Memory & Cognition
2006, 34 (1), 28-40

The effect of category focus at encoding on category frequency estimation strategies

MARIO PANDELAERE and VERA HOORENS
Katholieke Universiteit Leuven, Leuven, Belgium

We investigated whether category focus at encoding affects how people estimate category frequencies. Participants in three experiments viewed items of various categories. They estimated category frequencies after categorizing them into relevant versus irrelevant categories (Experiments 1–2) or after categorizing versus memorizing them (Experiment 3). Verbal protocols (Experiments 2A and 2B), response latencies (Experiments 2A and 2B), frequency estimate changes (Experiment 2B), and the relationships between objective and estimated category frequencies and instance recall (Experiments 1–3) showed that the participants mainly used availability to estimate category frequencies after memorizing instances (Experiment 3) or after categorizing them into irrelevant categories (Experiments 1–2). After categorizing items into relevant categories, the participants relied more often on stored category frequency information (Experiments 1–3).

People often need to estimate event or category frequencies. For instance, they may wonder how many parking tickets they have had lately or how many mammals they have seen in a certain zoo. In the present research, we investigated how people estimate category frequencies and, more precisely, whether *a focus on categorical information* while instances are encoded determines how people estimate *category frequencies*.

Category frequencies or category (set) sizes describe how many instances of categories have occurred or how many instances of certain categories that have occurred fall into a well-defined set. If someone wonders how many mammals he or she has seen in a certain zoo, he or she basically estimates the frequency of the category *mammals* and, more specifically, *mammals in zoo X*. Category frequencies differ from event frequencies in that the latter refer to the number of times a stimulus has occurred (e.g., Betsch, Siebler, Marz, Hormuth, & Dickenberger, 1999; Hanson & Hirst, 1988; Manis, Shedler, Jonides, & Nelson, 1993). When estimating the number of parking tickets one has had, one is estimating an event frequency.

Clearly, people can estimate event frequencies (e.g., Brown, 1995, 1997) or behavioral frequencies (Conrad, Brown, & Cashman, 1998; Menon, 1993) in various manners. Tversky and Kahneman (1973) proposed that people estimate category frequencies by using instance availability. One can distinguish between availability by ease and availability by number. Availability by ease refers to the estimation of frequencies on the basis of *how easy* it is to retrieve (or generate) instances of the category. Availability by number refers to frequency estimation on the basis of *how many* instances can be retrieved. As such, it implies counting retrieved instances and extrapolating this count to a frequency estimate (see Watkins & LeCompte, 1991).

Ease of retrieval and the number of retrieved instances are usually strongly related (but see, e.g., Schwarz et al., 1991). In fact, ease of retrieval has often been operationalized as the number of instances retrieved (e.g., Curt & Zechmeister, 1984; Lewandowsky & Smith, 1983; Tversky & Kahneman, 1973; Williams & Durso, 1986). Manis et al. (1993) even argued that it is the *best* operationalization. Therefore, estimating category frequencies through availability is supposed to create a strong correlation between the estimates and the number of instances recalled (e.g., Betsch et al., 1999; Bruce, Hockley, & Craik, 1991; Manis et al., 1993; Watkins & LeCompte, 1991). Moreover, the correlation should reflect a *direct*, rather than an indirect, relationship (e.g., Betsch et al., 1999; Bruce et al., 1991; Maley, Hunt, & Parr, 2000; Manis et al., 1993).

Many studies support the idea that people estimate category frequencies on the basis of availability. In their classic *famous people* experiment, Tversky and Kahneman (1973) read a list of 39 men's and women's names to their participants. Either the men or the women were famous. However, the list contained fewer names of the *famous gender* (19 names) than of the *nonfamous gender* (20 names). Famous names were remembered better than nonfamous ones. At the same time, the category frequency of the famous gender was estimated to be higher than the category frequency of the nonfamous gender. Tversky and Kahneman argued that the frequency estimates were derived from the names' availability. Further supporting the availability view, factors as diverse as instance salience and repetition (Lewandowsky & Smith, 1983), serial position

The authors thank Neil Bearden, Michael Kahana, Frank Van Overwalle, Luk Warlop, and two anonymous reviewers for their useful and constructive comments on an earlier draft of the manuscript. Correspondence concerning this article should be sent to M. Pandelaere, School for Mass Communication Research, E. Van Evenstraat 2, 3000 Leuven, Belgium (e-mail: mario.pandelaere@soc.kuleuven.be).

(Curt & Zechmeister, 1984), encoding time and instance typicality (Williams & Durso, 1986), and encoding effort, dispersion of instances from the same category across the list, and extra-list cuing (Greene, 1989) affect instance recall and category frequency estimates similarly.

Some studies have demonstrated a *direct* relation between instance recall and category frequency estimates (e.g., Curt & Zechmeister, 1984; Manis et al., 1993). For instance, using the famous people paradigm, Manis et al. found a positive relation between the number of names recalled of a gender and the category frequency estimate for that gender, even if the effect of fame was controlled for statistically. This implies that the number of names recalled and category frequency estimates were related directly (rather than spuriously, through a joint dependency on fame).

Despite this evidence, some studies suggest that people may *not* base category frequency estimates on instance availability. First, Betsch et al. (1999, Experiments 5 and 6) and Maley et al. (2000) found an *indirect* relation, at best, between category frequency estimates and the number of instances recalled. Second, Watkins and LeCompte (1991) found that the average number of instances recalled deviated from objective category frequencies more strongly than did the average estimated category frequencies. Third, Alba, Chromiak, Hasher, and Attig (1980) and Brooks (1985) found that people estimate category frequencies quickly and rather accurately. They could do so even if they were not forewarned to pay attention to category frequency while encoding the instances. According to the authors, this suggests that people base category frequency estimates on automatically stored category frequency information, rather than on availability.

One solution for the inconsistency is that people store and update category frequency information automatically only under certain conditions. Freund and Hasher (1989; see also Alba et al., 1980; Barsalou & Ross, 1986; Curt & Zechmeister, 1984) suggested that people accrue category frequency information automatically for any category that is activated while they encode the category instances. In that case, they do not need to rely on availability to estimate category frequencies (although they may choose to do so). For categories that are not activated while they encode the instances, people do not store category frequency information in memory. They *necessarily* use availability to estimate category frequencies. The accuracy of their estimates depends on the circumstantial factors that render some instances more available than others, possibly distorting their estimates accordingly.

Unfortunately, the evidence for the latter view is either indirect or difficult to interpret. If one focuses on indirect evidence first, several studies (Barsalou & Ross, 1986; Freund & Hasher, 1989; Hanson & Hirst, 1988) have shown that category frequency estimates are more *accurate* if the relevant categories have been activated at the encoding of the instances than if they have not. This has been taken to suggest that if the relevant category is activated at encoding, people base category estimates on stored frequency information. However, this inference does not necessarily follow. Activating a relevant category at encoding may simply enhance the effectiveness of the availability heuristic (see Beyth-Marom & Fischhoff, 1977) because it enhances the number of instances recalled (Conrad, Brown, & Dashen, 2003; Epstein, Dupree, & Gronikowski, 1979).[1] Since people tend to underestimate category frequencies, enhancing the number of instances recalled leads to higher and, hence, more accurate category frequency estimates (in an absolute sense).

