This article was downloaded by:[Stenberg, Georg]
On: 8 February 2008
Access Details: [subscription number 790480663]
Publisher: Psychology Press
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

European Journal of Cognitive Psychology

Publication details, including instructions for authors and subscription information:
<http://www.informaworld.com/smpp/title-content=t713734596>

The memorability of names and the divergent effects of prior experience

Georg Stenberg ^a, Johan Hellman ^a, Mikael Johansson ^b

^a Kristianstad University, Sweden

^b Lund University, Sweden

First Published on: 01 July 2007

To cite this Article: Stenberg, Georg, Hellman, Johan and Johansson, Mikael (2007)
'The memorability of names and the divergent effects of prior experience', European
Journal of Cognitive Psychology, 20:2, 312 - 345

To link to this article: DOI: 10.1080/09541440701398724

URL: <http://dx.doi.org/10.1080/09541440701398724>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

The memorability of names and the divergent effects of prior experience

Georg Stenberg and Johan Hellman
Kristianstad University, Sweden

Mikael Johansson
Lund University, Sweden

Pre-experimental familiarity can have paradoxical effects on episodic memory. Knowledge of the stimulus domain usually enhances memory, but word frequency—a presumed correlate of prior experience—is negatively related to recognition accuracy. The present study examined episodic recognition of names and its relation to two measures of pre-experimental knowledge, name frequency, and fame. Frequency was operationalised as the number of hits in a national telephone directory, and fame as hits on national mass media websites. Recognition accuracy was increased by fame, but diminished by frequency. Four experiments confirmed the findings, using yes/no recognition, ROC curves, and remember-know paradigms. Hit rates were consistently more strongly influenced by fame than by frequency, whereas the reverse was true for false alarm rates. These dissociations suggest that two different forms of semantic memory, specific and nonspecific knowledge, interact with episodic memory in separate ways.

The distinction between episodic and semantic memory has become well established, but research about the two phenomena largely proceed on separate tracks. Semantic memory studies are typically concerned with the structure of conceptual organisation, whereas episodic memory research cares mainly about the processes of encoding and retrieval. Yet there is considerable interdependence between the two branches, because events encoded in episodic memory are normally interpreted against a background of semantic knowledge, and when such foreknowledge is lacking, remembering suffers (Bartlett, 1932).

Correspondence should be addressed to Georg Stenberg, Dept. of Psychology, School of Behavioural Sciences, Kristianstad University, SE-291 88 Kristianstad, Sweden. E-mail: georg.stenberg@bet.hkr.se

As a first approximation to a generalisation, frequent prior experience of to-be-studied items should enhance episodic memory. From the perspective that memory is adapted to the environment, it is clearly rational to preserve in memory those items most likely to be encountered again (Anderson & Schooler, 1991, 2000; Dennis, 1995). Thus, items about which semantic knowledge has been amassed are more prone to be remembered after episodic encounters than less known items. In an early study (Allen & Garton, 1968), physics students recognised physics words from a list better than did arts students, but both groups of students recognised physics words better than common words. Similarly, in a more recent study (Chalmers, Humphreys, & Dennis, 1997), computer science students showed better episodic recognition of rare computer science terms than psychology students, although both groups showed an advantage for low-frequency over high-frequency words. The finding that relatively rare words are better recognised than relatively common words is a consistent and widely reproduced finding (reviewed by Chalmers & Humphreys, 1998; Reder et al., 2000), but it flies in the face of the generalisation that frequent prior experience enhances memory.

The *word frequency effect* has been a challenge for memory theories in more ways than one. Because most early models did not take pre-experimental experience into account, there was no explanation of the frequency effect. Second, if prior familiarity *was* explicitly addressed, theories predicted that it would raise *both* the hit rate *and* the false alarm rate. Instead, the typical finding is that hit rates are raised and false alarm rates are lowered. This pattern is often called the *mirror effect*, because the old and the new distributions are thought of as moving away from the criterion in opposite directions (Glanzer & Adams, 1985, 1990; Glanzer, Adams, Iverson, & Kim, 1993).

Word frequency, as measured by the probability of a word's occurrence in (usually newspaper) text, is expected to reflect experiential frequency, i.e., probability of encountering the word in daily life. That assumption has been called into question (Estes & Maddox, 2002). A more closely controlled manipulation of pre-experimental experience can be accomplished in a three-phase experiment, where stimuli are presented in a familiarisation phase before the usual study-test procedure. With both nonverbal and verbal materials, repeated presentations preceding the study and test phases have been seen to increase both hit rates and false alarm rates, sometimes leading to a net decline in sensitivity, d' (Estes & Maddox, 2002; Maddox & Estes, 1997).

The effect brought about by word frequency is different from prior familiarisation, although the specific functional relationship can vary depending on the range of frequencies included. If the range is wide and includes very low frequencies, the typical finding is an inverted-U relationship,

with maximum sensitivity at moderately low frequencies. The peak is reached by the combined effect of both high hit rates and low false alarm rates. In view of these divergent trends, Estes and Maddox (2002) concluded that the word frequency effect is a misnomer, because it reflects a quality different from experiential frequency, probably *lexicality* in a wide sense, which tends to be confounded with frequency.

The model used by Estes and Maddox (2002) belongs to the class of global matching models for recognition memory (Clark & Gronlund, 1996). A competing type of models, the *dual-process* class, has made a different interpretation. Two processes are available in recognition, according to these theories: recollection and familiarity, and they contribute in different ways to the overall word frequency effect (Arndt & Reder, 2002; Reder et al., 2000). The hit rate part reflects the fact that low-frequency words are more often recollected. The false alarm part can be ascribed to the greater familiarity of high-frequency words. Because there are two different contributions, they can be pried apart by experimental manipulations. Thus, if task demands necessitate extraordinarily fine discriminations to be made between targets and similar distractors, the data will show responses consistent with a high degree of recollection, reflected in the shape of the ROC curves (Arndt & Reder, 2002). If remember-know responses are recorded, high-frequency words will attract more know-responses than low-frequency words do, in keeping with their higher familiarity (Reder et al., 2000).

The present study is concerned with the kind of prior experience subsumed under the heading of general knowledge or semantic memory, and the effects this experience has on episodic memory. We use proper names as the stimulus material, these being the object of everyday semantic knowledge (as well as an oft-lamented source of memory lapses). We propose that knowledge of names takes two forms, one distinctive, such as knowing that Björn Borg is the name of a celebrated tennis player, and one nondistinctive, such as knowing that Tom Jones is a common name and Engelbert Humperdinck is not. These two forms of semantic memory bear a structural resemblance to two forms of episodic memory, recollection and familiarity. Furthermore, they interact with episodic memory in radically different ways. Distinctive semantic knowledge supports episodic memory, whereas the nondistinctive form may interfere with it.

Distinctive semantic memory shares with recollection a relative richness of associated detail; as applied to proper names it singles out a particular bearer of the name, and brings to mind known facts about that person. The name of a celebrity is associated with that person's looks, achievements, public appearances, etc., all of which can help to encode an episodic encounter with the name, such as seeing it in the newspaper. Nondistinctive memory, on the other hand, brings to mind a sense of familiarity, a sense of many previous encounters with the name, without any particular one coming

to the fore. Frequent names—such as Smith and Jones—often evoke such a sense of familiarity without bringing any particular bearer to mind. We aim to show that such familiarity is detrimental to episodic recognition memory. In that respect, the *name frequency effect*, which we aim to document, is a close relative of the word frequency effect. As mentioned above, it has been proposed that the word frequency effect is not really a result of frequency of encounters with the words in daily life (Estes & Maddox, 2002). Instead, lexicality, with which frequency may be confounded, could be the causative factor. Nonetheless, the present study provides further indications that the frequency effect is genuine, by demonstrating a parallel effect, using a different stimulus material.

Names (i.e., first name plus last name) that were used in this study varied orthogonally along two dimensions: frequency and celebrity. Both were meant to reflect environmental quantities. With *frequency* we refer to the relative number of persons bearing the name, by *celebrity* we mean the probability of the name being mentioned in the media. Operationally, frequency was measured as the number of hits in a computerised search of the national telephone directory. Celebrity was similarly defined as the number of hits in a search of the Internet pages of national news media. Conceptually as well as empirically, these are independent criteria, and four groups could be formed by cross-classifying high and low groups. Examples are given later, and the full lists of names are available on the Internet.¹ In the experiments, names were presented for study visually, and tested for recognition shortly thereafter.

