Memory & Cognition
1997, 25 (1), 1–10

Fluency and response speed in recognition judgments

RUSSELL A. POLDRACK and GORDON D. LOGAN
University of Illinois at Urbana-Champaign, Urbana, Illinois

Previous research has suggested that perceptual fluency can contribute to recognition judgments. In this study, we examined whether fluency in recognition is based upon the speed of preceding operations, as suggested by studies of perceptual fluency. Subjects studied items in both lexical decision and naming tasks, and were then tested on two blocks of lexical decision trials with probe recognition trials. Jacoby's process dissociation procedure was used, and results from this procedure suggested that recognition judgments in the task were based largely upon familiarity. However, the estimated discriminability available from response time distributions was significantly less than the observed recognition discriminability. Simulated memory operating characteristics confirmed this underdetermination of recognition by response times. The results demonstrate, contrary to previous suggestions, that fluency in recognition is not based upon speed.

Subjective feelings of familiarity are an important basis for recognition memory (Jacoby, 1991; Mandler, 1980). In this paper we investigate a possible mechanism for the effects of familiarity on recognition judgments that follow other processing events. A number of studies have found that recognition judgments can be correlated with response times (RTs) to stimuli immediately preceding the recognition trial. The best example of this comes from studies of perceptual fluency, in which the time needed to identify a stimulus is correlated with the probability of calling that item "old" on a recognition test (e.g., Johnston, Dark, & Jacoby, 1985). Items that are identified quickly are called "old" more often than those that are difficult to perceive because the subject attributes this increased speed (or fluency) to memory rather than to, for example, more stable characteristics of the item, such as its frequency. We will refer to this as the speed hypothesis: the claim that attributions of response speed can support a significant level of recognition memory. To examine this hypothesis, we present a model of speed-based recognition and test this model in a lexical decision experiment.

Evidence for the Speed Hypothesis

Several researchers have investigated the relationship between recognition memory and response speed (Feustel, Shiffrin, & Salasoo, 1983; Johnston et al., 1985; Johnston, Hawley, & Elliot, 1991; Watkins & Gibson, 1988). In these experiments, subjects first study a set of stimuli. At test, for each studied item and an equal number of new items, subjects must first identify the item under difficult perceptual conditions (either identification at short presentation speeds or clarification under masking). Subjects then perform an immediate recognition judgment on the item in clear view (we will refer to this as a successive recognition task). The repetition effect (see, e.g., Scarborough, Cortese, & Scarborough, 1977) leads to better (faster, more accurate) identification of those items that are repeated from the study list. The speed hypothesis predicts that items that are identified more quickly will be called "old" regardless of their actual recognition status. Since old items are processed more quickly than new items, most such decisions will be correct, but the fastest new items will be incorrectly called "old."

A number of studies have yielded evidence in favor of the speed hypothesis. Johnston et al. (1985) and Feustel et al. (1983) both found that items identified more quickly in a clarification task were more likely to be called "old"; most importantly, they found that false alarms ("old" responses to new items) were greatest for those new items identified most quickly. Johnston et al. found that this was most evident for nonwords at long repetition lags, where explicit memory would be weakest, suggesting that fluency-based recognition is dissociated from explicit recognition.

Watkins and Gibson (1988) claimed that the correlation between identification time and recognition performance was due to item effects, arguing that the items that were more easily identified were also more familiar because of stable characteristics such as word frequency, leading to a correlation. Johnston et al. (1991) provided evidence against this claim by showing that the effect occurred only when recognition judgments followed identification immediately and did not occur when identification trials and

R.P. is now at Stanford University. This research was supported in part by National Science Foundation grant BNS-91-09856 to G.D.L., and by a predoctoral National Research Service Award from NIMH to R.P. We would like to thank D. Patel and E. Ponnezhan for assistance in testing subjects. We would like to thank N. Cohen, G. Dell, E. Dzhafarov, G. Murphy, and E. Shoben for their helpful discussions of this research, and L. Jacoby, W. Johnston, M. Masson, and B. Whittlesea for helpful comments on an earlier version of the paper. Correspondence should be addressed to R. A. Poldrack, Department of Psychology, Jordan Hall, Stanford University, Stanford CA 94305 (e-mail: poldrack@psych.stanford.edu).

recognition trials were presented in a blocked manner. This showed that the correlation depended uniquely upon the ability of the subject to attribute response speed to familiarity.

However, the evidence for the speed hypothesis is not conclusive. Johnston et al. (1991) and Watkins and Gibson (1988) attempted to create simulated fluency by intentionally speeding the identification of some items in the perceptual identification task. Both were unable to create a fluency effect on recognition using this manipulation. Johnston et al. also attempted to create artificial fluency by speeding responses using semantic priming and were similarly unsuccessful; Whittlesea (1993) was able to produce fluency on a recognition test using a conceptual fluency manipulation (presenting a target word in a strong context), and the effect was much stronger when overall recognition performance was low. Whittlesea, Jacoby, and Girard (1990) also succeeded in creating artificial effects of fluency using different levels of visual masking at test; in subjects who were not aware of the differences between mask densities, there was a small (but significant) increase in "old" responses for items at the lower mask density (cf. Whittlesea, 1993). The most reasonable conclusion to be drawn from these studies is that the perceptual fluency effect on recognition is somewhat fragile, is most likely produced when fluency arises out of variations in normal processing, and is most likely to occur when overall recognition performance is relatively low (cf. Johnston et al., 1991), whereas conceptual fluency effects may be more robust.