To our knowledge, only one study that is relevant to our present question *directly* investigated the effect of categorization at encoding. Bruce et al. (1991, Experiment 1) showed their participants instances from 30 semantic categories. The participants read them aloud or classified them into one of the categories. Afterward, they estimated category frequencies and recalled as many instances as possible. To assess whether the participants used the availability heuristic, Bruce et al. calculated the partial correlation between the estimated category frequency and the number of instances that were recalled for each category, controlling for objective category frequencies. A positive partial correlation would indicate a direct relationship between recall and category frequency estimates and would suggest that the participants used an availability strategy. Bruce et al. indeed found a positive partial correlation in the reading condition. In the categorization condition, such a correlation occurred only if the participants recalled instances *before* estimating category frequencies. It did not occur if they recalled instances *after* estimating category frequencies. This suggests that people derive category frequency estimates from the availability of the instances, unless they have categorized the instances into relevant categories *and* are not encouraged to recall instances before giving their estimates.

However, the between-subjects correlations reported by Bruce et al. (1991) confound within-subjects and between-subjects variability (see Michela, 1990). Within each participant, recall and frequency estimate should be at least monotonically related if the participant uses the availability heuristic, even after controlling for objective frequency. However, two individuals who use the availability heuristic may differ with respect to the extrapolation rules they use (see Watkins & LeCompte, 1991). Consequently, two people who use the availability heuristic and who recall the same number of instances may extrapolate to different estimates (for individual differences in extrapolation rules, see Kverno, 2000). To the extent that substantial between-subjects variability occurs—as is usually the case in frequency estimation—this may lead to between-subjects correlations that drastically diverge from the more relevant within-subjects correlations. Moreover, it is not a priori clear whether between-subjects correlations overestimate or underestimate within-subjects correlations.

At first sight, a second study bears directly on our question. Conrad et al. (2003) examined how people estimate frequencies of unnatural categories (i.e., categories into which people do not generally categorize objects). They presented their participants with objects that belonged to various property categories (such as *colors* or *smell*). The instances were presented either alone (implicit property condition) or below a property category label (explicit property condition). The participants were asked to estimate the frequencies of the categories. In Experiment 1, they were asked to think aloud while doing so. In Experiment 2, their response latencies were observed.

The participants in the explicit property condition mentioned relying on an enumeration-based strategy less often and relying on a general frequency impression more often than did the participants in the implicit property condition. In both conditions, however, enumeration-based strategies were used most frequently (Experiment 1). The participants in the explicit property condition were generally faster than the participants in the implicit property condition. In the former condition, but not in the latter, response latencies increased as the objective category frequency increased (Experiment 2). The participants in the explicit property condition were also more accurate than the participants in the implicit property condition (Experiments 1 and 2). The authors suggested that a relatively easy process of on-target enumeration occurred in the explicit property condition. In the implicit property condition, a more difficult enumeration process occurred that yielded many off-target instances. On the surface, this implies that categorizing instances at encoding elicits on-target enumeration. The explicit versus implicit property manipulation does indeed seem to parallel categorizing versus not categorizing at encoding.

However, being provided with property labels is not necessarily the same as actually engaging in categorization. In particular, because Conrad et al.'s (2003) participants were told to study the word carefully, at least some of them may have concentrated on the instance information and paid little attention to the property labels. The limited attention they paid to the property information may have facilitated subsequent recall without changing their estimation strategy (although it may have changed its efficiency). This may have led to a high proportion of enumeration-based strategies (Experiment 1) and a clear-cut relationship between the number of instances to be recalled and response latencies (Experiment 2). Given the absence of a process measure that shows whether (and how many of) the participants categorized the instances (or associated the property with the instances), it is not clear whether the findings are really due to full-blown categorization.

In addition, Conrad et al.'s (2003) taxonomy does not distinguish between retrieving stored frequency information and availability by ease as separate estimation strategies. Indeed, both of these strategies are included in the category of *general impressions*. On the other hand, availability by number is split up into *enumeration* and *enumeration + adjustment*. In addition, besides mere guesses, the category of *unjustified answers* may include the retrieval of stored frequency information. Indeed, a think-aloud procedure may not yield any justification if people merely retrieve stored frequency information. To summarize, Conrad et al.'s (2003) taxonomy cannot be mapped onto the distinction between availability (by number or by ease) and the retrieval of stored frequency information.

To summarize, several studies are compatible with the hypothesis that categorizing stimuli into a relevant category influences the mechanism through which people estimate category frequencies. However, some of them support this hypothesis only indirectly (e.g., Freund & Hasher, 1989; Hanson & Hirst, 1988). Others are difficult to interpret (Bruce et al., 1991) or were not designed to examine the present question (Conrad et al., 2003). Therefore, the aim of the present experiments was to examine how category activation at encoding determines how people estimate category frequencies. We hypothesized that category activation at encoding *enables* people to estimate category frequencies without using the availability heuristic (i.e., by relying on stored category frequency information). If the relevant category is not activated during encoding, people *necessarily resort to* the availability heuristic.

In Experiments 1A and 1B, the participants' estimation strategies were inferred from the pattern of correlations between objective category frequencies, frequency estimates, and instance recall. Experiments 2A and 2B corroborated Experiment 1 with various process measures. Experiment 3 showed that the findings of Experiments 1 and 2 could also be obtained using the same famous people paradigm that typically supports the use of an availability strategy in category frequency estimation tasks.

EXPERIMENTS 1A AND 1B

Experiment 1 tested the hypothesis that people who categorize stimuli into relevant categories base their subsequent category frequency estimates on stored category frequency, whereas people who categorize stimuli into irrelevant categories use availability. The participants viewed a list of nouns referring to objects that typically have one color out of a limited set of colors (with different numbers per color) and that are or are not suited for human consumption (half suited and half unsuited). The participants categorized the items according to either color (relevant categorization) or suitability for human consumption (irrelevant categorization). After the presentation phase, the participants gave relative (Experiment 1A) or absolute (Experiment 1B) frequency estimates for the various color categories. Finally, they wrote down as many items as possible.

For each participant, we inferred whether he or she had based his or her category frequency estimates on stored category frequency information or on availability. We did so by calculating and comparing the rank order correlation between his or her objective and estimated category frequencies and the rank order correlation between the number of items he or she recalled and his or her estimated category frequencies.[2] If a participant used availability, any correlation between the objective and the estimated frequencies should be due to the joint effects of the objective frequencies on the number of instances recalled and of the number of instances recalled on the estimated frequencies. This implies that the correlation between instance recall and estimated category frequencies should be greater than or equal to the correlation between objective and estimated category frequencies. If a participant derived category frequencies from stored category frequency information, any correlation between instance recall and estimated category frequencies should be due to the joint effects of the objective frequencies on the estimated frequencies and of the objective frequencies on the number of instances recalled. Consequently, the correlation between recall and estimated category frequencies should be smaller than or equal to the correlation between the objective and the estimated category frequencies.

We chose to use color categories, rather than semantic categories, because objects sometimes automatically activate the semantic category to which they belong (Nelson, Fehling, & Moore-Glascock, 1979; Warren, 1972; Wickens, 1970), whereas colored objects usually activate the corresponding color category to a lesser extent (Wickens, 1970) or not at all (Underwood, 1965). To examine whether relevant categorization reduces the use of the availability heuristic, we needed to minimize the possibility of automatic categorization into the relevant category.