We aimed to show that celebrity and frequency had opposite-sign effects on recognition accuracy, but we also wanted to show that they affected at least partly different memory systems. Our expectation was that distinctive semantic knowledge (correlated with celebrity) would feed into the episodic *recollection* system. By providing material for detailed encoding, it would engender context-rich memories.

Nondistinctive semantic knowledge (correlated with frequency), on the other hand, would feed into the episodic *familiarity* system, causing confusion in the process. Because familiarity is context-free by definition, it needs to be attributed to a source, and if several sources are possible, confusion may arise. The familiarity arising from frequent occurrences in the pre-experimental environment cannot be easily distinguished from the familiarity arising from study within the experiment. Because of this, frequency raises the noise level in the recognition process, and makes the signal difficult to distinguish.

¹ <http://www.stenberg.ys.se/Projects/Names/Names.htm>

Chalmers and Humphreys (1998) distinguished between generalised and episode specific strength in episodic recognition of high- and low-frequency words. Both factors were varied experimentally, generalised strength by frequency of presentation, and episode specific strength by the presentation of definitions in the familiarisation phase. Whereas making words distinctive by supplying definitions supported recognition accuracy, mere familiarisation was of doubtful value for recognition accuracy; in some conditions the effect was detrimental. The distinction we wish to make is similar to that of Chalmers and Humphreys, although distinctive semantic memory in our sense need not be episode specific. Knowledge concerning, e.g., Björn Borg is genuinely semantic in the sense of being accumulated over many instances. Events in which information about him was presented have themselves faded away from memory, leaving only the information behind, much as ancestors, long dead, have contributed the genetic code of the living.

Undoubtedly, episodes may be remembered still by many who witnessed Borg's Wimbledon victories, and they may even form personally relevant memories, perhaps charged with emotion and autobiographical significance. The point we wish to make is that episodes are not a *necessary* component of distinctive semantic memory. Knowledge of famous people can be detailed and individuating without being woven into the fabric of our own lives.

The distinction between autobiographically relevant and irrelevant knowledge of famous people has been elucidated by Moscovitch and colleagues (Westmacott & Moscovitch, 2003). They showed that celebrities that were associated with events in participants' own lives were better remembered in an episodic recognition experiment than celebrities not so associated. Furthermore, self-relevance also improved performance in tests of semantic memory. Westmacott & Moscovitch's results demonstrate the interdependence of semantic and episodic memory. We wish to make a related although different proposal, namely that a dividing line runs within semantic memory itself, irrespective of autobiographical association. Both fame and frequency are likely to increase the probability that a name evokes a personal memory, yet they have—as we purport to show—radically different effects on within-experiment episodic memory.

EXPERIMENT 1

The first experiment assessed the effects of distinctive and general semantic memory on episodic memory, using a pool of names with which participants had varying and measurable foreknowledge.

Method

Participants

Forty-seven students (34 women) participated, and were compensated with a lunch voucher. Ages ranged from 16 to 40, with an average of 24 years. They were randomly allotted to two orienting tasks, resulting in 27 participants in the celebrity orienting task, and 20 in the frequency orienting task.

Procedure

Participants were tested in groups of 2–16 at a time. The experiment took place in a laboratory, where participants were seated in separate booths, each with a computer, on which stimuli were presented using E-prime software. After a brief oral instruction and some on-screen instructions, participants ran the experiment at their own pace. The whole session took about half an hour.

The present experiment was divided into three study–test cycles, each presenting 32 names for study. Studied names reappeared in the test mixed with an equal number of distractors. Assignment of names to the study set or the distractor set was randomly and independently determined for each participant, as was the presentation order. The test phase followed the study phase without delay.

During study, each stimulus was preceded by a 1 s fixation cross. The name was displayed, centred on the screen, for 2 s, and during this time window a response was to be given to the orienting question (presented at the top of the screen as a reminder): either “Is this person famous?” or “Is this name frequent?” (The latter question had been specified in the instruction text to mean “Are there more than 10 bearers of that name in Sweden?”).

During the test phase, names were presented until a response was given or until 4 s had elapsed, whichever happened first. After the response, a brief (0.5 s) feedback concerning correctness of the response was given.

Materials

A priori ratings. A set of 192 Swedish names was constructed. Names were either selected from the set of those currently (late 2005) popular in the media, or combined (first name plus last name) using frequency tables provided by the national census bureau, Statistics Sweden. The experimenters, when constructing the stimulus material, judged each name as either Famous or Nonfamous, and either Frequent or Infrequent. Forty-eight names of each type were selected. Examples of common, famous

names were: Göran Persson, Ingmar Bergman, and Björn Borg; and of uncommon, famous names: Ingvar Kamprad, Greta Garbo, and Zlatan Ibrahimovic. Common, nonfamous names were, e.g., Maria Axelsson, Sven Holmgren, and Gustav Eklund, and rare, nonfamous names, Ernfrid Hammar, Hildegard Sten, and Guje Gagner. Celebrity names included ones whose claim to fame extended over decades (Garbo, Bergman), as well as others of more recent renown. The present set, and an expanded, second set are available at: <http://www.stenberg.ys.se/Projects/Names/Names.htm>.

Internet searches. To verify the judgements, names were checked for frequency by looking up each name in the Swedish, nationwide telephone directory (www.eniro.se), and noting the number of hits. This number was log transformed, and used as the variable Frequency, which was dichotomised into Frequent and Infrequent.

Similarly, *celebrity* was checked by making site-specific lookups via the Google search engine. Each name was searched at six Swedish websites, affiliated with important media: four national newspapers (www.dn.se, www.svd.se, www.expressen.se, www.aftonbladet.se) and two television networks (www.svt.se and www.tv4.se). Searches were made by a Visual Basic program, using a programming interface published by Google (www.google.com). The number of hits was added across sites, and log transformed. Finally, the variable was dichotomised into Famous and Nonfamous.

Participant ratings. In addition to the search data, participants were invited to rate the names for frequency or celebrity. Each rating task was allotted to one half of the participants as an orienting task in the study phase. Apart from providing validation of the stimulus classification, it also served to examine whether attention directed towards one dimension of the stimuli would affect memory for the names.

A priori ratings of celebrity and frequency were confirmed by the ratings given by participants and the pattern of hits in Internet searches. Agreement (Kendall's tau) was .86 with participant classification, and .98 with Internet hits, and between the latter two the correlation was .84 (see Table 1).

Internet hits in the phone directory was unrelated to hits on the media sites ($r = .04$), but both were correlated with the total sum of Google hits, a possible indicator of experiential frequency ($r = .43$ for the phone directory, and $r = .71$ for the media; $n = 192$).

Results

To allow generalisation across both subjects and items, two sets of analyses were performed (Clark, 1973), one with subjects, and the other with items, as

TABLE 1
Agreement between classification of stimuli by *a priori* judgements (columns) and empirical classification by Internet search (rows, upper half), and participant judgements (rows, lower half)

		<i>Frequent</i>		<i>Infrequent</i>		<i>Total</i>
		<i>Famous</i>	<i>Nonfamous</i>	<i>Famous</i>	<i>Nonfamous</i>	
Internet search						
Frequent	Famous	47	3			50
	Nonfamous	1	45			46
Infrequent	Famous			47		47
	Nonfamous			1	48	49
Total		48	48	48	48	192
Participant judgements						
Frequent	Famous	43	3	4		50
	Nonfamous	1	43	1	1	46
Infrequent	Famous	4	1	41		46
	Nonfamous		1	2	47	50
Total		48	48	48	48	192

source of the random error term. Although this analysis strategy has been questioned as a general practice (Raaijmakers, Schrijnemakers, & Gremmen, 1999), it is arguably necessary with the present data set, because random selection of items was possible only within conditions (Frequency and Celebrity), not across condition boundaries. Furthermore, analysis by items permitted examination of issues not otherwise accessible. In the subjects-based analysis, the *a priori* classification of the names was used, and in the item-based analysis, we used the empirical classification derived from Internet searches. Thus, we could perform regression analyses, predicting item hit rates and false alarm rates from our ratio-scaled Internet data.