Arguments Against the Speed Hypothesis

Successive recognition studies have demonstrated a correlation between response speed and recognition judgments. However, these studies have not provided any measure of the relative amount of recognition discrimination that could be contributed by speed. For example, Johnston et al. (1985) found a significant relationship between identification speed and successive recognition. However, the differences in identification latency for "old" and "new" responses, and for old and new items, were quite small with respect to the standard deviation of identification times. In making judgments based upon single responses, subjects would be limited in effectiveness by the standard deviation of the distribution of RTs. Although an experimenter can describe a distribution with respect to its mean and standard error of the mean, subjects do not have the luxury of a sampling distribution; they must examine single values from the distribution, and thus their ability to discriminate between *old* and *new* items depends upon the variability of single items. It seems unlikely that the observed difference between old and new RT distributions in the Johnston et al. (1985) study could support the level of discrimination that was observed in their successive recognition task, given the level of variability (cf. Ratcliff, 1993). In the experiment presented here, we investigated this issue more thoroughly.

A Model of Speed-Based Recognition

We developed a simple process model of recognition based upon RTs in order to measure the degree to which recognition judgments could be based upon speed. This model is similar in character to the signal detection theory (SDT) model of recognition memory (Lockhart & Murdock, 1970), where recognition is accomplished by setting a criterion along a distribution of familiarity. We simply defined familiarity in this model in terms of processing speed in order to test the claim that recognition could operate on the basis of speed.

In order to make a recognition judgment under the speed model, subjects compare a single processing time to a criterion for saying "old." For a perceptual identification task with successive recognition, this processing time would be the identification time for the item immediately preceding the recognition trial. In our task, this was the lexical decision time for the trial immediately preceding the recognition judgment. Subjects set a criterion on the RT dimension for making *yes* judgments. If the processing time is faster than the criterion, the subject responds "old." If the time is slower than the criterion, the subject responds "new" (Figure 1). Discriminability between old and new items is then determined by the distance between the processing time distributions for old and new items, relative to their variability.

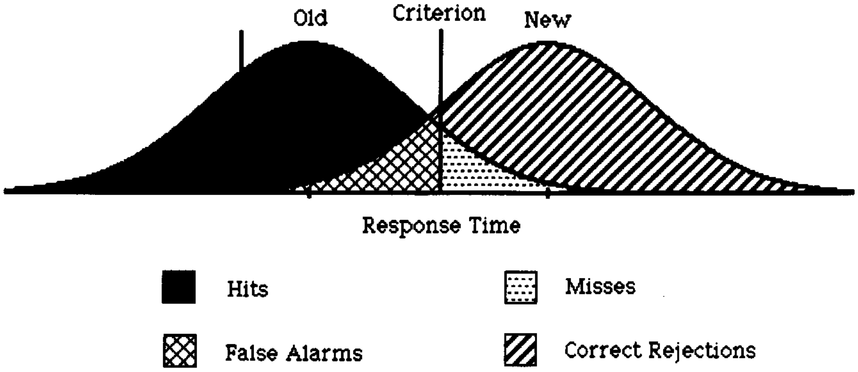

Figure 1. Illustration of response time distributions with fixed criterion, and the associated recognition outcomes according to the speed model.

***dRT*.** Within the model, a measure of discriminability between old and new RT distributions was developed, on direct analogy to the d' measure used in SDT. d' is defined as

$$d' = \frac{F_{\text{old}} - F_{\text{new}}}{SD_{\text{new}}}, \tag{1}$$

where F_{old} and F_{new} represent means of unobservable distributions of familiarity for old and new items respectively, and SD_{new} represents the standard deviation of familiarity for new items (Green & Swets, 1965). Intuitively, d' quantifies the distance between the standardized distributions of familiarity for old and new items, which translates directly into recognition performance. Our measure, called *dRT*, allows the analogous quantification of the amount of discriminability available from RTs, and was defined as follows:

$$dRT = \frac{M_{\text{new}} - M_{\text{old}}}{(SD_{\text{old}}\,/\,2 + SD_{\text{new}}\,/\,2)}, \tag{2}$$

where *M*s were the mean RTs and *SD*s are the standard deviations for RTs. By computing the percentage of observed d' that is accounted for by *dRT*, we can get a quantitative estimate of the degree to which recognition performance could possibly rely upon response speed (cf. Jacoby, 1991).

A critical feature of this measure was the use of the standard deviation in the denominator. When subjects make a judgment about the relative slowness of a single response compared to the overall distribution of response times, they have only one observation with which to make that judgment. The variation of the single sample for a subject in the model, then, is simply the variation of the distribution itself, rather than the standard error of the experimenter's sampling distribution of mean RT. This convention is also used in the definition of d' in SDT (Green & Swets, 1965).

Potential problems with *dRT*. The *dRT* measure suffers from two potential problems. First, the calculation of *dRT* as a measure of discriminability assumes that the old and new RT distributions are normal with equal variances. The normality assumption fails for RT distributions, which are positively skewed with long right tails. RTs have been described as convolutions of normal and exponential distributions (Ratcliff & Murdock, 1976), Weibull distributions (Logan, 1992), or gamma distributions (Luce, 1986) and clearly differ from the normal distribution. In addition, new items generally have more variation than do repeated items (Compton & Logan, 1991; Logan, 1988, 1992), resulting in failure to satisfy the equal variance assumption.

