Perceptual match effects in direct tests of memory: The role of contextual fan

LYNNE M. REDER, **DIMITRIOS K. DONAVOS**, and **MICHAEL A. ERICKSON**
Carnegie Mellon University, Pittsburgh, Pennsylvania

Abstract

The aim of the present study was to determine whether physical attributes of a memory representation would affect explicit memory performance and, if so, what type of factors would affect the size of a perceptual match effect. Subjects studied words in different, uncommon fonts and were later asked whether the word had been studied earlier. Words could be re-presented in the original font, a font studied with another word, or a font not seen earlier. In two additional experiments, we varied the numbers of words studied in the same unusual font (1 vs. 12 words per font). Recognition memory for the words was better if the test and study fonts matched, and this effect was larger for fonts not shared with other words. Moreover, old judgments were most likely to be classified as *remember* responses when words were re-presented in the same font when it had not been studied with other words. Although we found a significant effect of levels of processing, this factor did not interact with whether the font matched between study and test. These results are consistent with the predictions of the source of activation confusion model of memory and suggest that perceptual information operates according to the same memory principles as conceptual information.

There has been considerable interest in understanding the conditions under which perceptual similarity between the encoding stimulus and the test probe plays a critical role in subsequent memory performance (e.g., Cave, Bost, & Cobb, 1996; Curran, Schacter, & Bessenoff, 1996; Graf & Ryan, 1990; Light, LaVoie, Valencia-Laver, Owens, & Mead, 1992). Part of this interest may be attributable to the conjecture that perceptual similarity matters only for implicit memory (for reviews, see Richardson-Klavehn & Bjork, 1988; Roediger & McDermott, 1993, chap. 2). For example, both Jacoby (1983) and Rajaram and Roediger (1993) have shown that the closer the physical representation of the test probe to the representation of the target experienced during initial encoding, the better the performance on implicit (also known as indirect) tests of memory.

Repetition priming effects sometimes appear to be modulated by changes in type font (or changes between handwritten presentation of stimuli during study and typed presentation of stimuli at test; e.g., Jacoby & Hayman, 1987; Jacoby & Witherspoon, 1982; Roediger & Blaxton, 1987a, 1987b), whereas other times, level of facilitation in priming appears unaffected by such changes (e.g., Carr, Brown, & Charalambous, 1989; Clarke & Morton, 1983; Feustel, Shiffrin, & Salasoo, 1983; Graf & Levy, 1984). Roediger and McDermott (1993, chap. 2), after reviewing physical surface variations, including typography, concluded that "alterations in typography have much less effect than do changes in modality or symbolic form (pictures to words)" (p. 89).

In contrast, explicit (direct) tests of memory, such as recognition performance, did not seem to be enhanced by close physical matches between the item encoded during study and the test probe. Explicit memory tests, such as free recall, were helped more by prior exposure to a picture of the target than by seeing the printed word (Durgunoğlu & Roediger, 1987; Rajaram & Roediger, 1993; Weldon & Roediger, 1987). Overall, there has been less evidence that perceptual similarity between encoding and probe items affects explicit memory tests (e.g., Craik, Moscovitch, & McDowd, 1994; Murdock & Walker, 1969; Rajaram & Roediger, 1993), although the results of Graf and Ryan (1990) are a notable exception.

EXPLANATORY FRAMEWORKS

Why does physical similarity sometimes seem to help explicit memory? Understanding when and why physical similarity affects access to a memory trace has important theoretical implications for our understanding of the representation and processing of memory.

There has been considerable debate as to the appropriate explanatory framework to account for these memory dissociations—namely, that physical similarity seems so much more important for indirect than for direct tests of memory. On the one hand, some theorists have postulated that there exist qualitatively different memory *systems*—that is, an implicit and an explicit memory system—with different properties and mechanisms, so that the same variables could affect these two systems differentially (e.g., Curran et al., 1996; Schacter & Tulving, 1994; Tulving & Schacter, 1990). Perhaps the best known of these proposals involves the *perceptual representation system* (PRS) of Schacter and Tulving (1994), which states that the perceptual aspects of a memory experience are stored in the implicit memory system known as PRS, whereas the semantic aspects of an experience are stored in a different memory system.