By subjects. Hit rates and false alarm rates were recorded for each of the four *a priori* stimulus classes, averaged over items for each participant, and d' was computed (Snodgrass & Corwin, 1988). It was subjected to a $2 \times 2 \times 2$ analysis (Frequency $\times$ Celebrity $\times$ Orienting Task), the first two being within-participant factors, and the third a between-participant factor. Orienting Task had no effect, alone or in interaction, and it will not be further mentioned.

The d' measure (see Figure 1) showed a main effect of Frequency, because uncommon names were better recognised, $F(1, 45) = 42.19$, $p < .001$, $\eta_p^2 = .48$. Famous names were much better remembered than nonfamous, $F(1, 45) = 238.96$, $p < .001$, $\eta_p^2 = .84$. There was also an interaction between Celebrity and Frequency, $F(1, 45) = 6.24$, $p = .016$, $\eta_p^2 = .12$, due to a potentiated frequency effect among the celebrities.

Figure 1. Experiment 1: Values of d' (filled symbols) and C (unfilled symbols); circles: famous names; diamonds: nonfamous. Frequency is on the x-axis. Error bars show ± 1 standard error.

The criterion, C , was lower for famous names than for nonfamous, resulting in a Celebrity effect, $F(1, 45) = 67.38$, $p < .001$, $\eta_p^2 = .59$, and an interaction, Celebrity $\times$ Frequency, $F(1, 45) = 27.35$, $p < .001$, $\eta_p^2 = .37$ (Figure 1).

Hit rates showed an effect of Frequency, $F(1, 46) = 11.79$, $p = .001$, $\eta_p^2 = .20$, and a large effect of Celebrity, $F(1, 46) = 231.20$, $p < .001$, $\eta_p^2 = .83$, and an interaction, $F(1, 46) = 11.50$, $p = .001$, $\eta_p^2 = .20$.

False alarm rates were affected by Frequency, $F(1, 46) = 24.41$, $p < .001$, $\eta_p^2 = .35$, and by Celebrity, $F(1, 46) = 14.19$, $p < .001$, $\eta_p^2 = .24$, as well as by an interaction, $F(1, 46) = 25.83$, $p = .001$, $\eta_p^2 = .36$. Averages are given in Table 2.

Effect sizes for the effects of Frequency and Celebrity are also given in Table 2. There has been discussion as to which index of effect size is preferable in within-subjects designs, the η_p^2 or the η_G^2 , the latter being more comparable to between-subjects designs (Bakeman, 2005). We present both, to allow comparisons; the former in text and the latter in tables.

TABLE 2
Hit rates and false alarm rates in Experiment 1

	<i>Means</i>					
	<i>Low frequency</i>		<i>High frequency</i>		<i>Effect sizes, η^2_G</i>	
	<i>Nonfamous</i>	<i>Famous</i>	<i>Nonfamous</i>	<i>Famous</i>	<i>Frequency</i>	<i>Celebrity</i>
HR	.73	.90	.65	.90	.03	.63
FAR	.25	.14	.25	.26	.11	.08
HR (by items)	.74	.90	.65	.88	.07	.52
FAR (by items)	.25	.14	.24	.27	.04	.02
Mirror effect patterns	HR1	HR2	FAR2	FAR1		
Celebrity	.90	.69	.25	.20		
Frequency	.82	.78	.26	.20		

The lower part of the table verifies the mirror effect pattern by showing hit rates and false alarm rates collapsed across high and low levels of Frequency and Celebrity, respectively. The means are enumerated from left to right in the expected order of magnitude, under the assumption of a mirror effect. As the table shows, the data conformed to the expected pattern, inasmuch as there were mirror effects for both Frequency and Celebrity.

By items. Hits and false alarms were recorded for each stimulus, averaged over participants for each item, and item-wise hit rates and false alarm rates were computed. They were subjected to a 2×2 analysis (Frequency $\times$ Celebrity), both being between-items factors.

There was a reliable effect of Celebrity, $F(1, 188) = 206.04$, $p < .001$, $\eta^2_p = .52$, and of Frequency, $F(1, 188) = 13.92$, $p < .001$, $\eta^2_p = .07$ on hit rates. The interaction was also significant, $F(1, 188) = 6.53$, $p = .011$, $\eta^2_p = .03$.

False alarm rates showed an effect of Frequency, $F(1, 188) = 8.70$, $p = .004$, $\eta^2_p = .04$, that was stronger than the effect of Celebrity, $F(1, 188) = 4.35$, $p = .038$, $\eta^2_p = .02$. There was also an interaction, $F(1, 188) = 11.42$, $p = .001$, $\eta^2_p = .06$, due to particularly low false alarm rates for infrequent, famous names, compared to the other three groups.

Using the full range of predictor variables, instead of dichotomies, a multiple regression analysis was performed. First, *hit rate* was used as the dependent variable. Four potential predictors were entered into a stepwise regression: (a) the endorsement rate in the frequency orienting task (*freq_calls*), (b) the endorsement rate in the celebrity orienting task (*celeb_calls*), (c) the number of hits in the Internet search of the telephone directory (*freq_hits*), and (d) the number of hits in the Internet search of the media sites (*celeb_hits*). The latter two were log transformed for normality. The stepwise regression procedure settled for a model with three predictors:

Celeb_calls: $\beta = .52$, $t(188) = 8.40$, $p < .001$; *freq_calls*: $\beta = -.23$, $t(188) = -4.88$, $p < .001$; *celeb_hits*: $\beta = .28$, $t(188) = 4.58$, $p < .001$.

The same analysis was applied to the *false alarm rate*. Thus, a stepwise regression procedure was offered the same four potential predictors. Only one was selected as significant: *freq_hits*: $\beta = .18$, $t(188) = 2.51$, $p = .013$.

Thus, whereas hit rates were predicted by indicators of both frequency and celebrity (and more so by celebrity), false alarm rates were predicted by frequency alone.

Discussion

In keeping with the premise that prior experience furthers memory, famous names were retained much better than nonfamous names. However, completing the paradox of pernicious foreknowledge, frequent names were not well remembered at all; indeed, they fared far worse than very unusual names. The frequency effect was boosted in the group of famous names, but it was significant for famous and nonfamous alike.

Some of the infrequent names were combined from relatively rare constituent parts, and the resulting first+last name pairs could have had an unusual look and sound. To ascertain whether the frequency effect could be ascribed to this bizarreness aspect, a modified stimulus material was used in later experiments. It can be noted already at this point, however, that the frequency effect was in fact stronger among the celebrities, this being the basis of a significant interaction effect. In the famous names, any potential bizarreness would have been eroded by constant wear and use in the media, and to most native speakers these names would appear to be familiar household names.

Although both factors had quite marked effects on d' , separate influences could be noted on hit rates and false alarm rates. The impact of Celebrity on hit rates was huge, and that of Frequency paled by comparison. False alarms rates, on the other hand, were affected by Frequency, more so than by Celebrity, although both influences were relatively weak. It can be noted that the distinct patterns of effects on HR and FAR are some of the most important characteristics used to distinguish between familiarity and recollection (Reder et al., 2000).

EXPERIMENT 2

The purpose driving our further research in this area was, apart from replicating the basic findings, a wish to clarify the relation between the two dimensions of semantic memory and similar dimensions of episodic memory. In particular, we wished to make contact with the flourishing research on

familiarity and recollection, which has proposed methods of separating and measuring the two. An often used method is the *Remember–Know* paradigm, in which participants are questioned about the introspective quality of their memory. In the present context, high-celebrity names can be expected to elicit *Remember* responses. Analogously, frequency could possibly affect the rate of *Know* responses, perhaps especially those given erroneously, i.e., *Know* false alarms. We therefore adapted standard *Remember/Know* instructions (Rajaram, 1996) to the name memory task, which was otherwise presented as in Experiment 1. We anticipated some difficulty for the participants in performing the remember/know task with this particular material, arising from the possible confusion of pre-experimental familiarity with familiarity engendered within the experiment. In the instructions, we therefore emphasised that the subjective quality to which the label “*Know*” applied had nothing to do with previous knowledge (acquired, e.g., through the media), and that we asked participants to separate this way of “knowing” from the kind we wanted them to report, i.e., the familiarity produced by an earlier encounter within the experiment.

Method

Participants

Thirty-five students (26 women) participated in exchange for a lunch voucher. Ages ranged from 19 to 47, with an average of 23 years.