Another problem lies with the contributions of *nonfluent* processes to RTs. A *nonfluent process* is any process that does not show an item-specific speed-up but does contribute to RTs. Such processes as motor programming and execution are possibly nonfluent; they probably speed up with practice, but this speed-up might generalize to both old and new stimuli. To the extent that nonfluent processes add to the variation of the RTs but only fluent processes are responsible for the difference between old and new items as measured by *dRT*, *dRT* will underestimate the discriminability available from fluency. The components of variance of RT can be stated as follows:

$$SD^2_{\text{obs}} = SD^2_{\text{flu}} + SD^2_{\text{nonflu}}, \tag{3}$$

where these *SD*s are averaged over old and new items. Computation of *dRT* should then properly proceed using only the variation due to fluent processes in the denominator:

$$dRT = \frac{M_{\text{new}} - M_{\text{old}}}{SD_{\text{flu}}}. \tag{4}$$

From Equation 4, we derived the predicted level of SD_{flu} that would be necessary to support the observed level of recognition performance (d'_{obs}):

$$\text{Predicted } SD_{\text{flu}} = \frac{M_{\text{new}} - M_{\text{old}}}{d'_{\text{obs}}}. \tag{5}$$

To the extent that this value accounts for a small proportion of the observed variability, we can conclude that speed is inadequate to account for recognition judgments.[1]

Memory operating characteristics. In order to address the problems with the validity of *dRT* as a measure of discriminability, we computed simulated memory operating characteristics (MOCs) using observed RT distributions. MOCs represent the discriminability between old and new items over a range of levels of bias toward saying "old" (i.e., a range of criterion values) and are produced by plotting the *Z*-score for hit rate [$p(Hit)$] against the *Z*-score for the false alarm rate [$p(FA)$]. MOCs can be produced empirically by varying the hit rate using a payoff schedule, or by plotting $p(Hit)$ as a function of confidence ratings. We produced simulated MOCs from RT distributions as follows:

1. Specify a level of $p(Hit)$. In our computations, $p(Hit)$ took four possible values: {0.2, 0.4, 0.6, 0.8}.
2. Determine the value on the old item RT distribution that cut off $p(Hit)$ percent of the old item RT distribution. This value was taken as the criterion, C, for responding "old" on the recognition test. If subjects said "old" to all items as fast as or faster than C, they would have a hit rate of $p(Hit)$.
3. Determine the proportion of the new item RT distribution that was cut off by the criterion, C. This percentage is taken as probability of a false alarm, $p(FA$; Figure 1).
4. Complete Steps (2) and (3) for each level of $p(Hit)$.
5. Compute the *Z* values at each level for $p(Hit)$ and $p(FA)$, and construct an MOC by plotting $Z(p(Hit))$ against $Z(p(FA))$.

The level of discriminability afforded by the RT distributions is evident in the distance of the least squares regression line fitting each MOC from the unit line; the intercept of this best fitting line is an independent estimate of *dRT*.[2] The shape and slope of the MOC depends upon the shape of the underlying response time distributions; deviations from normality result in nonlinear MOCs, while differences in variation between old and new response time distributions result in MOCs with slopes different from one.

The Present Experiment

Here we report an experiment in which we tested the speed hypothesis using the model outlined above. This experiment extended the test of the speed hypothesis beyond the perceptual identification studies reported earlier to a processing task (lexical decision) where the nature of the RT distributions is more clearly understood (e.g., Logan, 1988). The experiment tested the speed hypothesis in a lexical decision task with successive recognition judgments, where some words were repeated at test. Lexical decision is a task that shows robust repetition effects (see, e.g., Logan, 1990; Scarborough et al., 1977); our model was applied to measure the degree to which these repetition effects could support the observed level of recognition. Jacoby's (1991) opposition procedure was employed as an independent measure of the degree to which recognition in the task is based upon familiarity.

In our experiment, subjects first studied words and nonwords in two study phases: lexical decision and naming. Subjects were then tested using a lexical decision task. On about 40% of trials, a recognition probe followed the lexical decision trial immediately. The speed hypothesis predicts that these recognition judgments should be based upon the lexical decision RT from the immediately preceding trial. In this case, recognition accuracy should be directly related to the size of the repetition effect for old items. To the degree that the process dissociation procedure shows that the task is based upon familiarity, failure of the speed hypothesis means that these familiarity effects must be based on something other than attributions of response speed.

METHOD

Participants

Forty-eight undergraduate students from the University of Illinois participated in the experiment as part of a class requirement. Twenty-four subjects participated in an inclusion condition, and 24 participated in an exclusion condition.

Materials

Words were selected randomly for each subject from a pool of 339 common words or matched nonwords, and assigned randomly to conditions. For each subject, two sets of 48 study stimuli (24 words and 24 nonwords) were chosen. One set was presented as stimuli in a lexical decision study block, and the other set was presented as stimuli in a naming study task. In addition, each study list included an additional 5 primacy and 10 recency buffer items. The words were nouns selected from the Kučera and Francis (1967) word frequency norms, with a mean frequency of 75.27 per million and a range of 8 to 787 per million. All nonwords were produced by changing one or two letters from the matched word, and all were pronounceable (see Logan, 1990, for more details on the stimulus set).

Procedure

Stimuli were displayed in lowercase on a Dell Super VGA monitor, and stimulus presentation and response collection were controlled by a Dell 210 microcomputer. Subjects were instructed that they would see words and nonwords, and would make lexical decision, naming, and recognition responses to these stimuli. Subjects responded in the lexical decision task by pressing the "Z" or "X" keys on the keyboard with their preferred hand, and made recognition responses vocally. The experiment consisted of one session lasting approximately 35–40 min.