On the other hand, another group of researchers has taken this same memory dissociation (role of physical similarity in the two types of memory tests) and argued that it is unnecessary to posit different memory systems; rather, they argue that these effects can be understood in terms of whether the encoding and the test processes engage the same set of cognitive operations (Kolers, 1975, 1979; Kolers & Ostry, 1974). This view, referred to as *transfer appropriate processing* (TAP), has been supported by results described in Blaxton (1989), Roediger, Srinivas, and Weldon (1989), and Graf and Ryan (1990). The crux of this challenge has been based on (1) finding dissociations among implicit memory tasks (e.g., Roediger et al., 1989) and (2) finding perceptual similarity effects in explicit memory experiments (e.g., Light et al., 1992).

Graf and Ryan (1990) have found that matching the font between encoding and test facilitated memory for both direct and indirect tests but that this effect was modulated by whether the encoding task focused on perceptual features (enhancing the effect) or conceptual features (minimizing the effect). It is important to note that Curran et al. (1996) found that perceptual specificity effects were unaffected by encoding task, thereby questioning the TAP account of these perceptual similarity effects. In sum, there appears to be considerable uncertainty in terms of when and why visual specificity effects occur in some explicit memory situations and not in others.

OUR PROPOSAL

In this paper, we will offer a new account of when and why perceptual match of features can enhance memory performance. Unlike the PRS view, our postulation is that perceptual information is represented and processed in a fashion analogous to that for conceptual information—for example, they strengthen and decay at the same rate and spread activation in the same way. Like the TAP framework, our belief is that allocation of attention affects how much information of a particular type is represented (e.g., perceptual vs. elaborative); however,

we posit that the perceptual match effects described above depend on the ability to access a stored memory trace, not to match encoding and retrieval processes. More specifically, we propose that the extent of the benefit of a perceptual match in a memory task depends on the relative distinctiveness of the perceptual features. It is important to operationalize distinctiveness, and that is done as part of the theoretical framework described below.

Distinctiveness in the Source of Activation Confusion Model of Memory

Before defining distinctiveness within the source of activation confusion (SAC) model, it is important to explain how recognition is assumed to operate. SAC adopts the notion adopted by a number of memory theorists (e.g., Jacoby & Dallas, 1981; Mandler, 1980) that recognition judgments can be based on either of two processes: familiarity or recollection. Details and formal specification of the representation and processing assumptions of this dual process theory can be found elsewhere (e.g., Reder, Angstadt, Cary, Ayers, & Erickson, 2002; Reder et al., 2000; Reder & Schunn, 1996; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997).

According to SAC, during the study phase of an experiment, a subject may lay down a trace encoding the fact that a particular word was studied in the experimental context (consult Figure 1). Later, when asked to recognize whether a word had been studied, the subject will respond affirmatively either by retrieving the trace that the word was studied in the experimental context or by inferring that the word was studied, owing to the sense of familiarity associated with the queried term. Familiarity is affected by the current level of activation of the concept node. Recollection requires that the *episode node* associated with the activated concept receive sufficient activation from the concept and context nodes to pass over threshold. The amount of activation sent from a source depends on the competition—that is, the number and strength of competing associations from the sending concept.

We posit that distinctiveness is inversely related to the number of contextual associations of the concept in question. In other words, concepts with *fewer contextual associations are more distinctive*. When there are fewer competing contextual associations, it is easier to retrieve the encoding context, because more of the source activation will flow back to the associated event node. Using implemented computational models, Reder et al. (2000) demonstrated that the recognition advantage of words of lower normative frequency can be explained as being a result of their fewer contextual associations.

Representing the Font of the Letters That Encode the Word

Figure 1 presents a schematic illustration of how we assume that font is represented in a visual presentation of a study word. A given font is represented as a single concept node that is linked to the episode node that encodes seeing a particular word in that font. The dashed line from that font node to the node for the word represents the *process* of perceiving the word in that font, not an associative memory connection. That process link is also schematic in that the connection from the font node must pass between the letters that are orthographic constituents of the lexical entry associated with that concept to the concept node for the word.

Our assertion of a generalized font concept has empirical support from Hofstadter (Hofstadter, 1985; Hofstadter & McGraw, 1995), who showed that people are able to generalize the form of other letters intended by the designer by a sample of the letters in that font. We assume that a font is represented as a node, just as one can represent semantic concepts—for example, a conceptual node for dogs. We hypothesize that these font nodes behave according to the same memory principles as all other types of information associated with a memory trace. They are strengthened in exactly the same way as any other type of concept, with associated benefits,

and are vulnerable to the same principles of interference inherent in other types of competing associations.

