category size (or item frequency) may be stored in memory, a proposal also advanced by Underwood (1969). This seems unlikely with the unnatural categories used in this experiment. A more likely factor is knowledge of the lexicographic structure of the language. For example, subjects may have known that "Aleph" is a much more popular initial letter than "Yud," a factor which could have tempered their reliance on availability information in the unaided task (rows a, b, and c). Similar use of analytical information might also explain the accuracy of the unaided estimates in the kibbutz task, for there are more Hebrew words beginning with "Mem" than with "Beit," and more with "Beit" than with "Hei."

Even unsuccessful attempts to produce instances may be informative. Some kibbutz-task subjects reported that the search process reminded them that there are many kibbutzim with the names of the form "Beit____" ("House of ____ ," "Beit"

being both the name of the letter and the word for "house") or of the form "Ha____" ("Ha" being the Hebrew word for "the"), even though they were not able to produce any specific instances. Quite possibly, "Kuf" subjects in the country task increased their frequency estimates following the production of relatively few instances because of a feeling that there were quite a few obscure countries "in there" but unretrievable; "Yud" subjects in the country task may have lowered their aided estimates because of the opposite feeling. Such information is reminiscent of that reported in "tip of the tongue" experiments (Brown & McNeill, 1966). If this sort of information is to be considered an aspect of availability, then that term should be expanded to include the "ease with which instances almost come to mind."

One subject for whom the instance-producing strategy of frequency estimation did seem relevant asked for an extension of the 5-sec estimation period, begging, "Just wait until I can think of at least one."

REFERENCES

BROWN, R., & McNEILL, D. The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 325-337.

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

UNDERWOOD, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559-573.