Procedure

Participants were tested in groups of 2–16 at a time. The experiment took place in a laboratory, where participants were seated in separate booths, each with a computer. After a brief oral instruction, they were given further on-screen instructions, and then ran the experiment at their own pace. This experiment was interleaved with a different, unrelated experiment, such that study of all (64) to-be-remembered names came first, followed by the other experiment (an Iowa gambling task), and finally a memory test for all the names (128). The retention period—about 10 min—was thus filled with a distracting task, but no names or other verbal material appeared in it. The whole session took about 40 min.

During study, each name was shown for 2 s, and no overt task was assigned during this period, except to memorise. In the test block, each name was presented along with two on-screen buttons, marked “Yes” and “No”, to be mouse-clicked during a 5 s period in response to the question “Did you see this name before in the experiment?”. In case of a “Yes” response, a three-button selection screen followed, with choices marked “Jag minns det” (“I remember”), “Det känns bekant” (“I know”), and “Jag gissar”

("I guess"). These alternatives had been extensively explained at the outset, using a modified version of the Rajaram (1996) instructions. "No" responses were not followed by any further selection.

Materials

An amended and expanded stimulus set of 288 names was used in this experiment. The main changes from the previous set were (a) the inclusion of 96 new names; (b) substitution of some infrequent, nonfamous names that could give rise to a bizarreness effect—all names (first + last name combination) had to have at least one directory-listed bearer as a requisite for inclusion (exceptions were made for nonlisted or deceased famous persons such as Greta Garbo); (c) an update of the fame and frequency data was performed through a renewed web search (February 2006), about 6 months later than the previous one—for the set of identical items in the two sets ($n = 132$), correlations were .95 and .87 for the phone directory hits and the media hits, respectively.

The number of phone directory hits was unrelated to the number of media hits ($r = .01$), but both were positively correlated with a third variable, which can be thought as a proxy for experiential frequency, the total number of Google hits ($r = .32$ and $r = .82$, respectively, all variables log transformed).

The selection of 64 targets and 64 distractors out of the 288 item pool was made randomly and independently for each participant, with the constraint that the four types of names be equally represented.

Results

As in the previous experiment, analysis was performed both by-subjects and by-items. Starting with the former, we made conventional analyses of d' , hit rates and false alarm rates, all in a 2×2 design (Frequency $\times$ Celebrity).

The remember-know data yielded two sets of variables, the remember rate (r), computed as the proportion of remember responses out of all old items, and the IRK-know rate (Table 3). The latter was computed as suggested by Jacoby and colleagues (Jacoby, Jones, & Dolan, 1998; Kelley & Jacoby, 1998), i.e., as the proportion of know responses out of targets not given a "remember" response. These variables, r and IRK_k , were subjected to a 2×2 analysis (Frequency $\times$ Celebrity). Additionally, we computed rFA and kFA , i.e., the proportion of false alarms given remember and know responses. The rate of false remember responses has been the focus of theoretical interest (Wixted & Stretch, 2004), and the rate of false know responses interested us because we suspected that frequency might have an influence on it.

TABLE 3
Hit rates and false alarm rates in Experiment 2

	<i>Means</i>					
	<i>Low frequency</i>		<i>High frequency</i>		<i>Effect sizes, η^2_G</i>	
	<i>Nonfamous</i>	<i>Famous</i>	<i>Nonfamous</i>	<i>Famous</i>	<i>Frequency</i>	<i>Celebrity</i>
HR	0.66	0.89	0.63	0.82	0.05	0.47
HR (by items)	0.68	0.89	0.62	0.83	0.02	0.27
r	0.46	0.79	0.32	0.68	0.20	0.68
IRK_k	0.54	0.86	0.41	0.76	0.15	0.60
FAR	0.16	0.11	0.22	0.22	0.17	0.02
FAR (by items)	0.14	0.12	0.23	0.22	0.07	0.00
rFAR	0.04	0.02	0.04	0.06	0.05	0.01
kFAR	0.07	0.05	0.10	0.11	0.10	0.00
Mirror effect patterns	HR1	HR2	FAR2	FAR1		
Celebrity	.86	.65	.19	.17		
Frequency	.78	.73	.22	.14		

r: remember; k: know; FAR: false alarm rate; HR: hit rate; IRK: independent remember-know. The lower part of the table verifies the mirror effect pattern by showing hit rates and false alarm rates collapsed across high and low levels of Frequency and Celebrity, respectively. See note to Table 1.

By subjects. The d' measure (Figure 2) was affected by both Frequency, $F(1, 34) = 25.45$, $p < .001$, $\eta^2_p = .43$, and Celebrity, $F(1, 34) = 86.55$, $p < .001$, $\eta^2_p = .72$. There was also an interaction, $F(1, 34) = 7.55$, $p = .010$, $\eta^2_p = .18$, due to larger differences between frequent and infrequent names among the famous. Despite differences in the stimulus material, all aspects of these results were well reproduced from the previous experiment.

Participants set the criterion, C , higher for nonfamous names, again replicating the Celebrity effect from the previous experiment, $F(1, 45) = 28.87$, $p < .001$, $\eta^2_p = .46$. There was no Frequency effect and no interaction.

Hit rates showed a large effect of Celebrity, $F(1, 34) = 89.36$, $p < .001$, $\eta^2_p = .72$, only a marginal effect of Frequency, $F(1, 34) = 3.74$, $p = .062$, $\eta^2_p = .10$, and no interaction, $F < 1$.

False alarm rates were affected by Frequency, $F(1, 34) = 20.66$, $p < .001$, $\eta^2_p = .38$, with no Celebrity effect and no interaction (both $p > .10$). More false alarms were made to frequent than to infrequent names.

Remember responses. Remember responses to old items were affected by both Frequency, $F(1, 34) = 21.67$, $p < .001$, $\eta^2_p = .39$, and Celebrity, $F(1, 34) = 199.59$, $p < .001$, $\eta^2_p = .85$, with no interaction.

Figure 2. Experiment 2. Values of d' (filled symbols) and C (unfilled symbols); circles: famous names; diamonds: nonfamous. Frequency is on the x-axis. Error bars show ± 1 standard error.

Remember responses to new items were quite few, but interesting, considering their incompatibility with threshold theories of remembering. They were more numerous to frequent names than to infrequent, Frequency, $F(1, 34) = 5.45$, $p = .026$, $\eta_p^2 = .14$. Celebrity had no effect, alone or in interaction.

Know responses. Know responses to old items, computed according to IRK assumptions (Jacoby et al., 1998) reflected the same pattern as remember responses, i.e., they were affected by both Frequency, $F(1, 34) = 12.82$, $p = .001$, $\eta_p^2 = .27$, and Celebrity, $F(1, 34) = 141.36$, $p < .001$, $\eta_p^2 = .81$, with no interaction.

Know responses to new items reflected the pattern of all types of false alarms in being more common to frequent names than to infrequent: Frequency, $F(1, 34) = 7.95$, $p = .008$, $\eta_p^2 = .19$. There were no other effects.

By items. In the interest of brevity, we report only regression analyses over items, not the ANOVA resulting from dichotomising the independent variables, although the latter was also performed and showed the same

pattern as the by-subjects analysis, i.e., hit rates were affected by both Celebrity and Frequency, but false alarm rates were affected by Frequency alone.

In the first regression analysis, hit rate was designated the dependent variable, and the potential predictors—(a) the log-transformed number of hits in the Internet search of the telephone directory (*freq_hits*), and (b) the log-transformed number of hits in the Internet search of the media sites (*celeb_hits*)—were introduced in a stepwise analysis. Both predictors were accepted as significant, *celeb_hits*: $\beta = .53$, $t(285) = 10.79$, $p < .001$; *freq_hits*: $\beta = -.18$, $t(285) = -3.64$, $p < .001$.

In a similar analysis, using false alarm rate as the dependent variable, only frequency was found to be a significant predictor, *freq_hits*: $\beta = .31$, $t(285) = 5.48$, $p < .001$.

Discussion

Using a partly different name set, we replicated the basic findings from the first experiment; names are better retained if they are famous beforehand, but famous or not, common names fare worse than unusual ones. Attempting to resolve this paradox of foreknowledge, we found that hit rates and false alarm rates showed different patterns of effects. Dissociation of hits and false alarms is a hallmark of dual-process theories of recognition. Whatever the particular brand of dual-process, a shared assumption is that recollection is active only in recognising old items and can have very little effect on false alarms. The FA rate is instead shaped by familiarity, i.e., familiar items attract more false recognition responses. The quality of familiarity may or may not be helpful in making true recognition responses as well, and the particular blend of recollection and familiarity that goes into shaping the hit rate is specific to each task and context.