Method

Participants. Seventy-two students (29 of them male, 43 female; mean age = 21.3 years) from the economics department of the University of Leuven volunteered to participate in Experiment 1A. In Experiment 1B, 75 first-year and second-year psychology students (11 of them male, 64 female; mean age = 19 years) participated as a course requirement.

Materials. We selected six color categories (yellow, green, red, white, black, and brown) for which we were able to generate a sufficient number of instances. Forty students who did not participate in the main experiment were presented with 120 items that typically have one of these colors. They wrote down the color next to each item. An item was selected if at least 32 students (80%) wrote down the intended color. Given that an insufficient number of black and brown items could be selected, we combined them into the category brown/black. Consequently, the actual experiment comprised five categories. The items can be found in the Appendix.

We constructed five list types according to a randomly selected 5 × 5 Latin square in such a way that the five color categories were paired with one of five objective frequencies (2, 4, 6, 8, or 10 instances). Within list types, each color category was paired with a different frequency. Across list types, each color category was paired with each frequency only once.

Procedure. The participants arrived at the lab in groups of 3–6 and were seated at computers. The instructions on the screen explained that they would see a number of words that they had to classify. The participants in the *relevant* categorization condition read that they had to do so according to the typical color of the objects the words referred to. The participants in the *irrelevant* categorization condition read that they had to classify the objects according to suitability for human consumption.[3] For each participant in the irrelevant categorization condition, the computer selected one list type. Suitable items were drawn randomly, so that exactly half of each color category referred to something suited for human consumption. Each participant in the relevant categorization condition received the same list as a participant in the irrelevant categorization condition.

In Experiment 1A, the participants responded by pressing a key that was covered by a patch of the corresponding color (relevant categorization) or by pressing either J or N (irrelevant categorization; J for suited, N for unsuited). In Experiment 1B, the participants responded by clicking the mouse on one of either five or two buttons that appeared on the computer screen. The buttons were labeled with the color names (relevant categorization) or with *suited for human consumption* and *not suited for human consumption* (irrelevant categorization).

After the categorization task, the participants were instructed that they had seen words that referred to objects of five colors, which were summed up. In Experiment 1A, they rank ordered the colors by typing them in in decreasing category frequency order. In Experiment 1B, they entered absolute category frequency estimates for each color. Finally, all the participants wrote down all the items that they recalled for each color category.

Results

One participant in Experiment 1A was excluded because he forgot a category in the rank-ordering task. For each of the other participants, we calculated pairwise Kendall correlations between objective category frequencies, number of instances recalled, and estimated category frequencies (see Table 1). We inferred that a participant had used availability if the correlation between instance recall and estimated category frequency was greater than the correlation between objective and estimated category frequencies. We inferred that a participant had retrieved

Table 1
Percentages for Inferred Estimation Strategy and Self-Mentioned Estimation Strategy, as a Function of Categorization Relevance

	Experiment 1A		Experiment 1B		Experiment 2A		Experiment 2B	
	Irrelevant (n = 36)	Relevant (n = 35)	Irrelevant (n = 39)	Relevant (n = 36)	Irrelevant (n = 83)	Relevant (n = 76)	Irrelevant (n = 51)	Relevant (n = 52)
Inferred estimation strategy								
Availability	72.2	26.5***	56.8	34.3†	50.6	35.6†	41.7	27.5
Stored frequency information	16.7	73.5***	35.1	62.9*	34.6	63.0***	27.1	68.6***
Guessing	11.1	0*	8.1	2.9	14.8	1.4**	31.2	3.9***
Number of participants for whom no inference was made	0	1	2	1	2	3	3	1
Self-mentioned estimation strategy								
Availability	–	–	–	–	64.8	19.7***	71.7	11.8***
Stored frequency information	–	–	–	–	15.5	40.9***	6.5	68.6***
Availability and stored frequency information	–	–	–	–	11.3	38.0***	0	17.6***
Guessing	–	–	–	–	8.4	1.4*	21.7	2.0***
Number of protocols that could not be coded	–	–	–	–	12	5	6	1

Note—Percentages are calculated excluding participants that could not be classified or whose protocol could not be coded. Superscripts indicate significantly different percentages within a row and experiment. †$p < .06$. *$p < .05$. **$p < .01$. ***$p < .001$.

stored category frequency information if the correlation between instance recall and estimated category frequency was smaller than the correlation between objective and estimated category frequencies. One participant in Experiment 1A and 3 in Experiment 1B could not be classified, because the relevant correlations were identical. We inferred that the participants had simply guessed (1) if the correlations were extremely low (below .25) or (2) if their frequency estimates were identical for all the categories, even though they had recalled an unequal number of items. Eight participants (4 per experiment) were thus classified as *merely guessing*. For the participants who could be classified, the mean absolute difference between the two relevant correlations was .26 in Experiment 1A and .20 in Experiment 1B. The resulting frequencies are displayed in Table 1.

A chi-square test revealed an association between categorization relevance and inferred estimation strategy [availability, stored category frequency information, or guessing; Experiment 1A, $\chi^2(2, N = 70) = 26.62$, $p < .001$; Experiment 1B, $\chi^2(2, N = 72) = 5.82$, $p < .055$; see Table 1]. More participants relied on availability or guessed in the irrelevant categorization condition than in the relevant categorization condition. In contrast, more participants relied on stored frequency information in the relevant categorization condition than in the irrelevant categorization condition.

Interestingly, the mean rank order correlation between objective and estimated category frequencies was higher after relevant than after irrelevant categorization, suggesting that the participants in the relevant categorization condition were more accurate from a *relative* point of view [Experiment 1A, $t(69) = 5.75$, $p < .001$; Experiment 1B, $t(72) = 3.19$, $p < .01$]. The other two correlations did not differ between conditions [Experiment 1A, both ts < 1.11, both ps $> .27$; Experiment 1B, both ts < 1.65, both ps $> .10$].

Discussion

In Experiment 1, we investigated whether categorization relevance at encoding determines how people estimate category frequencies. On the basis of the correlations between objective category frequencies, category frequency estimates, and recall, we classified the participants as estimating category frequencies on the basis of availability or stored category frequency information. The participants who had categorized the stimuli into categories that were relevant to the frequency estimation task tended to rely on stored category frequency information, rather than on availability. In contrast, the participants who had categorized the stimuli into irrelevant categories tended to rely on availability (or guessing), rather than on stored frequency information. These findings suggest that category focus relevance affects how relative (Experiment 1A) and absolute (Experiment 1B) category frequencies are estimated.

Because we classified participants for whom the relevant correlation coefficients differed even to a small degree, some participants may have been misclassified.[4] To check for this possibility, we reanalyzed the data after eliminating the participants for whom the difference was below .10 (Experiment 1A, $n = 12$; Experiment 1B, $n = 26$) and, in a second wave, all the participants for whom it was below .20 (Experiment 1A, additional $n = 22$; Experiment 1B, additional $n = 21$). The association between categorization criterion and inferred strategy increased slightly as the elimination threshold increased, showing that our findings are robust against misclassification.