It is our claim that a common font, such as the one that is used to print this document, has a strong node to represent the font but is associated with so many other contexts besides this document that it is not distinctive (i.e., it has a very high contextual fan). Figure 1 also illustrates the hypothesized memory representation for words studied in a font that is used to present many words in the experiment. For space reasons, we give an example of two words studied with the same font ("Gadzoox") but have many lines emanating from the font node to indicate that more words were studied with it.

According to SAC, when a word is presented in a font at test, activation will spread to associated contexts from both sources (i.e., the concept and font nodes), as well as from the context node. If the font is the same as the one used during study, there will be a link from the font node to the relevant episode node. The fewer contextual associations out of the font node (i.e., the less *fan*), the more activation can spread to the relevant episode node, making it more likely to get over threshold to enable a recollection. The phenomenon that fan or competing associations will affect memory performance has been demonstrated many times (e.g., Anderson, 1974; Cantor & Engle, 1993; Kimberg, 1994; Lewis & Anderson, 1976; S. B. Peterson & Potts, 1982; Reder & Ross, 1983; Reder & Wible, 1984; Zbrodoff, 1995) and has been demonstrated in this theoretical framework, with words of different normative and experimental frequency (Reder et al., 2000), as well as with pseudowords (Reder et al., 2002). The idea that the fan of the font should matter has not been tested before.

The two fonts used by Graf and Ryan (1990) were more distinctive than Helvetica or Palatino, but their fonts were each used with half of the words studied during the experiment, so the fan out of each was very high. We hypothesize that if each font is relatively novel and is associated with only one word during study, the effect of matching font between study and test will be greater. In our first experiment, we sought to conduct a partial replication of Graf and Ryan. Like Graf and Ryan, we hoped to show that perceptual match effects can be obtained in direct tests of memory. They concluded that the orienting task modulated whether or not a perceptual match effect occurred. We predicted that when the effect of perceptual match was allowed to be stronger, by using fonts with less contextual fan, the perceptual match effect would obtain regardless of the orienting task during encoding.

We employed the same two orienting tasks as those used by Graf and Ryan (1990)—namely, judging legibility or pleasantness. A noncontroversial prediction in this partial replication is to find a main effect of orienting task, so that performance will be better if the words are encoded using conceptual processing, such as judging pleasantness, than if the encoding involves only a superficial task, such as judging legibility.

EXPERIMENT 1

Method

Subjects—The subjects were 30 Carnegie Mellon undergraduate s enrolled in psychology courses, who participated as partial fulfillment of a course requirement. Fifteen were randomly assigned to the legibility group, and 15 to the pleasantness group.

Design and Materials—The design consisted of three factors: orienting task during encoding (judging legibility or pleasantness of the word), word status (judged or not during encoding), and font-matching condition at test (original font, swapped font, or novel font). The first factor was manipulated between subjects, and the other two factors were manipulated within subjects. The third factor could not be crossed with both levels of the second because

original font was not defined for nonstudied words. During encoding, a font was seen with only one word, and each word was seen only once. During test, one half of the studied words were re-presented with their original encoding font and half were different, one fourth of the words presented with a swapped font, and one fourth with a novel (not previously used) font. In order to maintain the same proportion of old to new fonts for new items as that used for old items, three quarters of the new words were presented in an swapped font and one quarter in a novel font.

The design involved a total of 144 words and 108 distinctive fonts. Of the 144 words, one half were randomly sampled to make up the encoding list, and all were presented at test. All the words were concrete nouns 3–14 letters in length ($M = 5.4$) and of moderate frequency ($M = 75$, range, 40–120 occurrences per million according to Kuèera & Francis, 1967).

The 108 fonts were collected from various Internet sources and were selected on the basis of the following criteria: The fonts seemed distinctive (i.e., seemed dissimilar to other fonts selected for the experiment, as well as to common fonts, such as Helvetica), were easily readable, and were similar in size to the other fonts used in the experiment. Examples of fonts used in the experiment are shown in Figure 2.

For each subject, fonts were randomly assigned to words, and both words and fonts were randomly assigned to conditions. Since this random assignment was performed separately for each subject, any stimulus effects would become part of the subject error term.

Procedure—The subjects were tested individually in sessions that lasted approximately 30 min. The session consisted of two phases: an encoding phase and a word recognition test phase. Prior to the encoding phase, the subjects were told that the purpose of the experiment was to explore what attributes of a word make it either more or less legible (for the legibility group) or more or less pleasant (for the pleasantness group). They were not told that they were to remember the words for a subsequent test.