So far, we have traced a parallelism. Fame, like recollection, exerts its effect mainly on hit rates. Frequency, although a less potent force overall, is the dominant influence, and sometimes the only influence, on false alarm rates.

We pursued this parallelism further with remember-know methodology, speculating that the different qualities of recognition associated with frequency and celebrity might be introspectively accessible. We had some misgivings on that score, acknowledging the burden we placed upon the participants in distinguishing pre-experimental from experimental familiarity. Our misgivings proved to be well-founded. The pattern of effects was essentially the same over d' , remember and know responses, suggesting that participants used the different responses mainly to express different degrees of confidence, not different qualities of experience. Indeed, one active line of

criticism against remember-know research is that this is exactly what happens: “Know” expresses just a lower degree of confidence than “remember”, and apart from that, qualities of experience do not enter into the decision (Hirshman & Henzler, 1998). We do not wish to enter that debate, except to affirm that the remember-know distinction was less useful for our purposes.

Having had response confidence brought to our attention, we now turn to a class of methods based on explicit confidence responses, the ROC (receiver operating characteristics) approach.

EXPERIMENT 3

Experiment 3 was performed to examine the memory characteristics of the name stimuli over a set of criteria. By inviting responses graded from high to low confidence, information can be gleaned about the shape of the ROC curves.

The ROC is a function relating hit rates to false alarm rates over a range of criteria. It has received increasing interest as an indicator of memory processes in recent years, because it gives more detailed information about the variation of accuracy when different criteria are adopted. The usefulness of ROC data for memory studies has been pointed out by many (Heathcote, 2003), although there is no complete consensus on how analysis of such data should proceed. Many approaches refer to the z -transformed ROCS (i.e., data converted by the inverse of the cumulative normal distribution), because these will be roughly linear (see Figure 3). We will also refer to these for convenience, although our analyses do not use linear regression fitting, but maximum likelihood fitting of the original data (Harvey, 2005).

Three types of parameters can be extracted: first, the distance between the old and the new distributions, which is a measure of accuracy, akin to the conventional d' . In the z -transformed graphs, this is reflected in the intercept of the regression line. There is a choice between scaling the distance by the standard deviation of the new or the old distribution or a compromise (d_a), but either way, the interpretation is relatively straightforward. With our present data, we expect the intercept to reflect the effects of frequency and celebrity on accuracy, as documented in the first two experiments.

The second parameter is the slope of the regression line. It reflects the relation between the standard deviations of the new and the old distributions, and because the new distribution is often assumed to have $\sigma = 1$, the slope is simply the inverse of the SD of the old distribution, $1/\sigma_{\text{old}}$. In some studies, slope has been seen to decrease with increased accuracy (Glanzer, Kim, Hilford, & Adams, 1999), and in others it has remained constant across conditions (Ratcliff, McKoon, & Tindall, 1994). There is still an ongoing debate about the proposed constancy-of-slopes generalisation

Figure 3. ROC curves (upper half) and z -transformed ROC-curves (lower half) for Experiment 3. Markers: circles: low frequency, low celebrity; diamonds: low frequency, high celebrity; triangles: high frequency, low celebrity; stars: high frequency, high celebrity.

(Ratcliff et al., 1994). In the present study, slopes will also be examined for effects paralleling those on intercepts, such that wherever accuracy increases, slope decreases. If, on the other hand, the constancy of slopes assumption is correct, we will see no change, or a greatly attenuated decrease. Our interest in the present study is not primarily in this matter, but in the possible dissociation of two types of prior experience. Therefore, we will attend to the

effects of frequency and celebrity on slopes, and in particular whether they are parallel or divergent.

The third type of parameter is the shape of the ROC. Although difficult to capture in a single quantity, the shape reflects possible deviations from the assumed signal detection model. Some theories of recognition memory explicitly posit such deviations. Dual process theories (Yonelinas, 2002) assume that a recollection process sometimes introduces an all-or-none element into the otherwise graded and probabilistic recognition process. This produces a concave-upward bent on the otherwise linear z -ROC. On the other hand, other processes of an artefactual character, such as high-confidence guessing, can produce a concave-downward deviation from linearity. In fact, this type of downward curvature is arguably more common (Glanzer et al., 1999; Heathcote, 2003) than the upward curvature that signals dual-process (but see, e.g., Arndt & Reder, 2002). We will not pursue the matter further, except to note that we observed a slight downward curvature in some of our conditions (see Figures 3 and 5).

Regardless of what type of parameter one chooses to focus on, model fitting can proceed along different paths. We will sketch two modes of operation. The first—and standard—mode operates on individual data for each experimental condition separately. It produces parameter estimates that are readily accessible and can easily be submitted to conventional statistical analysis. It has the drawback that single-subject data are relatively sparse, and several cells in the data sheet may be empty. Such analysis is sensitive to noisy data and may even introduce a systematic bias (Schooler & Shiffrin, 2005).

An alternative is to operate on aggregated data from all participants, i.e., a group ROC. This eliminates problematic discontinuities and safeguards against undue influence of minor artefacts. The question arises, however, as to how the variability, or conversely, the reliability of an effect is to be estimated. A solution can be found in recently developed, computer intensive bootstrapping methods (Martinez & Martinez, 2002; Schooler & Shiffrin, 2005). By resampling with replacement from the original data, new samples are constructed in which variability reflects that of the raw data. With enough resampling (typically > 1000 times), confidence intervals for any parameter of interest can be arrived at. Thus, contrasts corresponding to ANOVA effects can be computed, and the corresponding confidence intervals can be examined as to whether they include zero or not. Alternatively, another recent statistical development can be put to use. Killeen (2005) has proposed an alternative to null hypothesis testing in p_{rep} , the probability of replication. It is defined as the probability of a new experiment of similar power arriving at an effect of the same sign. Effects conventionally described as significant typically have a p_{rep} of .90 or more.

The p_{rep} is easily computed from bootstrapping samples by tallying the number of same-sign effects.

Orthogonally to the group versus single-subject question, a choice has to be made whether to fit each condition separately (e.g., deep vs. shallow study) or all conditions simultaneously. Not all software allows the second choice, but L. Harvey's Rscore program (Harvey, 2005) permits a large number of signal conditions. The first choice is the path most often taken, but the second one uses the data more efficiently. The empirical fact speaking in favour of simultaneous fitting is the finding that people tend to use the same criteria in judging stimuli throughout a session. In fact, some evidence suggests they are extremely reluctant to change their criteria even when different classes of words are marked with different colours (Stretch & Wixted, 1998). Treating each condition separately does not take advantage of these shared criteria and may result in suboptimal model fitting. However, the evidence concerning shared criteria is complex (Benjamin & Bawa, 2004; Dobbins & Kroll, 2005) and does not at this point allow definitive conclusions. In the present study, we present the outcome of both the standard and the aggregated approaches, as they have been described here. This means that we fit a model with shared criteria in the aggregated approach, and use separate criteria in the standard approach. If we can arrive at similar conclusions along these different paths, the assumptions concerning criteria placement are probably not decisive.

Method

Procedure

This experiment was performed as a classroom experiment, in which the stimuli were presented on a large screen, using PowerPoint and a projector, at a rate of 2 s per name. In the ensuing memory test, participants rated names on a 6-point scale, ranging from "I am sure the name was shown" to "I am sure it was not shown before", marking their choices in a booklet. The experiment was divided into two blocks, each containing 64 studied names and a test where they were mixed with 64 distractors. The four types of names were present in equal proportions.

Participants and materials

The experiment took place on two occasions, in classes at two different universities, as part of courses on memory. On the first occasion, 16 participants were tested, and on the second, another 12 (two of which were excluded because of low scores, 65% correct being the cutoff for inclusion). Mean ages were 44 and 27 years, respectively. The stimulus materials were slightly different; the first occasion used an expanded and modified set from

Experiment 1; the second occasion used names drawn from the final set used in Experiments 2 and 4. Despite these slight differences, we analysed the two sets of data together, for simplicity of presentation. Group was included as a between-subjects variable in the analyses, and it turned out that the pattern of effects was quite similar, as shown by the absence of interactions with the Group factor. There were no main or interaction effects pertaining to this factor.