Category frequency estimates appear to be more accurate (at least in a relative sense) after a relevant categorization of instances than after an irrelevant categorization. Indeed, the average within-subjects rank order correlation between objective category frequencies and estimated category frequencies was higher in the former than in the latter condition. This is consistent with the finding that encoding into a relevant category yields more accurate category frequency estimates than does encoding into an irrelevant category (see Hanson & Hirst, 1988) or not categorizing at all (Bruce et al., 1991; see also Barsalou & Ross, 1986; Freund & Hasher, 1989). An additional analysis revealed that the average rank order correlation between objective category frequencies and estimated category frequencies was higher for the participants who had relied on stored frequency information than for the participants who had relied on availability [Experiment 1A, $M = .77$ vs. $M = .41$, $t(68) = 5.65$, $p < .001$; Experiment 1B, $M = .80$ vs. $M = .57$, $t(66) = 3.84$, $p < .001$]. This finding supports the common assumption that relying on stored category frequency leads to more accurate category frequency estimates than does relying on information availability. It also supports the practice of inferring estimation strategy differences from accuracy differences (e.g., Barsalou & Ross, 1986; Freund & Hasher, 1989).

In Experiment 1, we inferred how participants estimated category frequencies from the pattern of correlations between objective and estimated frequencies and instance recall. In Experiment 2, we further examined how processing goals at encoding affect how people estimate category frequencies by including additional process measures.

EXPERIMENTS 2A AND 2B

By including various process measures in Experiment 2, we aimed at testing more comprehensively whether processing goals at encoding affect how people estimate category frequencies. We also examined whether category frequency estimates that are based on availability are derived from availability by number or availability by ease. Both aims were achieved by replicating Experiment 1. Given that Experiments 1A and 1B yielded comparable results, Experiment 2 focused on absolute estimates.[5]

Both Experiments 2A and 2B included retrospective protocols (for a similar procedure, see Conrad et al., 1998). Their use was based on the assumption that right after estimating category frequencies, participants are aware of how they arrive at them and that they are able to roughly describe their estimation strategy.

Experiment 2A also included response latencies for each estimate. These would provide additional evidence for different estimation strategies in the relevant and irrelevant categorization conditions. Moreover, they would allow pitting availability by number against availability by ease. If people derive category frequency estimates from availability by number, high estimates should be associated with *longer* latencies than are low estimates, since it takes longer to recall many instances than just a few (for a similar reasoning, see Brown, 1995; Conrad et al., 1998; Conrad et al., 2003). If people derive frequency estimates from availability by ease, high estimates should be associated with *shorter* latencies than are low estimates, since instances that come to mind easily also come to mind quickly. Because most of the participants in the irrelevant categorization condition were expected to rely on availability to estimate category frequencies, we predicted that the frequency estimates and their latencies would be related in that condition. The direction of this relationship would tell whether the participants used availability by ease (negative relation) or availability by number (positive relation). In contrast, most of the participants in the relevant categorization condition were expected to rely on stored frequency information to estimate category frequencies. There is no reason to expect a relationship between frequency estimates and these estimates' latencies if estimates are based on stored category frequency information that is simply there to retrieve. Consequently, we predicted that the (positive or negative) relationship between the frequency estimates and their latencies would be stronger in the irrelevant categorization condition than in the relevant categorization condition.

In Experiment 2B, we measured how often and in which direction the participants would change their estimates after they had initially given them. We assumed that if the participants had the opportunity to adjust their category frequency estimates, they would change them more often when they had based their initial estimates on availability than when they had based them on stored frequency information. In the former case, searching memory for instances of one category might bring to mind instances of another category. This might prompt the participants to increase their estimate for the former category. If the participants with an irrelevant categorization goal at encoding used availability more often than did those with a relevant categorization goal, therefore, the former participants should change their estimates more often than would the latter ones. In addition, the changes that did occur in the irrelevant categorization condition should usually entail increases, whereas the changes that occurred in the relevant categorization condition might entail both increases and decreases (reflecting calibration of previous estimates).

It should be noted that allowing the participants to change their estimates precluded measuring response latencies for individual estimates. However, we did record the total latency. We expected that the participants in the relevant categorization condition would give their frequency estimates more quickly than would the participants in the irrelevant categorization condition. This prediction was based on the assumption that *retrieving* stored frequency information takes less time than recalling instances.

Method

Participants. Experiment 2A was run with 159 students (47 of them male, 112 female; mean age = 21.3 years) from various departments of the University of Leuven. They were paid €6.50 for their participation in this and an unrelated experiment. Experiment 2B was run with 104 students (53 of them male, 51 female; mean age = 20.1 years) from the economics department. They participated in fulfillment of a course requirement.

Procedure. Experiment 2A was identical to Experiment 1B, except for two points. First, we unobtrusively measured the response latencies for the frequency estimates. In order to do so, we let the category labels appear on the screen one by one and disappear after the participants had given an estimate for that category. The order of the labels was counterbalanced using a Latin square design. Second, at the end of the session, the participants wrote down how they had made their estimates. We cued them with three possibilities: a pure guess, an (exhaustive) retrieval of relevant instances, and a direct knowledge of the number of instances within each category. The instructions read as follows. "We would like you to write down how you have estimated the number of yellow, green, red, white, and brown/black items you have seen. What did you think of? Did you rely on some kind of feeling, did you know the answer without thinking very hard, did you try to remember all items of each color category when you made your estimates?"

In the frequency estimation task in Experiment 2B, in contrast, the participants saw all the categories simultaneously. They were allowed to change their estimates until they were satisfied. The nature and frequency of these changes were unobtrusively recorded. Second, because Experiment 2A yielded several protocols that could not be coded, we slightly clarified the instructions for the retrospective protocols. We asked the following. "We would like you to write down how you have estimated the number of yellow, green, red, white, and brown/black items you have seen. Did you know the response without thinking very hard or did it take you some time? Did you try to remember a few or all items of each color category, did you guess, did you know approximately how many items had appeared? On what did you base yourself to make the estimates?"

Results

Inferred strategy. Five participants in Experiment 2A and 4 participants in Experiment 2B could not be classified. The data of 1 participant in Experiment 2B were dropped because this participant had misunderstood the instructions. Thirteen participants in Experiment 2A and 17 participants in Experiment 2B were classified as *guessing*. For the participants that could be classified as relying on availability or on stored frequency information, the mean absolute differences between the two correlations were .26 (Experiment 2A) and .22 (Experiment 2B). For the resulting percentages, see Table 1.

Having versus not having categorized instances in relevant categories at encoding influenced the participants' estimation strategy [Experiment 2A, $\chi^2(2, N = 154) = 18.36, p < .001$; Experiment 2B, $\chi^2(2, N = 99) = 22.70, p < .001$]. They relied on guessing and on availability more often in the irrelevant categorization condition than in the relevant categorization condition (although for availability the difference was not significant in Experiment 2B). In contrast, the participants in the relevant categorization condition used stored frequency information more often than did the participants in the irrelevant categorization condition.

Correlations between objective and estimated frequency and recall. The participants who had processed the instances with a relevant category focus gave more accurate category frequency estimates than did the participants who had processed them with an irrelevant category focus. This may be derived from the mean within-subjects rank order correlation between objective and estimated category frequencies, which was higher in the relevant categorization condition than in the irrelevant categorization condition [Experiment 2A, $t(156) = 6.60$, $p < .001$; Experiment 2B, $t(98) = 5.82$, $p < .001$; see Table 2].

The mean within-subjects rank order correlation between the number of instances recalled and the estimated category frequency was also higher in the relevant categorization condition than in the irrelevant categorization condition [Experiment 2A, $t(156) = 3.16$, $p < .01$; Experiment 2B, $t(97) = 2.34$, $p < .05$]. This effect disappeared after eliminating the *guessers* [Experiment 2A, $t(144) = 1.58$, $p = .12$; Experiment 2B, $t(83) = 0.07$, $p = .94$]. This suggests that it was mainly due to the higher proportion of *guessers* in the irrelevant categorization condition than in the relevant categorization condition. The mean within-subjects rank order correlation between the number of instances recalled and the objective category frequency did not differ between conditions [Experiment 2A, $t(157) = 1.28$, $p = .20$; Experiment 2B, $t(100) = 0.65$, $p = .51$].