During the encoding phase, a series of 72 words, each presented in a different, unique font, was presented to the subjects, who were asked to judge them. In the legibility group, they were instructed to "ignore the meaning of the word and to only pay attention to the font that the word was presented in." The subjects in the pleasantness group were told to "ignore the font that the word was presented in and to only pay attention to the meaning of the word." The subjects made their judgments on a 4-point Likert scale in which 1 indicated *very legible/pleasant* and 4 indicated *very illegible/unpleasant*. Words were displayed for 2 sec before the Likert scale was displayed that prompted the subjects to make a rating. Responses were entered on the computer keyboard, with the keys labeled for easy identification.

In the subsequent test phase, individual words were displayed on the computer screen, and the subjects were instructed either to press a key labeled "N" if they believed the displayed word (irrespective of the font) had not been presented during the encoding phase or to press "O" if they remembered seeing the displayed word during the encoding phase. This procedure continued until all 144 words (72 studied and 72 new words) had been judged.

Results and Discussion

Table 1 presents the proportions of hits and false alarms and mean d' as a function of encoding task (pleasantness or legibility). The proportions are also partitioned by font match condition: in the case of old words, whether the font used at test matched the font used during encoding (original), had been used with another word (swapped), or was presented in a font not used during encoding (novel). In the case of new words, only the latter two categories applied. Given that the false alarm rate for the *original* font condition is undefined, d' for the original font

condition was calculated using the false alarm rate of the *swapped* condition. We conducted analyses of variance (ANOVAs) on hits and d' scores, using encoding task and font-matching condition as factors. (We also analyzed false alarms as a function of orienting task.) The analyses were conducted both with and without the *novel* condition. The comparison of our results to those of Graf and Ryan (1990) is more straightforward when the *novel* condition is omitted, but both analyses are listed for completeness (the *F* statistic that includes the *novel* condition is listed second).

As one would expect from manipulations of levels of processing (Craik & Lockhart, 1972), subjects who rated pleasantness correctly recognized more old words than did the subjects judging legibility [$F(1,28) = 11.34$, $MS_e = 0.021$, $p < .01$; $F(1,28) = 13.58$, $MS_e = 0.039$, $p < .01$]. There were also fewer false alarms in the pleasantness condition than in the legibility condition, but this difference was not reliable. The signal detection analysis supports the same conclusion: The subjects were reliably more accurate in the pleasantness condition [$F(1,28) = 9.91$, $MS_e = 0.866$, $p < .01$; $F(1,28) = 12.57$, $MS_e = 1.467$, $p < .01$].

Comparison of d' for just the swapped and the novel fonts reveals an interaction of encoding task and font type [$F(1,28) = 4.50$, $MS_e = 0.141$, $p < .05$], so that the novel font yielded a higher d' for the words in the pleasantness judgment task and the swapped font yielded a higher d' for the subjects in the legibility judgment task; however, neither specific contrast (within encoding task) was reliable, and the pattern was not replicated in the experiments to be described later.

Of greater interest is the effect of font matching. As was predicted, there was a main effect on hits [$F(1,28) = 18.28$, $MS_e = 0.010$, $p < .001$; $F(2,56) = 27.82$, $MS_e = 0.010$, $p < .001$], reflecting the better performance when the font of the word at test matched the font used to present the word during encoding. The signal detection analysis reinforced these conclusions, showing a reliable effect of matching font on d' [$F(1,28) = 14.17$, $MS_e = 0.123$, $p < .001$; $F(2,56) = 9.07$, $MS_e = 0.139$, $p < .001$].

Figure 3 compares the results of Graf and Ryan (1990; in the right panel) and of our analogous conditions (in the left panel). Although Graf and Ryan did not report an interaction between encoding task and font match condition (original vs. swapped), their conclusions were based on a significant difference owing to font matching in the legibility condition and the absence of a difference in the pleasantness condition. In our experiment, the same conditions showed significant effects on d' of both encoding task and font matching condition (reported above), and the interaction was not reliable ($F < 1$).

In sum, like Graf and Ryan (1990), we found a main effect of orienting task; however, we also found a main effect of font-matching condition, but no interaction of orienting task with the font-matching effect. It is worth noting that a proponent of the TAP framework might argue that our failure to find an interaction was due to lack of power, despite our two main effects. Thus, the question naturally arises of whether our main effect of font matching was driven totally by the legibility orienting task condition, or whether, unlike Graf and Ryan, we found a reliable effect of font type in the pleasantness orienting task condition. Even in the pleasantness condition, the original font condition produced better performance than did the swapped font condition [$F(1,14) = 4.58$, $MS_e = 0.152$, $p = .05$].