Analysis

Hit rates and false alarm rates were computed to allow comparison with the other experiments. The middle point of the rating scale was used as the criterion.

ROC curves were plotted both on probability axes and on z -transformed axes (see Figure 3). Fitting the ROCs was performed by L. Harvey's program RscorePlus (Harvey, 2005), which uses maximum likelihood estimation.

Results

Variables were submitted to 2 (Frequency) $\times$ 2 (Celebrity) $\times$ 2 (Group), the latter being a between-participant factor, introduced because tests were performed on two occasions with slightly different materials. It will be mentioned only if significant.

Hit rates were affected by both Celebrity, $F(1, 24) = 67.46$, $p < .001$, $\eta_p^2 = .74$, and Frequency, $F(1, 24) = 17.17$, $p < .001$, $\eta_p^2 = .42$, and there was also an interaction between these two factors, $F(1, 24) = 6.02$, $p = .022$, $\eta_p^2 = .20$.

False alarm rates were affected by Frequency, $F(1, 24) = 13.22$, $p = .001$, $\eta_p^2 = .36$, and Celebrity, $F(1, 24) = 6.47$, $p = .018$, $\eta_p^2 = .21$, and the interaction between them, $F(1, 24) = 4.42$, $p = .046$, $\eta_p^2 = .16$ (Table 4).

Standard ROC analysis. The measure of accuracy for ROCS, d_a , which is similar to d' , showed effects of Frequency, $F(1, 24) = 55.84$, $p < .001$, $\eta_p^2 = .70$, and Celebrity, $F(1, 24) = 96.39$, $p < .001$, $\eta_p^2 = .80$, as well as their interaction, $F(1, 24) = 8.50$, $p = .008$, $\eta_p^2 = .26$. The interaction was caused by a larger effect of frequency among nonfamous than among famous names.

The standard deviation for the old distribution, which equals the inverse of the slope of the z -ROC, was tested in a similar ANOVA. It showed a main effect of Celebrity, $F(1, 24) = 4.31$, $p < .049$, $\eta_p^2 = .15$, and no other effect.

Aggregated analysis. Using Matlab's bootstrap function, 10,000 samples of size $n = 26$ were drawn with replacement from the data. Each sample gave

TABLE 4
Experiment 3

	<i>Means</i>					
	<i>Low frequency</i>		<i>High frequency</i>		<i>Effect sizes, η_G^2</i>	
	<i>Nonfamous</i>	<i>Famous</i>	<i>Nonfamous</i>	<i>Famous</i>	<i>Frequency</i>	<i>Celebrity</i>
HR	0.69	0.86	0.55	0.83	0.16	0.55
FAR	0.16	0.11	0.27	0.17	0.18	0.14
d_a	1.43	2.20	0.67	1.84	0.39	0.66
<i>SD</i> of old dist.	1.48	1.75	1.42	1.73	0.00	0.06
Mirror effect patterns	HR1	HR2	FAR2	FAR1		
Celebrity	.85	.62	.22	.14		
Frequency	.78	.69	.22	.14		

d_a is a measure of accuracy, the ROC equivalent of d' . *SD* is the standard deviation of the old distribution, i.e., the inverse of slope. ROCs have been fitted to individual data for each condition separately with the maximum likelihood method. For the lower part of the table (the mirror effect patterns), see note to Table 1.

rise to an averaged group ROC, which was submitted to the RScorePlus (Harvey, 2005) program with instructions to fit data for all four types of stimuli together. The result is illustrated in Figure 4.

Of the 10,000 samples, 93% gave acceptable fits ($p > .05$) of the model and were used. From the output of the program, eight means and eight standard deviations (Old/New $\times$ High/Low Frequency $\times$ High/Low Celebrity) were extracted for each sample and used to compute two contrasts (Low minus High Frequency and High minus Low Celebrity) for both intercepts, $(\mu_{old} - \mu_{new})/\sigma_{old}$, and slopes, $\sigma_{new}/\sigma_{old}$. The vectors of contrast values were sorted and the 2.5 and 97.5 percentiles were identified. The numbers of positive and negative values were tallied, and p_{rep} was computed as the proportion having the same sign as the mean.

Slopes. The Frequency contrast showed an average close to zero: 0.03 (95% CI: -0.31 to 0.37), and a p_{rep} of .57, close to chance. Celebrity, on the other hand, averaged -0.31 (95% CI: -0.63 to 0.04), with a p_{rep} of .96, a reliable effect. Slopes were lower in the High Celebrity conditions.

Intercepts. Both Frequency, $m = 0.89$, $p_{rep} = .999$, and Celebrity, $m = 1.42$, $p_{rep} = 1.00$, had highly replicable effects on intercepts, i.e., on accuracy. This fact confirmed findings from the earlier experiments.

Figure 4. Experiment 3. New (dark) and old distributions of the four simultaneously fitted conditions. Vertical lines mark the criteria.

Discussion

ROC curves traced the same trends concerning accuracy as the earlier experiments, now over a wider range of criteria. Famous names were better retained than nonfamous ones, and infrequent names held an advantage over frequent ones. As before, hit rates were determined by both Celebrity and Frequency, more so by the former. False alarm rates were also influenced by both, but more so by the latter.

The shape and locations of the distributions showed, in general terms, a mirror effect, i.e., those types of stimuli that had great memory strength when they were old, were weak when they were new—i.e., positioned far to the left on the familiarity/memory strength axis. This held true for both famous versus nonfamous names, and for infrequent versus frequent names.

There was a difference between the two factors in their effects on the slopes of the z -ROCs. Celebrity affected the slopes, and frequency did not. There has been debate concerning the degree to which slopes vary with accuracy (Heathcote, 2003). One finding has been that variations in materials that affect accuracy also affect slopes. Other, manipulated experimental variables such as study time, often affect accuracy while leaving slopes unchanged.

We found that one aspect of prior experience, the specific knowledge associated with famous names, had an effect on slopes, whereas the nonspecific kind did not. This could be due to the fact that high celebrity raises the level of memory strength—and with it the standard deviation—specifically in the old distribution without affecting the new. High familiarity, on the other hand, raises the familiarity of both the old and the new distributions—and with it the standard deviation—leaving the ratio unaffected.

EXPERIMENT 4

The fourth experiment aimed at reproducing the effects, especially the ROC data, in a new sample of individually tested participants. With the large material divided into shorter blocks, with computer administration and individual testing, we hoped to improve the quality of the ROC data by encouraging use of the full scale of response categories.

Method

Experiment 4 was conducted in the context of ERP (event-related potentials) recording, and electrophysiological data will be reported elsewhere. The focus here is on the behavioural responses and the ROC analyses based on them.

Participants

Twenty-four students at Lund University (14 women) participated in the experiment, which was conducted in a laboratory at the University Hospital. Each participant was tested individually during an approximately 1-hour long session, and received a cinema ticket voucher in compensation. Age of participants averaged 24.8, with a range of 19–42.

Procedure

The pool of 288 names provided stimuli for four blocks, each with 36 studied names and 36 distractors. The four study–test blocks were run

consecutively with subject-terminated pauses between them. Assignment of names to study or distractor status was made randomly for each participant, subject to the constraint that each of the four types of names be represented equally in every block.

Stimuli were presented on a computer screen by an E-prime program. Each name was displayed for 2 s in the study phase, preceded by a 1 s fixation cross. In the test phase, a name was first shown for 2 s while responding was disabled (to give ERPs without motor artefacts). Then a prompt appeared above the name ("Have you seen this name before in the experiment?"), and two response buttons appeared below it, marked "Yes" and "No". A mouse click on either response button terminated this display, which was followed by a new prompt ("How sure are you?") with three response buttons, marked "Quite sure", "Relatively sure", and "Not sure". A maximum of 7 s was allowed for response selection, but a response terminated the display, usually much sooner.

Analysis

A 6-point scale, ranging from "Quite sure new" to "Quite sure old" was constructed from the responses and used to plot the ROC and *z*-ROC graphs in Figure 5. Hit rates and false alarm rates were computed, and the ROC data were further analysed.

Harvey's RScorePlus (Harvey, 2005) extracted the parameters d_a and SD for each condition and each participant in the standard analysis. Further, using the aggregated scores of the whole group, it fitted all four conditions simultaneously, resulting in the data on which Figure 6 is based. Bootstrapping produced 10,000 samples, which were analysed as in the previous experiment.