Retrospective protocols. Two coders who were blind with respect to the categorization task scored the protocols. They indicated whether the participants referred to stored category frequency information, availability, or guessing. The coders were instructed to indicate *stored frequency information* if the participant mentioned somehow knowing that one category had occurred more often than another category or mentioned having the impression that a given category had occurred a few times or a lot. They specifically had to indicate *stored frequency information* if the participant additionally denied having retrieved or having tried to retrieve instances and denied having guessed. The coders were instructed to indicate *availability* if the participant mentioned (1) having retrieved instances and having counted them—possibly adding a few instances to allow for forgetting and reporting the (adjusted) count—or (2) having tried to retrieve instances but not necessarily having counted them. Although we originally had instructed the coders to distinguish between availability by ease and availability by number, the descriptions that the participants gave of their estimation strategy were unfortunately too vague to allow such a distinction. In fact, most of the participants who referred to retrieval simply mentioned having tried to retrieve or having retrieved instances but did not mention whether they relied on ease or on number to make an estimate. Finally, the coders were instructed to indicate *guessing* if the participant mentioned having had no clue or having guessed and if no trace was found of *availability* or *stored frequency information*. The coders were allowed to indicate more than one strategy per protocol. They were also allowed not to indicate any strategy (rendering the protocol *unclassifiable*). Discrepancies were resolved through discussion.

Table 2
Mean Kendall Correlations (With Standard Deviations) Between Objective Category Frequency, Instance Recall, and Estimated Category Frequency, as a Function of Categorization Relevance

	Experiment 1A				Experiment 1B				Experiment 2A				Experiment 2B			
	Irrelevant ($n = 36$)		Relevant ($n = 35$)		Irrelevant ($n = 39$)		Relevant ($n = 36$)		Irrelevant ($n = 83$)		Relevant ($n = 76$)		Irrelevant ($n = 51$)		Relevant ($n = 52$)	
Mean Correlation Between	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Objective category frequency and recall	.61	.27	.63	.26	.67	.27	.76	.17	.68	.23	.72	.20	.64	.28	.68	.27
Objective category frequency and estimated category frequency	.41	.30	.77	.22	.58	.32	.78	.19	.40	.41	.75	.21	.35	.38	.72	.25
Recall and estimated category frequency	.58	.35	.66	.25	.64	.29	.72	.23	.50	.37	.67	.30	.44	.41	.60	.26

Excluding *unclassifiable* protocols, categorization relevance affected the self-mentioned estimation strategies [Experiment 2A, $\chi^2(3, N = 142) = 39.05, p < .001$; Experiment 2B, $\chi^2(3, N = 97) = 73.03, p < .001$; see Table 1]. The participants in the irrelevant categorization condition mentioned having relied on availability or having guessed more often than did the participants in the relevant categorization condition. The participants in the relevant categorization condition mentioned having relied on stored frequency information (either by itself or in combination with availability) more often.

The classification on the basis of retrospective protocols was significantly associated to the classification obtained using the correlations [excluding cases that could not be classified using either procedure; Experiment 2A, $\chi^2(6, N = 138) = 23.81, p < .001$; Experiment 2B, $\chi^2(6, N = 94) = 23.92, p < .001$]. However, the degree of association was only modest [Cramér's *V*: Experiment 2A, .29; Experiment 2B, .36], for several reasons. First, self-mentioned strategies allowed the possibility of classifying a participant as having used both stored category frequency information and availability, whereas the correlational inference procedure did not. In addition, the participants with low correlations were inferred to have been guessing. However, the majority of these participants claimed to have relied on availability. Either social desirability affected their claims or, for them, availability led to poor estimates. The latter possibility is more in line with our observation that availability generally leads to less accurate estimates than does relying on stored category frequency information. Moreover, social desirability cannot explain why self-mentioned guessing was more prominent in the irrelevant categorization condition than in the relevant categorization condition. Finally, 30% of the participants who mentioned having used availability or stored frequency information exclusively were inversely classified using their correlations. Probably, this was due to the large sample fluctuation of correlations calculated over only five observations (the five categories for each participant), making misclassifications rather likely. In fact, given the limited number of categories, the classification on the basis of the correlations may be less reliable than the classification on the basis of the retrospective protocols.

Estimate changes (Experiment 2B). As compared with the participants in the relevant categorization condition, the participants in the irrelevant categorization condition did not change their minds more often immediately after entering their estimates—that is, before clicking into the response box of another category [$M = 0.57$, $SD = 0.92$, vs. $M = 0.56$, $SD = 1.07$; $F(1,101) = 0.01$, $p = .96$]. After clicking into the response box of another category, however, they returned to change a previous estimate more often [$M = 1.00$, $SD = 1.96$, vs. $M = 0.37$, $SD = 0.89$; $F(1,101) = 4.51, p < .05$].

For the participants who changed their estimates, we calculated direction scores for each category. This score was +1 if the final estimate for the given category was larger than the initial one, −1 if it was lower, and 0 if it was identical to it. For each participant, a total direction score was obtained by summing the five direction scores. Mean total direction scores were greater than zero (implying that the changes mostly involved increases) in the irrelevant categorization condition [$M = 1.92$, $SD = 1.82$; $t(23) = 5.17, p < .001$], but not in the relevant categorization condition [$M = -0.10$, $SD = 1.41$; $t(19) = 0.32, p = .75$]. The difference between the conditions was significant [$F(1,42) = 16.40, p < .001$].

Estimate latencies. We log-transformed the latencies to remove skewness and analyzed the log-transformed latencies,[6] using a mixed model (i.e., a multilevel model). The estimates were regressed onto the log-latencies, taking into account the fact that the five within-subjects observations were not independent. The higher the estimate, the shorter its latency [Experiment 2A, $F(1,634) = 9.08$, $p < .001$]. This effect was qualified by an interaction with categorization relevance [Experiment 2A, $F(1,634) = 8.64$, $p < .001$]. Higher estimates were associated with shorter latencies in the irrelevant categorization condition [slope = −0.04; $t(634) = 3.82, p < .001$], but not in the relevant categorization condition [slope = −0.0005; $t(634) = 0.06, p = .95$].

In addition, estimates were generally given more quickly in the relevant than in the irrelevant categorization condition, both when individual estimate latencies were examined [Experiment 2A, $M = 8.84$, $SD = 0.46$, vs. $M = 9.27$, $SD = 0.57$; $t(157) = 5.25, p < .001$] and when total latencies were examined [Experiment 2B, $M = 10.21$, $SD = 0.41$, vs. $M = 10.69$, $SD = 0.72$; $F(1,101) = 17.34$, $p < .001$]. The total latency remained smaller in the relevant categorization condition after controlling for the observed difference in estimate changes [$F(1,100) = 14.67, p < .001$; $M = 10.27$ vs. $M = 10.64$].