The session began with 48 lexical decision study trials, on which the subject performed only lexical decision. The 48 target study items (24 words and 24 nonwords) were surrounded by 5 primacy and 10 recency buffer items, which were not repeated thereafter. After a short break, subjects performed 48 naming study trials (24 words and 24 nonwords), which were surrounded by 10 primacy and 10 recency items that did not appear again. On naming trials, subjects pronounced stimuli into a microphone that triggered a voice-activated relay. An experimenter controlled the presentation of stimuli and recorded the naming accuracy of the subject.

Following the study trials, subjects performed two blocks of lexical decision trials. Each block consisted of 96 studied items (which appeared in each block) and 96 new items (which were unique to each block). On a subset of trials, a recognition probe was presented immediately after the lexical decision response. Probes were presented following 40 new items (20 words and 20 nonwords), 20 old items (10 words and 10 nonwords) from the lexical decision study block and 20 old items (10 words and 10 nonwords) from the naming study block. (Because of a program error, 10 new words, rather than 20, were probed in the second test block.) Subjects in the inclusion condition were instructed to say "old" on recognition probes if they remembered having studied the item in either of the study blocks. Subjects in the exclusion condition were instructed to say "old" only to items that they remembered from the naming study block; they were instructed to say "new" if they remembered having studied the item in the lexical decision study block.

On each trial, a 500-msec signal ("+") was followed by the presentation of the word or nonword. The stimulus remained on the screen until the subject made the lexical decision task or naming response. On the probed recognition trials, the stimulus remained on the screen and the question "Old or New?" appeared; subjects responded vocally, and an experimenter in the room recorded the recognition responses. Subjects were instructed that speed and accuracy were both important, and they were told to concentrate on the lexical decision part of the experiment and not to try to anticipate the recognition probes. The computer emitted a beep if the subject made a lexical decision error.

RESULTS AND DISCUSSION

The alpha level for all statistical tests was .05. RTs greater than 1.5 *SD*s above the mean were excluded from the analysis (a total of 8.5% of trials).

Response Time

Study. Mean study RTs are presented in Table 1. These data were analyzed using a 2 (group: inclusion vs. exclusion) × 2 (study block: lexical decision vs. naming) × 2

Table 1
Mean Study Response Times for Words and Nonwords Averaged Over Subject Groups

	Item Type			
	Words		Nonwords	
Study Block	*M*	*SE*	*M*	*SE*
Lexical decision	613	15.0	708	20.1
Naming	566	21.5	641	40.0

Table 2
Mean Test Response Times for Words and Nonwords

	Study Type					
	Old (Lexical Decision)		Old (Naming)		New	
Block	*M*	*SE*	*M*	*SE*	*M*	*SE*
			Inclusion			
1						
Word	635	19.8	641	19.6	667	19.5
Nonword	719	21.9	720	24.2	694	21.3
2						
Word	624	20.7	616	18.5	649	20.2
Nonword	699	22.0	703	23.0	680	20.3
			Exclusion			
1						
Word	709	28.6	717	26.7	733	25.6
Nonword	798	34.0	828	37.6	779	30.5
2						
Word	670	27.0	684	29.1	715	25.4
Nonword	753	26.2	767	25.6	726	24.8

(lexicality: word vs. nonword) mixed analysis of variance (ANOVA). There was a significant effect of study block [$F(1,46) = 9.03$, $MS_e = 17{,}399.3$], reflecting faster responses on naming trials than on lexical decision trials. There was also a significant effect of lexicality [$F(1,46) = 44.55$, $MS_e = 7{,}834.9$], reflecting faster responses for words than for nonwords. No other effects were significant ($ps > .1$).

Test. Mean test RTs are presented in Table 2. These data were analyzed using a 2 (group: inclusion vs. exclusion) × 3 (study type: lexical decision, naming, and unstudied) × 2 (lexicality: word vs. nonword) × 2 (test blocks) mixed ANOVA. There was a significant effect of group [$F(1,46) = 4.72$, $MS_e = 147{,}238$]; subjects in the inclusion task were faster overall than subjects in the exclusion task. There was a significant effect of test block [$F(1,46) = 15.48$, $MS_e = 8{,}079.1$], showing a decrease in RTs from Block 1 to Block 2. The effect of lexicality was significant [$F(1,46) = 77.24$, $MS_e = 8{,}365.4$], reflecting the fact that responses to words were faster than responses for nonwords. The lexicality × list interaction was significant [$F(2,92) = 23.22$, $MS_e = 2{,}341.0$]. This effect reflected the fact that studied words were faster than new words but studied nonwords were slower than new nonwords. This negative repetition effect for nonwords is likely due to the role of familiarity in lexical decision; repeating a nonword makes it more familiar, which makes the item harder to reject as a nonword (see, e.g., Feustel et al., 1983). No other effects were significant ($ps > .1$).

Accuracy

Study. Study accuracy data are presented in Table 3. These data were analyzed using a 2 (group: inclusion vs. exclusion) × 2 (study block: lexical decision vs. naming) × 2 (lexicality: word vs. nonword) mixed ANOVA. There was a significant effect of study block [$F(1,46) = 11.68$, $MS_e = 0.004$], reflecting greater accuracy in the naming block than in the lexical decision block. There was also a significant effect of lexicality [$F(1,46) = 16.25$, $MS_e = 0.005$]; responses to words were more accurate than responses to nonwords in both study blocks. The interaction was not significant.