Results

Hit rates were affected by Frequency, $F(1, 23) = 32.82$, $p < .001$, $\eta_p^2 = .59$, and Celebrity, $F(1, 23) = 62.64$, $p < .001$, $\eta_p^2 = .73$, and their interaction, $F(1, 23) = 18.47$, $p < .001$, $\eta_p^2 = .45$. False alarm rates were affected by Frequency, $F(1, 23) = 27.51$, $p < .001$, $\eta_p^2 = .55$, and Celebrity, $F(1, 24) = 10.36$, $p = .004$, $\eta_p^2 = .31$, with no interaction (Table 5).

Standard analysis. The measure of accuracy for ROCs, d_a , the analogue of d' , showed effects of Frequency, $F(1, 23) = 46.41$, $p < .001$, $\eta_p^2 = .67$, and Celebrity, $F(1, 23) = 81.21$, $p < .001$, $\eta_p^2 = .78$, as well as their interaction, $F(1, 23) = 5.09$, $p = .034$, $\eta_p^2 = .18$. As before, famous and infrequent names

Figure 5. ROC curves (upper half) and z -transformed ROC-curves (lower half) for Experiment 4. Markers: circles: low frequency, low celebrity; diamonds: low frequency, high celebrity; triangles: high frequency, low celebrity; stars: high frequency, high celebrity.

were better remembered than their nonfamous and frequent counterparts. The interaction indicated a larger frequency effect among the nonfamous.

Standard deviation of the old distribution, the inverse of slope of the z -ROCs, showed only an effect of Celebrity, $F(1, 23) = 5.26$, $p = .031$, $\eta_p^2 = .19$, all other F s < 1 .

Aggregated. Of 10,000 bootstrapped samples, 94% gave acceptable fits. Contrasts for high-low celebrity and low-high frequency were computed, the order of the terms arranged to place higher performance first.

Figure 6. Experiment 4. New (dark) and old distributions of the four simultaneously fitted conditions. Vertical lines mark the criteria.

Slopes. The replicability was only marginal, $p_{rep} = .86$ for both Frequency and Celebrity, whereas .90 could probably be considered the lower limit for a significant effect. The direction of the effect was such that high celebrity produced lower slopes (and higher accuracy). The direction of the Frequency effect, on the other hand, was such that low frequency produced *higher* slopes (and higher accuracy). The means were +0.17 (frequency) and -0.18 (celebrity).

Intercepts. Both contrasts evidenced highly reproducible effects; both $p_{rep} = 1.00$. Means of the contrasts were 0.98 and 1.32, for Frequency and Celebrity, respectively.

TABLE 5
Data from Experiment 4

	<i>Means</i>					
	<i>Low frequency</i>		<i>High frequency</i>		<i>Effect sizes, η_G^2</i>	
	<i>Nonfamous</i>	<i>Famous</i>	<i>Nonfamous</i>	<i>Famous</i>	<i>Frequency</i>	<i>Celebrity</i>
HR	0.76	0.89	0.61	0.85	0.27	0.60
FAR	0.11	0.07	0.15	0.12	0.25	0.13
d_a	1.78	2.50	1.18	2.18	0.34	0.65
SD of old distribution	1.37	1.57	1.38	1.84	0.01	0.05
Mirror effect patterns	HR1	HR2	FAR2	FAR1		
Celebrity	.87	.69	.13	.10		
Frequency	.83	.73	.14	.09		

For the lower part of the table, see note to Table 1.

Item analysis. In a regression analysis over items, hit rate was entered as the dependent variable, and the potential predictors—(a) the log-transformed number of hits in the Internet search of the telephone directory (*freq_hits*), and (b) the log-transformed number of hits in the Internet search of the media sites (*celeb_hits*)—were introduced in a stepwise analysis. Both predictors were accepted as significant, *celeb_hits*: $\beta = .50$, $t(285) = 10.24$, $p < .001$; *freq_hits*: $\beta = -.30$, $t(285) = -6.27$, $p < .001$.

In a similar analysis, using false alarm rate as the dependent variable, frequency was found to be a significant predictor, *freq_hits*: $\beta = .29$, $t(285) = 5.10$, $p < .001$, and so was *celeb_hits*: $\beta = -.14$, $t(285) = -2.43$, $p = .016$.

Discussion

As in all the other experiments, we found a partial dissociation, in that frequency affected false alarm rates more than celebrity did, and celebrity affected hit rates more than frequency did. This held true for both analysis over subjects and over items.

Thus, although both factors had large effects on net accuracy (d_a), the patterns of effects on the components of performance were different. As to the ROC data, celebrity had an impact on the standard deviations of the distributions in the standard analysis, whereas frequency had none. (In the aggregated analysis, both had only weak effects, however, in opposite directions.) We interpret this as indicating that high frequency impairs performance by raising the level—and importantly, the variance—of familiarity in both old and new distributions.

In other words, we find our results to be compatible with a pattern where frequency primarily raises the variance of old and new distributions alike, and impairs performance as a result. Celebrity, on the other hand, raises mean and variance of the old distribution specifically, improving performance in the process.

GENERAL DISCUSSION

We studied the effects of two types of prior experience on the memorability of names. Both were environmental variables, presumably related to experiential frequency, such that higher values on the two variables meant higher probabilities of encountering the names in daily life, given the relatively homogeneous cultural environment of our participants. The variables—name frequency and celebrity—were measured both by environmental statistics and by participant ratings, with satisfactory reliability.

Although similar in their relation to experiential frequency, the variables had completely different effects on memory accuracy. The effects were different not only in size, but more strikingly, in direction. Increasing name frequency lowered accuracy, a pattern reminiscent of the much-studied word frequency effect. Increasing name celebrity, on the other hand, raised accuracy. In this latter respect, our data resembled earlier studies, where knowledge of the stimulus domain has been seen to improve accuracy. Examples include superior memory for chess positions in chess masters (de Groot, Gobet, & Jongman, 1996), and the memorial advantage for words taken from a student's major field of study (Allen & Garton, 1968; Chalmers et al., 1997). Wine experts show superior recognition memory for wine-related odours (Parr, White, & Heatherbell, 2004), and experienced golfers show enhanced memory for specific putts, but only when routinisation of the putting task is disturbed (Beilock, Wierenga, & Carr, 2002).

The divergent effects of prior experience are probably quite general, rather than specific to names. The pattern of sharpened memory within preferred domains of knowledge can arguably be ascribed to increased opportunities for elaborative encoding, although this would need independent evidence. The other pattern is perhaps more counterintuitive, although we can easily find examples in daily life of the familiarity that breeds disregard. In fact, many memory failures are the result of poor encoding of run-of-the-mill events, or neglect by habit. As the history of the word frequency effect shows, this common phenomenon has proven recalcitrant for several memory theories.

Recently, neuroscience has turned up evidence (Fernandez & Tendolkar, 2006) that a part of the medial temporal lobe, the rhinal cortex, acts as a gatekeeper to the elaborated encoding orchestrated by the hippocampus.

If familiarity is high, as judged by the rhinal cortex, the event is deemed uninteresting and it is denied access to deepened encoding. If news value is high, on the other hand, all the facilities of the memory system are called upon to engrave the new event. In a recent fMRI study, stimuli that had been primed beforehand, underwent less encoding and showed depressed retention in relation to novel stimuli (Wagner, Maril, & Schacter, 2000).

As the present data show, some well-known objects escape the ban of the gatekeeper. If a name is famous, be it ever so common, it enjoys the privilege of the newsworthy, and gets to be encoded deeply. What determines whether an object, although familiar, can pass the scrutiny of the gatekeeper? We have proposed that specific semantic knowledge, i.e., the fact that it is individuated, is decisive. If a name is ripe with unique and detailed associations, a web of potential retrieval cues can be established at encoding, possibly bound together by hippocampal pointers. If, in addition, rhinal cortex judges the item to be novel, this seal of approval further facilitates encoding. The fact that we found fame and novelty to interact overadditively in Experiments 1 and 2 suggests that they both affect the same processing stage, possibly the hippocampal encoding mechanism.

Apart from the difference in direction of the general memory effect, we also found that the two types of prior experience had different profiles in their impact on memory components. This can be summarised in two points:

- Celebrity influenced hit rates more than frequency did, and frequency influenced false alarm rates more than celebrity did.
- Celebrity tended to have some effect on ROC slopes, but frequency did not.