Two further analyses supported the view that the difference in estimation times was probably due to the different strategies being used. First, the individual estimation log-latencies (Experiment 2A) differed across the four types of self-mentioned strategies [$F(3,138) = 12.04, p < .001$].[7] Post hoc Tukey comparisons revealed that the participants who mentioned having exclusively relied on availability ($M = 9.37$, $SD = 0.59$) were slower than the participants who mentioned having relied on stored frequency knowledge ($M = 8.76$, $SD = 0.39$) or on both stored frequency knowledge and availability ($M = 9.00$, $SD = 0.51$). The latter two groups also differed significantly. The participants who mentioned having guessed ($M = 8.87$, $SD = 0.42$) did not differ from any of the other three groups, probably because of low power (since very few participants mentioned guessing). Second, the log-transformed total estimation times (Experiment 2B) also differed across self-mentioned strategies [$F(3,93) = 11.95, p < .001$]. Post hoc Tukey comparisons revealed that the participants who mentioned having relied exclusively on availability ($M = 10.84$, $SD = 0.70$) were slower than the participants who mentioned having relied on stored frequency knowledge ($M = 10.17$, $SD = 0.36$), on both stored frequency knowledge and availability ($M = 10.31$, $SD = 0.49$), or on guessing ($M = 10.09$, $SD = 0.40$). The latter three groups did not differ significantly, also probably because of low power (since very few participants mentioned guessing or relying on both stored frequency knowledge and availability).

Discussion

On the basis of the pattern of correlation between objective and estimated frequency estimates and instance recall, we again inferred whether availability (irrelevant categorization) or stored frequency information (relevant categorization) was used to estimate category frequencies. Availability seemed to be more heavily used in the irrelevant categorization condition than in the relevant categorization condition. More participants seemed to rely on stored frequency information, and fewer participants seemed to rely on mere guessing, in the relevant categorization condition than in the irrelevant categorization condition. The results of Experiments 2A and 2B thus support the conclusion that category focus at encoding affects whether availability or stored frequency information is used to estimate category frequencies.

The results of the additional process measures support this conclusion. According to the retrospective protocols (Experiments 2A and 2B), the participants in the relevant categorization condition had relied on stored frequency information more often than had those in the irrelevant categorization condition. The latter participants said that they had relied on availability or that they had guessed more often. In addition, both individual and overall latencies were longer in the former condition than in the latter (Experiments 2A and 2B). Individual response latencies were related to the size of the estimates in the irrelevant categorization condition, but not in the relevant categorization condition (Experiment 2A). Finally, the participants in the irrelevant categorization condition changed their estimates for a category, after thinking about another category, more often generally and, particularly, more often in an upward direction than did the participants in the relevant categorization condition (Experiment 2B).

The conclusion seems warranted, therefore, that the activation of relevant categories at encoding enables people to accumulate category frequency information that they can store in memory and retrieve for future use. If the relevant categories are not activated during encoding, they inevitably resort to availability or to guessing. Experiment 2A allowed distinguishing between availability by ease and availability by number. The higher the estimates in the irrelevant categorization condition were, the shorter the response latencies became. This is more consistent with availability by ease than with availability by number. In addition, although the latencies were higher in the irrelevant categorization condition than in the relevant categorization condition, they were nevertheless still rather low in an absolute sense in the irrelevant categorization condition (Experiment 2A, M = 14.1 sec, SD = 9.8 sec, median = 10 sec; Experiment 2B, for five estimates, M = 29.9 sec, SD = 15.3 sec, median = 24.1 sec). These low latencies are also more consistent with availability by ease than with availability by number.

Our findings partially replicate the findings of Conrad et al. (2003). Their participants were slower and less accurate in an implicit property condition (which was equivalent to our irrelevant categorization condition) than in an explicit property condition (which resembled but was not identical to our relevant categorization condition). We obtained a similar pattern. The participants in their implicit property condition also said that they had relied more often on some form of enumeration and less often on a *general impression* than did the participants in their explicit property condition. Insofar as *a general impression* included automatically stored frequency information, this finding also was similar to ours.

However, Conrad et al. (2003) found a positive relationship between estimates and latencies in the explicit property condition and no relationship in the implicit property condition. We found no relationship in the relevant categorization condition and a negative relationship in the irrelevant categorization condition. One explanation for this difference may be based on procedural differences. We *explicitly asked* the participants to categorize the instances into relevant or irrelevant categories. Conrad et al. (2003) *assumed* that the participants in the explicit property condition categorized the instance into relevant categories. It is possible that at least some of these participants did not consciously categorize the instances but that the category labels merely helped the participants to (exhaustively) recall the instances.

Experiments 1 (A and B) and 2 (A and B) suggest that category focus at encoding affects category frequency estimation strategies. However, many studies that have supported the availability hypothesis have used (some variant of) the famous people paradigm (e.g., Manis et al., 1993; McKelvie, 1995, 1997; McKelvie & Drumheller, 2001; Tversky & Kahneman, 1973). Because it differs dramatically from ours, we decided to test our hypothesis with the famous people paradigm as well.

EXPERIMENT 3

In Experiment 3, we tested the prediction that category focus at encoding reduces reliance on availability when category frequencies are estimated in the famous people paradigm. After reading a list of 10 male and 10 female names, participants should remember more male names if the list included more famous men than famous women than if it included more famous women than famous men. If they base their category frequency estimates on availability, they should, therefore, estimate the number of men to be higher in the former case than in the latter one. Since we expected that only the participants in the memorization condition would base their frequency estimate on availability, we expected the estimated number of male names to be affected by the number of famous men on the list in the memorization condition, but not in the categorization condition. In other words, we expected an interaction between list type (more famous men vs. more famous women) and processing goal at encoding (memorization vs. categorization).

Method

The participants (12 of them male and 68 female first-year students; mean age of 18.2 years) viewed a list consisting of 10 male and 10 female names (first + last name). The names appeared one by one in the middle of a computer screen. Each list was constructed so that half of the names belonged to famous people (on the basis of a pretest on 120 first-year students): eight famous men and two famous women (famous men set) or two famous men and eight famous women (famous women set). Half of the participants received a famous men set, whereas the other half received a famous women set. The participants in the memorization condition had to memorize the names. They were told that the names would appear at the rate of one name every 3 sec. This presentation rate might render the task difficult, but they were urged to do their best. The participants in the categorization condition had to categorize, within 3 sec, each name on gender by pressing the V button for female names and the M button for male names. It was stressed that they should read the full name (first *and* family name), because a pilot study had shown that participants with categorization instructions tend to focus on the first name. This strategy would reduce the differential availability of the famous names, because no single first name was famous by itself (e.g., contrast *Albert* with *Albert Einstein*). After the presentation phase, the participants estimated the number of names they had seen (total frequency estimate) and the number of male names on the list. They then wrote down as many names as they could remember. They were encouraged to write down the entire name, although they were allowed to write down a first or a last name only. Finally, for each recalled name, the participants indicated whether it belonged to a man or a woman. This allowed us to score the perceived gender of last names that were recalled without a first name.

Results and Discussion

A name was scored as recalled if the entire name or just the first or last name was recalled. We did not correct for intrusions. If a participant bases his or her estimate on his or her recall, the distinction between correct recalls and intrusions is irrelevant, because participants do not make it themselves (e.g., Brown, 1997; Watkins & LeCompte, 1991; Williams & Durso, 1986).