Test. Test accuracy data are presented in Table 4. These data were analyzed using a 2 (group: inclusion vs. exclusion) × 2 (test block) × 2 (lexicality: word vs. nonword) × 3 (study type: lexical decision, naming, or new) mixed ANOVA. There was a significant effect of lexicality [$F(1,46) = 6.39$, $MS_e = 0.004$], reflecting greater accuracy for words than for nonwords. The effect of study type was significant [$F(2,92) = 4.02$, $MS_e = 0.002$]. Planned comparisons showed that old items from the lexical decision study block had higher accuracy at test than did new items ($p < .01$); no other comparisons were significant. The lexicality × list interaction was significant [$F(2,92) = 24.82$, $MS_e = 0.001$], reflecting the fact that studied words were more accurate than new words, but studied nonwords were less accurate than new nonwords. These data rule out a speed–accuracy tradeoff as an explanation for repetition effects. The four-way interaction was marginally significant ($p < .06$); the meaning of this interaction is unclear. No other effects were significant ($ps > .1$).

Recognition

Recognition results were analyzed using the d' measure of discriminability (Green & Swets, 1965). These data

Table 3
Mean Study Accuracy for Words and Nonwords, Averaged Over Subject Groups

	Item Type			
	Words		Nonwords	
Study Block	*M*	*SE*	*M*	*SE*
Lexical decision	0.955	0.008	0.915	0.014
Naming	0.986	0.004	0.946	0.009

Table 4
Mean Test Accuracy (Proportion Correct) for Words and Nonwords

Block	Old (Lexical Decision) M	Old (Lexical Decision) SE	Old (Naming) M	Old (Naming) SE	New M	New SE
			Inclusion			
1						
Word	0.973	0.006	0.973	0.005	0.952	0.009
Nonword	0.952	0.012	0.932	0.020	0.960	0.008
2						
Word	0.988	0.004	0.978	0.006	0.941	0.008
Nonword	0.944	0.013	0.951	0.010	0.964	0.008
			Exclusion			
1						
Word	0.987	0.006	0.975	0.008	0.951	0.008
Nonword	0.949	0.014	0.957	0.005	0.970	0.006
2						
Word	0.988	0.005	0.977	0.006	0.947	0.008
Nonword	0.968	0.010	0.953	0.009	0.970	0.007

are presented in Table 5. The data were analyzed using a 2 (group: inclusion vs. exclusion) × 2 (lexicality) × 2 (study task: lexical decision vs. naming) × 2 (test block) mixed ANOVA. The main effect of block was significant [$F(1,46) = 25.25$, $MS_e = 0.61$], reflecting greater discriminability in the second test block than in the first. This main effect was accompanied by a block × group interaction [$F(1,46) = 15.98$, $MS_e = 0.61$]. This effect reflects a greater increase in d' across test blocks for inclusion than exclusion subjects. The main effect of study task was significant [$F(1,46) = 20.40$, $MS_e = 0.41$], showing greater discriminability for items studied in the naming block than for items studied in the lexical decision block. The main effect of lexicality was also significant [$F(1.46) = 4.11$, $MS_e = 0.55$]; recognition was better for words than for nonwords. Both of these effects are qualified by a significant study task × lexicality interaction [$F(1,46) = 8.06$, $MS_e = 0.29$], revealing a greater difference between study tasks for nonwords than for words.

Two higher order interactions were also present in the d' analysis. The significant block × lexicality × study task interaction [$F(1,46) = 11.97$, $MS_e = 0.27$] reflected the fact that the increase in d' from Block 1 to Block 2 was much smaller for nonwords studied in the naming block than for other lexicality × study task combinations. The significant block × lexicality × group interaction [$F(1,46) = 7.99$, $MS_e = 0.69$] reflected the fact that the increase in d' from Block 1 to Block 2 was not different for subject groups for nonwords, but was much greater for the inclusion condition than for the exclusion condition for words. No other effects were significant ($ps > .1$).

Process Dissociation

The process dissociation procedure (Jacoby, 1991) was used to estimate the roles of fluency and recollection in

Table 5
d' and dRT Values for Inclusion and Exclusion Groups by Block and Study Type

Block	Words d'	Words SE	Words dRT	Words SE	Words % d'	Nonwords d'	Nonwords SE	Nonwords dRT	Nonwords SE	Nonwords % d'
					Inclusion					
1										
Lexical decision	1.29	0.093	0.35	0.078	27.1	1.14	0.129	−0.18	0.043	n/a
Naming	1.31	0.127	0.28	0.072	21.4	1.74	0.230	−0.22	0.071	n/a
2										
Lexical decision	2.29	0.220	0.26	0.105	11.4	1.78	0.164	−0.16	0.060	n/a
Naming	2.46	0.230	0.33	0.105	13.4	1.84	0.139	−0.23	0.076	n/a
					Exclusion					
1										
Lexical decision	1.53	0.156	0.20	0.077	13.1	0.92	0.126	−0.09	0.084	n/a
Naming	1.55	0.182	0.14	0.072	9.0	1.70	0.159	−0.29	0.079	n/a
2										
Lexical decision	1.31	0.204	0.38	0.099	29.0	1.34	0.140	−0.18	0.085	n/a
Naming	1.67	0.193	0.33	0.094	19.8	1.70	0.139	−0.31	0.072	n/a

Note—% of d' was not calculated for nonwords because dRT was negative, making the measure meaningless for these data.