The first of these points was evident in the measures of effect size, which showed consistency across the four experiments. However, to test the statistical significance of this fact further, we formed the contrast of high versus low celebrity, and low versus high frequency and computed these contrasts over hit rates and false alarm rates for each individual. The outcome is shown in Figure 7. Furthermore, 2×2 ANOVAs were computed with these contrast scores, and all four experiments evidenced the critical effect, a significant interaction (all ps at least $< .02$), showing the celebrity contrast being larger for hit rates, and the frequency contrast being larger for false alarm rates. The mean replicability of this effect, p_{rep} (Killeen, 2005), is better than .97. The fact that hit rates and false alarm rates can be dissociated has been interpreted as evidence in favour of two-process theories of recognition memory (Reder et al., 2000).

In the examination of ROCs, we found indications that celebrity affected slopes, but frequency did not. Earlier literature has discussed why some variables (especially those related to materials) increase both accuracy and

Figure 7. Means of the contrasts high-low celebrity and high-low frequency, computed for hit rates and false alarm rates in the four experiments.

ROC slopes, whereas others affect accuracy without changing slopes, and no definitive consensus has been reached (Heathcote, 2003). In any event, the finding adds to the evidence that the two variables exert different effects.

The frequency effect can tentatively be characterised as belonging to semantic memory, or possibly to a very long-term form of conceptual priming. This is not denying the fact that there are related phenomena in the animal learning literature. Conditioning can be impeded by *Kamin blocking*. This refers to the finding that a stimulus can be rendered ineffective as elicitor of a conditioned response with which it is paired, if it (the stimulus) has been familiarised beforehand. In anthropomorphic terms, the animal discounts the stimulus as being useless as a predictor, because of its prior experience with it in a noncausal role. A very similar phenomenon, *latent inhibition*, has been examined in some human learning studies (Lubow & Gewirtz, 1995) that have demonstrated how stimuli can be rendered ineffective for learning (i.e. conditioning) by frequent presentations before conditioning starts, evidently a case of the familiarity that breeds discounting.

A recent theory has awarded familiarity a third place in the memory hierarchy, subordinate to declarative memory, but on an equal footing with semantic and episodic memory (Moscovitch et al., 2005), and Fernandez and Tendolkar (2006) have reviewed the evidence for the rhinal cortex acting as a familiarity detector. The phenomena of familiarity versus novelty are evidently very important in regulating encoding resources up and down in

the course of everyday experience, thereby deciding what becomes preserved in memory and what disappears without a trace.

In summary, we have found indications of two types of memory, tentatively described as semantic. They are both related to the prior experience of names, but they exert diametrically opposed effects on episodic recognition. One type is related to general familiarity. At the time of study, it tends to inhibit deep encoding by signalling ordinairiness, thereby denying access to encoding resources. At the time of retrieval, it confuses the recogniser by engendering a sense of familiarity that is hard to distinguish from the sought-after memory. By this mechanism, the false alarm rate is raised. Further, it raises the level of variability in both the old and the new distributions.

The distinctive type of semantic memory, on the other hand, resembles knowledge of subject matter, or expertise within a stimulus domain. It facilitates encoding by providing links to individuating features, highly specific associations that serve to make the memory distinctive. This effect on encoding raises hit rates without notably affecting false alarm rates.

Original manuscript received December 2006

Revised manuscript received March 2007

First published online July 2007

REFERENCES

- Allen, L. R., & Garton, R. F. (1968). the influence of word-knowledge on the word-frequency effect in recognition memory. *Psychonomic Science*, 10(12), 401–402.
- Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2(6), 396–408.
- Anderson, J. R., & Schooler, L. J. (2000). The adaptive nature of memory. In E. Tulving & F. Craik (Eds.), *The Oxford handbook of memory*. Oxford, UK: Oxford University Press.
- Arndt, J., & Reder, L. M. (2002). Word frequency and receiver operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(5), 830–842.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379–384.
- Bartlett, F. (1932). *Remembering*. Oxford, UK: Oxford University Press.
- Beilock, S. L., Wierenga, S. A., & Carr, T. H. (2002). Expertise, attention, and memory in sensorimotor skill execution: Impact of novel task constraints on dual-task performance and episodic memory. *Quarterly Journal of Experimental Psychology: Section A*, 55(4), 1211–1241.
- Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language*, 51(2), 159–172.
- Chalmers, K. A., & Humphreys, M. S. (1998). Role of generalized and episode specific memories in the word frequency effect in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(3), 610–632.
- Chalmers, K. A., Humphreys, M. S., & Dennis, S. (1997). A naturalistic study of the word frequency effect in episodic recognition. *Memory and Cognition*, 25(6), 780–784.

- Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour*, 12, 335–359.
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin and Review*, 3(1), 37–60.
- De Groot, A. D., Gobet, F., & Jongman, R. W. (1996). *Perception and memory in chess: Studies in the heuristics of the professional eye*. Assen, The Netherlands: Van Gorcum & Co.
- Dennis, S. (1995). The impact of the environment on the word frequency and the null list strength effects in recognition memory [Electronic version]. *Noetica*, 1. Retrieved 19 March 2007 from <http://www2.psy.uq.edu.au/CogPsych/Noetica/Issue1/Environment.html>
- Dobbins, I. G., & Kroll, N. E. A. (2005). Distinctiveness and the recognition mirror effect: Evidence for an item-based criterion placement heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(6), 1186–1198.
- Estes, W. K., & Maddox, W. T. (2002). On the processes underlying stimulus-familiarity effects in recognition of words and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(6), 1003–1018.
- Fernandez, G., & Tendolcar, I. (2006). The rhinal cortex: “Gatekeeper” of the declarative memory system. *Trends in Cognitive Sciences*, 10(8), 358–362.
- Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory and Cognition*, 13(1), 8–20.
- Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(1), 5–16.
- Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, 100(3), 546–567.
- Glanzer, M., Kim, K., Hilford, A., & Adams, J. K. (1999). Slope of the receiver-operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(2), 500–513.
- Harvey, L. O. (2005). RscorePlus: A Macintosh/Windows program that fits a multiple-distribution, variable-criterion signal detection theory model to confidence-rating data (Ver. 5) [Computer program]. http://psych.colorado.edu/~lharvey/Software%20Zip%20Files/RscorePlus_Win.zip
- Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(6), 1210–1230.
- Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, 9(1), 61–65.
- Jacoby, L. L., Jones, T. C., & Dolan, P. O. (1998). Two effects of repetition: Support for a dual-process model of knowledge judgments and exclusion errors. *Psychonomic Bulletin and Review*, 5(4), 705–709.
- Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica*, 98(2–3), 127–140.
- Killeen, P. R. (2005). An alternative to null-hypothesis significance tests. *Psychological Science*, 16(5), 345–354.
- Lubow, R. E., & Gewirtz, J. C. (1995). Latent inhibition in humans: Data, theory, and implications for schizophrenia. *Psychological Bulletin*, 117(1), 87–103.
- Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(3), 539–559.
- Martinez, W. L., & Martinez, A. R. (2002). *Computational statistics handbook with Matlab*. London: Chapman Hall.
- Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., et al. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: A unified account based on multiple trace theory. *Journal of Anatomy*, 207(1), 35–66.

- Parr, W. V., White, K. G., & Heatherbell, D. A. (2004). Exploring the nature of wine expertise: What underlies wine experts' olfactory recognition memory advantage? *Food Quality and Preference*, 15(5), 411–421.
- Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, 41(3), 416–426.
- Rajaram, S. (1996). Perceptual effects on remembering: Recollective processes in picture recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(2), 365–377.
- Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(4), 763–785.
- Reder, L. M., Nhuyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(2), 294–320.
- Schooler, L. J., & Shiffrin, R. M. (2005). Efficiently measuring recognition performance with sparse data. *Behavior Research Methods*, 37(1), 3–10.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117(1), 34–50.
- Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1397–1410.
- Wagner, A. D., Maril, A., & Schacter, D. L. (2000). Interactions between forms of memory: When priming hinders new episodic learning. *Journal of Cognitive Neuroscience*, 12(6), 52–61.
- Westmacott, R., & Moscovitch, M. (2003). The contribution of autobiographical significance to semantic memory. *Memory and Cognition*, 31(5), 761–774.
- Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin and Review*, 11(4), 616–641.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441–517.