The proportion of male names recalled was subjected to a 2 (fame: famous men set or famous women set) × 2 (instruction: memorization vs. categorization) between-subjects ANOVA ($MS_e = 0.0192$).[8] The proportion of male names recalled was higher for a famous men set ($M = .71$, $SD = .15$) than for a famous women set ($M = .32$, $SD = .12$) [$F(1,76) = 156.37$, $p < .001$]. The instruction × fame interaction was not significant [$F(1,76) = 0.75$, $p > .39$; see Table 3].

To increase the power of the analysis on estimated category frequencies, we eliminated any variability due to individual differences by transforming the estimates into relative ones. Each participant's estimated number of male names was divided by his or her total frequency estimate.[9]

A 2 (fame: famous men set or famous women set) × 2 (instruction: memorization vs. categorization) between-subjects ANOVA ($MS_e = 0.0151$) on the relative estimates yielded a main effect of fame [$F(1,76) = 19.33$, $p < .001$]. The participants who had received the famous men set thought they had seen more male names ($M = .60$, $SD = .12$) than did the participants who had received the famous women set ($M = .48$, $SD = .14$). However, this effect was qualified by a fame × instruction interaction [$F(1,76) = 10.04$, $p < .01$].[10] Planned comparisons revealed that in the memorization condition, the famous men set led to higher estimates of male names than did the famous women set [$t(76) = 5.35$, $p < .001$]. In contrast, in the categorization condition, the famous men set did not lead to higher estimates of male names than did the famous women set ($t < 1.00$, $p > .38$; see also Table 3). Given that famous names were more available than nonfamous names, the participants in the memorization condition seem to have derived their category frequency estimates from availability, whereas the participants in the categorization condition did not. Experiment 3 thus supported the conclusions of Experiments 1 and 2 with the famous people paradigm.

GENERAL DISCUSSION

Our research shows that category focus at encoding affects how people estimate category frequencies. If people categorize stimuli into relevant categories, they tend to base their category frequency estimates on stored frequency information (Experiments 1–3). If they categorize them into irrelevant categories (Experiments 1 and 2) or merely memorize the stimuli (Experiment 3), they tend to base their category frequency estimates on availability or resort to guessing. These conclusions are based on experiments with two different designs, two sets of stimuli (gendered names and colored objects), and a variety of measures (retrospective protocols, estimate changes, and correlations between objective and subjective frequencies, response times, and recall).

Admittedly, some of the participants in Experiments 1 and 2 who focused on relevant categories seemed to have based their category frequency estimates on availability. Clearly, people who make a relevant categorization do not *have to* base category frequency estimates on stored frequency information. Still, people *can* use stored frequency information only if they have classified the instance into relevant categories. Whether they use availability despite the presence of stored category frequency information

Table 3
Relative Estimated Number of Male Names and Proportional Recall of Male Names as a Function of Fame and Instruction (With Standard Deviations)

	Relative Estimated Number				Relative Recall			
	Memorization		Categorization		Memorization		Categorization	
Set	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Famous men	.63	.13	.57	.10	.74	.14	.67	.16
Famous women	.42	.15	.53	.10	.33	.15	.31	.10

may depend on situation features, such as having engaged in recall prior to the estimation task (see, e.g., Betsch et al., 1999; Bruce et al., 1991) or being instructed to think carefully (e.g., Haberstroh & Betsch, 2002). Moreover, individuals may differ in their confidence in stored frequency information. As Brown (1997) argued, "[some] people [may] prefer concrete information (in the form of instance counts . . .) to vague information [such as stored frequency information]" (p. 911). In addition, people who have access to stored frequency information may base their estimates on that information but may supplement it with availability information. In fact, in Experiment 2, some participants in the relevant categorization condition mentioned that they had used both stored frequency information and availability. Anecdotally, most of them indicated that they had roughly estimated the frequencies of the large categories and that they had tried to recall the items of the smaller categories, possibly to fine-tune their estimates for those categories.

Of the participants whose attention was not directed to the relevant categories, some appeared to have used stored frequency information (Experiments 1 and 2). One explanation may be that they had spontaneously categorized the instances despite our use of color categories to discourage such spontaneous categorizations. Interestingly, some studies in which semantic categories were used—for which spontaneous categorizations are even more likely (cf. Nelson et al., 1979; Warren, 1972; Wickens, 1970)—showed that participants did not base their category frequency estimates on availability either (e.g., Alba et al., 1980; Watkins & LeCompte, 1991).

Our findings support the view that people who categorize stimuli into relevant categories store frequency information for these categories. One may wonder how such information is represented in memory. Hypotheses may be derived from theories about the estimation and the mental representation of frequencies of occurrence (i.e., event frequencies). Despite the fact that they have to do with different entities (individual stimuli vs. categories), frequencies of occurrence and category frequencies may be similar. In fact, category frequencies may be thought of as frequencies of occurrence *of a category*. If category labels appear alongside its instances *and* receive full attention, category frequencies are even identical to frequencies of occurrence. In most real-life situations that require category frequency estimation, however, the category label is *not* presented. In such a case, the basic difference between frequencies of occurrence and category frequencies is that the referents of the former appear explicitly, whereas the referents of the latter do not. In a sense, then, category frequencies are frequencies of *implicit* occurrences.

To date, the MINERVA2 model (Hintzman, 1988) is the most elaborate and successful theory of how people estimate frequencies of occurrence and how they represent this information in memory. Its basic assumption is that each time an instance occurs, a separate memory trace is laid down (cf. the multiple-trace hypothesis). When people estimate how many times an instance occurred, they probe their memory with the representation of that instance. As a result, memory traces are activated in parallel to the extent that they resemble the probe. Memory traces that are very similar to the probe are strongly activated, whereas memory traces that are very dissimilar to the memory probe are hardly activated at all. The sum of the activations of the individual memory traces represents a signal of familiarity. The stronger it is, the higher the frequency estimate.

To apply the MINERVA2 model to category frequency estimation, we need to assume that if people focus on a relevant category at encoding, the memory traces contain information on the category alongside information on the instances. To estimate how often a category occurred, people may probe their memory with the representation of that category and use the resulting signal of category familiarity to estimate the frequency. Category frequency may, then, be represented as the number of memory traces that refer to the category, provided that they do contain category information. If people do not categorize the instances, memory traces contain no category information. Hence, no signal of familiarity may be generated. People then need to resort to guessing or instance availability to estimate category frequencies.

Alternatively, categorization may set up a counter for the category that is incremented each time an instance is encountered (cf. Underwood's [1969] theory of event frequency information). In that case, category frequency may simply be read off the counter. For people who do not categorize the instances, no counter is set up. They cannot rely on stored frequency information to estimate category frequencies.

Clearly, our experiments were not designed to distinguish between theories of the mental representation of category frequencies. It is, however, an interesting avenue for future research. In addition, future research may be needed to examine the generalizability of our results. Our experiments showed that category focus at encoding does, indeed, affect how people subsequently estimate category frequencies for relatively simple lists. It remains to be seen to what extent this remains the case for more demanding lists, although we would suspect that increased complexity would render availability even less likely for people who had a category focus at encoding.

REFERENCES

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

Barsalou, L. W., & Ross, B. H. (1986). The roles of automatic and strategic processing in sensitivity to superordinate and property frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 116-134.

Betsch, T., Siebler, F., Marz, P., Hormuth, S., & Dickenberger, D. (1999). The moderating role of category salience and category focus in judgments of set size and frequency of occurrence. *Personality & Social Psychology Bulletin*, **25**, 463-481.