Table 6
Process Dissociation Procedure Results by Study Type and Block

Block	Words		Nonwords	
	Recollection	Familiarity	Recollection	Familiarity
1	0.196	0.596	0.125	0.486
2	0.438	0.689	0.333	0.675

recognition performance. These data are presented in Table 6. This analysis showed that recognition performance in the present experiment had a large familiarity component that overshadowed the recollection component in every block for both words and nonwords, and by a factor of more than three in the first block. The relative contribution of fluency decreased in the second test block, probably due to repetitions of studied items from the first test block. These data demonstrate that a familiarity-based process (i.e., fluency) supports recognition judgments in the task.

Speed Model Analyses

Having found that familiarity was the predominant basis of recognition judgments in the task, we examined the degree to which this familiarity might be based upon attributions of response speed. We first calculated *dRT* values according to Equation 2; these are presented in Table 5 along with the observed d' values. The difference between *dRT* and d' indicates the degree to which the observed recognition performance could possibly be supported by response speed. The ratio of *dRT* to d' is also presented in Table 5. These values are uniformly small and suggest that response speed could have supported very little of the observed recognition performance. However, because recognition performance was at least partially attributable to recollection, these ratios may underestimate the proportion of fluency-based recognition discriminability that can be explained by RT differences.

The speed model is also undermined by a dissociation between RTs and recognition results for words and nonwords. Repetition effects for words were positive, leading to predictions of above-chance recognition performance. However, repetition effects for nonwords were negative, in which case the speed model predicts below-chance recognition accuracy. These predictions were not supported by the recognition data. Although d' for nonwords was significantly less than d' for words overall, performance for nonwords was still well above chance in all conditions.

The relationship between recognition and speed was examined directly by comparing *dRT* and d' in a repeated measures ANOVA with lexicality as an additional within-subjects factor. The main effect of measure (d' vs. *dRT*) was significant [$F(1,47) = 330.1, MS_e = 0.354$]; d' values were greater ($M = 1.60$) than *dRT* values ($M = 0.04$). The effect of lexicality was also significant [$F(1,47) = 44.26, MS_e = 0.112$], reflecting greater d' and *dRT* values for words than for nonwords. The measure × lexicality interaction was also significant [$F(1,47) = 13.26, MS_e = 0.103$]. The difference between *dRT* and d' was greater for nonwords than for words; however, d' significantly exceeded *dRT* for both words and nonwords ($ps < .001$).

In order to further examine the speed hypothesis, we compared the observed variability of RTs to the variability that would be necessary in order to support the observed recognition performance, according to the fluency model (Equation 5). Predicted and observed standard deviations were computed for each subject; these values are presented in Table 7, along with the proportion of observed variance accounted for by the predicted variance. These values were compared using a 2 (predicted vs. observed) × 2 (block) × 2 (group: inclusion vs. exclusion) mixed ANOVA. There was a significant main effect of predicted versus observed [$F(1,43) = 80.08, MS_e = 6{,}562.8$]. Predicted *SD* was significantly lower ($M = 20.8$) than observed *SD* ($M = 130.2$). No other effects were significant. Thus, the observed *SD* values were outside of the range of values necessary to support speed-based recognition performance.

Memory operating characteristics. To confirm the *dRT* analysis, simulated memory operating characteristics (MOCs) were constructed from the test RT distributions for items studied in the lexical decision study block. These are presented in Figures 2 and 3 for inclusion and exclusion groups separately, along with a point representing observed recognition performance and the best fitting linear function for the MOC. In each case, these points fell far outside the MOC, showing that the observed level of recognition discriminability was greater than that available from the RT distributions. The y-intercept of the linear MOC function (in normalized coordinates) is an estimate of *dRT* (Green & Swets, 1965); these estimates were similar in magnitude to those obtained from Equation 2, and in all cases fell outside the confidence intervals for observed recognition d'. It is interesting to note that the operating characteristics obtained in the present study were linear with a slope very near the unit slope. Although this finding is difficult to interpret (Lockhart & Murdock, 1970), it does suggest that there were no large-scale violations of the assumptions underlying the

Table 7
The Predicted Level of SD_{flu} (Standard Deviation of the Fluency Distribution) Required to Support the Observed Recognition Performance, Along With the Observed Standard Deviations for Words and the Percentage of Variance Accounted for

Block	*SD*		% Variance
	Predicted	Observed	Accounted for
		Inclusion	
1	27.5	112.7	6.0%
2	16.9	118.6	2.0%
		Exclusion	
1	22.4	152.5	2.2%
2	17.1	133.3	1.6%

Note—Nonword data are not included because the observed repetition effects were negative.

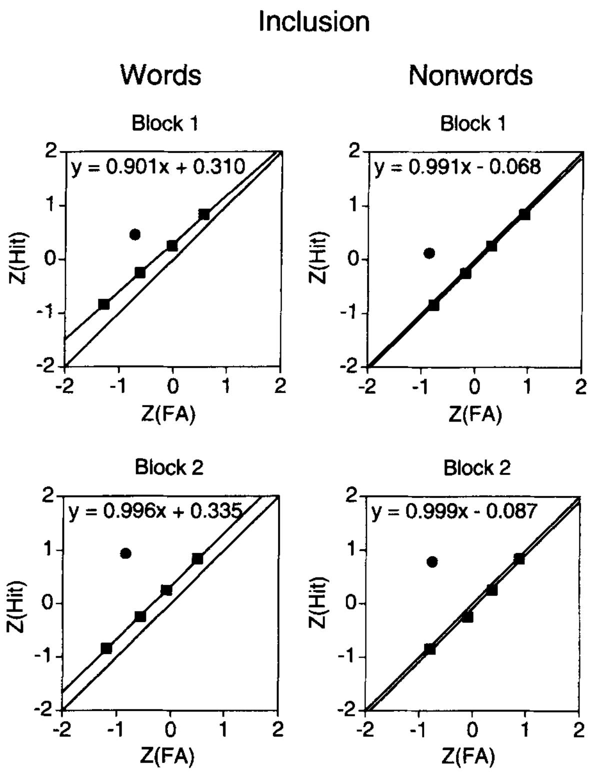

Figure 2. Simulated memory operating characteristics (MOCs) for inclusion group by block and lexicality, along with observed recognition memory performance (denoted by a circle). The equation for the best fitting linear function is included for each MOC.

normalized operating characteristic. The MOC data show that, although familiarity played a large part in recognition judgments, little of this performance can be accounted for in terms of speed.

GENERAL DISCUSSION

In a successive recognition study, we found that recognition performance was largely dependent upon familiarity according to Jacoby's (1991) process dissociation procedure. However, our model-based analyses showed that the observed recognition performance could not be supported by attributions of response speed to familiarity. Using the *dRT* measure of discrimination by fluency, we found that speed could account for only a small proportion of the observed recognition discrimination for both words and nonwords. The construction of MOCs confirmed the findings of the *dRT* analysis. We conclude that response speed cannot underlie the fluency effect in successive recognition tasks.

What Might Fluency Be?

Our results show that response speed can support only small portions of observed recognition performance. Against our conclusions one might argue that fluency is based upon something other than response speed. However, studies of fluency and recognition have defined fluency experimentally in terms of RTs (e.g., Johnston, Dark, & Jacoby, 1985; Johnston, Hawley, & Elliott, 1991); such a redefinition would throw out an important literature on fluency and recognition. In addition, the definition of fluency in terms of speed is advantageous because it allows one to bring to bear a great deal of previous work on RTs to this problem (e.g., Luce, 1986).

We might ask, however, how else (other than response speed) fluency could be represented within the existing range of RT models. One possibility is that fluency could be based upon the speed of a given processing stage rather than on the entire RT (Ratcliff, 1993). This might allow the subject greater discriminability than that which is possible from overall RTs (by excluding the variation arising from these nonfluent processing stages). The nonword results in the present experiment argue in favor of this sort of model; fluency would arise from the processing stage in which repeated words or nonwords are facilitated, excluding the decision stage in which repeated nonwords are slowed. Examination of this possibility would entail the formidable task of constructing a model of the processing stages involved in the task and then using the finishing time distributions from one stage to predict recognition performance. Alternatively, in a latent

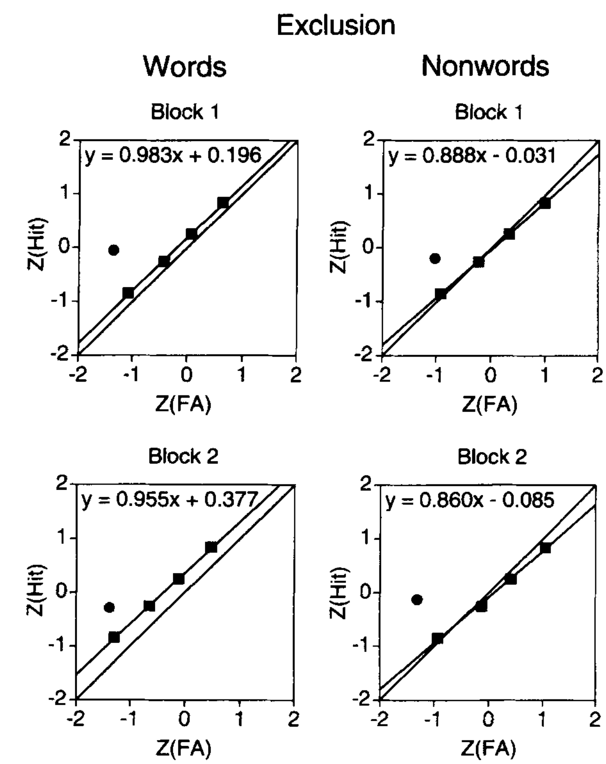

Figure 3. Simulated memory operating characteristics (MOCs) for exclusion group by block and lexicality, along with observed recognition memory performance (denoted by a circle). The equation for the best fitting linear function is included for each MOC.

network theory of RTs (Schweickert, 1978), fluency might be accounted for in terms of slack, which is the amount of time by which a process can be lengthened without affecting overall RT. Changes in the amount of slack might be observable by the subject without affecting observed RTs, with greater slack being interpreted as increased fluency. Further investigation is needed to examine these possibilities.