Beyth-Marom, R., & Fischhoff, B. (1977). Direct measures of availability and judgments of category frequency. *Bulletin of the Psychonomic Society*, **9**, 236-238.

Brooks, J. E. (1985). Judgments of category frequency. *American Journal of Psychology*, **98**, 363-372.

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 1539-1553.

Brown, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 898-914.

Bruce, D., Hockley, W. E., & Craik, F. I. M. (1991). Availability and category-frequency estimation. *Memory & Cognition*, **19**, 301-312.

Conrad, F. G., Brown, N. R., & Cashman, E. R. (1998). Strategies for estimating behavioural frequency in survey interviews. *Memory*, **6**, 339-366.

Conrad, F. G., Brown, N. R., & Dashen, M. (2003). Estimating the frequency of events from unnatural categories. *Memory & Cognition*, **31**, 552-562.

Curt, C. L., & Zechmeister, E. B. (1984). Primacy, recency, and the availability heuristic. *Bulletin of the Psychonomic Society*, **22**, 177-179.

Epstein, M. L., Dupree, D. A., & Gronikowski, L. A. (1979). Encoding specificity and contextual similarity. *Bulletin of the Psychonomic Society*, **14**, 177-180.

Freund, J. S., & Hasher, L. (1989). Judgments of category size: Now you have them, now you don't. *American Journal of Psychology*, **102**, 333-352.

Greene, R. L. (1989). On the relationship between categorical frequency estimation and cued recall. *Memory & Cognition*, **17**, 235-239.

Haberstroh, S., & Betsch, T. (2002). Online strategies versus memory-based strategies in frequency estimation. In P. Sedlmeier & T. Betsch (Eds.), *Etc. Frequency processing and cognition* (pp. 205-220). Oxford: Oxford University Press.

Hanson, C., & Hirst, W. (1988). Frequency encoding of token and type information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 289-297.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.

Kverno, K. S. (2000). Trait anxiety influences on judgments of frequency and recall. *Personality & Individual Differences*, **29**, 395-404.

Lewandowsky, S., & Smith, P. W. (1983). The effect of increasing the memorability of category instances on estimates of category size. *Memory & Cognition*, **11**, 347-350.

Maley, J. E., Hunt, M., & Parr, W. (2000). Set-size and frequency-of-occurrence judgements in young and older adults: The role of the availability heuristic. *Quarterly Journal of Experimental Psychology*, **53A**, 247-270.

Manis, M., Shedler, J., Jonides, J., & Nelson, T. E. (1993). Availability heuristic in judgments of set size and frequency of occurrence. *Journal of Personality & Social Psychology*, **65**, 448-457.

McKelvie, S. J. (1995). Bias in estimated frequency of names. *Perceptual & Motor Skills*, **81**, 1331-1338.

McKelvie, S. J. (1997). The availability heuristic: Effects of fame and gender on the estimated frequency of male and female names. *Journal of Social Psychology*, **137**, 63-78.

McKelvie, S. J., & Drumheller, A. (2001). The availability heuristic with famous names: A replication. *Perceptual & Motor Skills*, **92**, 507-516.

Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research*, **20**, 431-440.

Michela, J. L. (1990). Within-person correlational design and analysis. In C. Hendrick & M. Clark (Eds.), *Review of personality and social psychology: Vol. II. Research methods in personality and social psychology* (pp. 279-311). Newbury Park, CA: Sage.

Nelson, T. O., Fehling, M. R., & Moore-Glascock, J. (1979). The nature of semantic savings for items forgotten from long-term memory. *Journal of Experimental Psychology: General*, **108**, 225-250.

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 & Social Psychology*, **61**, 195-202.

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

Underwood, B. J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology*, **70**, 122-129.

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

Warren, R. E. (1972). Stimulus encoding and memory. *Journal of Experimental Psychology*, **94**, 90-100.

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

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review*, **77**, 1-15.

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

NOTES

1. In our experiments as well, relevant categorization routinely led to better recall than did irrelevant categorization [Experiment 1A, $M = 12.9$, $SD = 2.5$, vs. $M = 11.2$, $SD = 2.9$, $t(69) = 2.5$, $p < .05$; Experiment 1B, $M = 14.6$, $SD = 2.7$, vs. $M = 13.0$, $SD = 2.7$, $t(73) = 2.6$, $p < .05$; Experiment 2A, $M = 14.4$, $SD = 2.6$, vs. $M = 12.6$, $SD = 2.5$, $t(101) = 3.6$, $p < .001$; Experiment 2B, $M = 14.7$, $SD = 2.8$, vs. $M = 12.7$, $SD = 3.6$, $t(153) = 3.7$, $p < .001$].

2. We agree with Watkins and LeCompte (1991) that the extrapolation from the recalled number of instances to category frequency estimates is monotonically increasing but not necessarily linear. Therefore, we calculated Kendall rank order correlations, rather than Pearson correlations.

3. Methodologically, it would have been better if the irrelevant categorization condition also involved five categories. Unfortunately, we were unable to come up with two categorization criteria each involving five categories with sufficient stimuli.

4. Although the classification procedure is error free, the data used to classify the participants are not. As a result, some participants may be misclassified.

5. Using absolute estimates seems to entail a more conservative test, since the effect of categorization relevance on the inferred strategies was slightly, although not significantly, smaller in Experiment 1B than in Experiment 1A [$\chi^2(1, N = 132) = 2.54$, $p = .11$].

6. Since we analyzed log-latencies, reported means and standard deviations also refer to log-latencies (i.e., are not back-transformed).

7. Consistent with our observation that the inferred strategies may contain more misclassifications than do the self-mentioned strategies, additional analyses conditionalizing on inferred strategy, rather than on self-mentioned strategy, yielded comparable but less significant results.

8. The results of the proportional recall data parallel those of the raw recall data, with one exception. In the analysis on the raw recall data, a main effect of instruction was obtained [$F(1,40) = 4.81$, $MS_e = 1.82$, $p < .05$]. More (male and female) names were recalled after memorization ($M = 7.87$, $SD = 2.15$) than after categorization ($M = 6.67$, $SD = 1.88$).

9. In line with previous work, the participants underestimated the *total* number of names they had seen [$M = 16.90$; $t(79) = -5.64$, $p < .001$]. A 2 (instruction: memorization or categorization) × 2 (fame: famous men set or famous women set) between-subjects ANOVA revealed no significant effects (all Fs < 1.64, all ps $> .20$). So, the total frequency estimate apparently did not depend on instruction or fame.

10. The fame × instruction interaction was also significant when an ANOVA was conducted on the raw estimated number of male names.

(*Continued on next page*)

APPENDIX
Stimulus Pool Used in Experiments 1A, 1B, 2A, and 2B

Category	Brown/Black	Green	Red	White	Yellow
Fit for consumption	wholemeal bread	spinach	strawberry	milk	french fries
	candy syrup	lettuce	lobster	lump of sugar	lemon
	liquorice	gherkin	tomato	salt	banana
	coffee	sprouts	ketchup	whipped cream	corn
	plain chocolate	broccoli	cherry	cauliflower	cheese
Not fit for consumption	tar	cactus	fire engine	gypsum	canary
	raven	crocodile	blood	paper	sun
	mussel shell	pine tree	poppy	snow	sunflower
	coal	grass	can of cola	polystyrene foam	dandelion
	tire	grasshopper	clown nose	judo suit	straw

(Manuscript received May 30, 2003;
revision accepted for publication February 14, 2005.)