Recent studies of the neural basis of repetition priming also suggest likely possibilities for the underlying basis of fluency effects. Neural imaging studies have shown that priming is associated with decreases in regional cerebral blood flow to the cortical areas involved in performing a task, for example, occipital areas for visual tasks (Buckner et al., 1995) and frontal areas for conceptual tasks (Demb et al., 1995. It is currently unclear how these changes relate to the phenomenological experience of fluency, but the data do suggest that subjects might experience changes in different forms of processing as qualitatively different forms of fluency. Event-related potential studies of repetition priming also suggest bases for the qualitative experience of fluency. Repetition leads to an attenuation of the N400 potential, which is thought to reflect semantic processing or semantic memory retrieval (Rugg, 1990), and this repetition effect differs in its topographic distribution for within-modality and between-modality repetition (Domalski, Smith, & Halgren, 1991); this modality effect suggests that the effect may have both perceptual and conceptual components that could be experienced separately. These findings are in line with the imaging results in suggesting that fluency might result from the phenomenal experience of the reduced level of processing that accompanies stimulus repetition. Thus, subjects may attribute a reduced amount of processing either to prior experience or to other sources, depending upon the constraints of the task being performed (Jacoby, Kelley, & Dywan, 1989; Whittlesea, 1993).

Formalizing the Concept of "Fluency"

The discovery of a procedure by which to separate the effects of familiarity and recollection in recognition tasks has led to a consensus that both are independently important in recognition (Jacoby, 1991), and a primary source of familiarity is attributions of fluency. The basis of fluency effects, however, has remained markedly unspecified. By ruling out speed as a foundation for the effects of fluency, our findings suggest that theories of memory need to specify new mechanisms by which fluency might occur. Defining fluency in terms of intuitive concepts like "ease," without specifying the mechanism by which this fluency is generated, results in theories that defy testing and falsification. The analytic method laid out in the testing of our speed model can perhaps provide a framework for testing other alternative theories of fluency. In addition, our analysis is applicable to testing any theory that involves comparisons between RT distributions. Further work with this model should bring increased rigor to studies of theoretical concepts like fluency.

REFERENCES

Buckner, R., Peterson, S. E., Ojemann, J. G., Miezin, F. M., Squire, L. R., & Raichle, M. E. (1995). Functional anatomical studies of explicit and implicit memory retrieval tasks. *Journal of Neuroscience*, **15**, 12-29.

Compton, B. J., & Logan, G. D. (1991). The transition from algorithm to retrieval in memory-based theories of automaticity. *Memory & Cognition*, **19**, 151-158.

Demb, J. B., Desmond, J. E., Wagner, A. D., Stone, M., Lee, A. T., Glover, G. H., & Gabrieli, J. D. E. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *Journal of Neuroscience*, **15**, 5870-5878.

Domalski, P., Smith, M. E., & Halgren, E. (1991). Cross-modal repetition effects on the N4. *Psychological Sciences*, **2**, 173-178.

Feustel, T. C., Shiffrin, R. M., & Salasoo, A. (1983). Episodic and lexical contributions to the repetition effect in word identification. *Journal of Experimental Psychology: General*, **112**, 309-346.

Green, D. M., & Swets, J. A. (1965). *Signal detection theory and psychophysics*. New York: Wiley.

Grice, G. R. (1968). Stimulus intensity and response evocation. *Psychological Review*, **75**, 359-373.

Jacoby, L. L. (1991). A process dissociation framework: Separating intentional from automatic uses of memory. *Journal of Memory & Language*, **30**, 513-541.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391-422). Hillsdale, NJ: Erlbaum.

Johnston, W. A., Dark, V., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **11**, 3-11.

Johnston, W. A., Hawley, K. J., & Elliot, J. M. G. (1991). Contributions of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **17**, 210-233.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

Lockhart, R., & Murdock, B. B. (1970). Memory and the theory of signal detection. *Psychological Bulletin*, **74**, 100-109.

Logan, G. D. (1988). Toward an instance theory of automaticity. *Psychological Review*, **95**, 492-527.

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology*, **22**, 1-35.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 883-914.

Luce, R. D. (1986). *Response times*. New York: Oxford University Press.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, **114**, 510-532.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**,190-214.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition*, **18**, 367-379.

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 1-17.

Schweickert, R. (1978). A critical path generalization of the additive factor method: Analysis of a Stroop task. *Journal of Mathematical Psychology*, **18**, 105-139.

Watkins, M. J., & Gibson, J. M. (1988). On the relationship between perceptual priming and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 477-483.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1235-1253.

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory & Language*, **29**, 716-732.

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, **5**, 102-122.

NOTES

1. Ratcliff (1993) made the similar point that subjects might be able to make decisions about their own response speed on the basis of decision stages that are independent of output processing.

2. The simplest form of the fluency model would be one in which the criterion remains fixed across judgments. This fixed-criterion model predicts that all items called "old" should be faster than all items called "new," with a slight overlap possible due to variance in other components of processing. However, this prediction is called into question by the data from Johnston et al. (1985); the RT distributions for items called "old" and items called "new" showed a substantial degree of overlap. A variable-criterion model can account for the overlap in distributions. In this model, the subject's criterion for saying "old" varies according to some distribution (see, e.g., Grice, 1968). However, such a model is much less tractable analytically than the fixed-criterion model, because a stochastic component must be introduced. Wickelgren (1968) generalized SDT to the case of variable criteria and showed that discriminability in the variable criterion case will always be smaller than discriminability in the fixed-criterion case (because the variance of the criterion appears in the denominator of the d' calculation). This means that the fixed-criterion model is the limiting best case of all variable-criterion models. On the basis of this argument, we used the fixed-criterion model to generate MOCs.

(Manuscript received October 7, 1995;
revision accepted for publication February 13, 1996.)