Given that we obtained a sizeable main effect of orienting task, the difference between our results (i.e., finding an effect of matching font for both orienting tasks) and Graf and Ryan's (1990; i.e., finding it only for the perceptual orienting task) should not be attributed to a weakness in our orienting task manipulation. So why was there a difference between our results and theirs? Two explanations occur to us. As was explained earlier, we believe that Graf and Ryan's font manipulation was weaker than ours because they used only two different fonts, whereas we used a different, unusual font for each study word. Another explanation, although

less interesting, is that their font-matching effect was masked by a ceiling effect in the pleasantness orienting task condition. Not surprisingly, according to levels of processing (Craik & Lockhart, 1972), subjects' performance was much better in the pleasantness judgment task than in the legibility judgment task condition in both our study and Graf and Ryan's. Given that Graf and Ryan's subjects outperformed ours in the pleasantness condition, perhaps our subjects were not vulnerable to the ceiling effect problem, whereas theirs were. Experiment 2 was designed to test both the font fan and the ceiling effect explanations.

EXPERIMENT 2

The same two orienting tasks were used as those in Experiment 1, but in several other respects Experiment 2 differed from Experiment 1. Words were again presented in unusual fonts; however, for half of the words that were judged during the orienting task, the font used to display that word was a font that also was used to present 11 other words. Words displayed in fonts shared by 12 words during encoding were described as being in the *high-fan font* condition. We consider the high-fan font condition to be roughly analogous to the Graf and Ryan (1990) paradigm. For the remainder of the words, each presentation font used during encoding was unique to that word. These words were thus assigned to the *low-fan font* condition that replicates the design used in Experiment 1.

The other difference from Experiment 1 was the addition of a second task, whose purpose, with respect to this study, was to vary the delay between encoding and test. At issue was whether the delay between study and test would influence recognition accuracy and whether that variable would differentially affect the impact of font fan on font matching, depending on orienting task.

Method

Subjects—The subjects were 28 Carnegie Mellon undergraduates, who were paid $8 for their participation. The pleasantness group included a total of 13 subjects (6 in the immediate and 7 in the delayed recognition conditions). The legibility group included a total of 15 subjects (8 in the immediate and 7 in the delayed recognition conditions) .

Design and Materials—The design consisted of five factors, the first three being the same as those in Experiment 1: orienting task (judging legibility vs. pleasantness), word status (judged vs. not seen during encoding), and font-matching condition (original vs. swapped vs. novel). Two new factors were introduced: font fan (high-fan font vs. low-fan font) and delay (immediate vs. delayed test). Orienting task and delay were the only between-subjects factors. The level of font fan for a given swapped font matched the level of font fan used with that word during encoding.

The words were not identical to those used in Experiment 1 but were drawn from the same database, using the same criteria. The fonts were from the same pool as that generated for Experiment 1. Construction of lists was quite similar to Experiment 1, except for the incorporation of the font fan factor.

Procedure—The subjects were tested individually in sessions that lasted approximately 80 min. The session consisted of three phases: an encoding task, a filler task, and the recognition task. The latter two tasks were counterbalanced across subjects for each encoding group. The filler task was word identification of briefly presented words. The stimuli included words judged during the encoding phase, as well as words that had not been rated. Words and fonts were randomly assigned to appear in either the identification or the recognition task—that is, they did not appear in both tasks. The encoding task instructions and the procedure during the recognition phase were the same as those in Experiment 1. Assignment to judgment task and

delay condition was randomly determined, with the constraint that an equal number of subjects was assigned to each condition.

Results and Discussion

Table 2 presents the mean proportions of hits, false alarms, and d' scores as a function of encoding task, font-matching condition, and font fan and by whether recognition immediately followed encoding or was delayed by an intervening task. As before, d' was calculated so that the false alarms for the original font condition were based on swapped fonts, because original fonts do not exist for new words. For novel fonts, high versus low fan is not defined, so there is only one entry for that factor.

As with Experiment 1, we conducted analyses over all three levels of matching (including the novel font condition), but for many analyses it does not make sense to include that level, inasmuch as fan is not defined for novel font. (F statistics from analyses that are possible with the novel font condition are listed second.) The four-way ANOVAs on d' used encoding task, font-matching condition, fan of font, and delay of test as factors.

Not surprisingly, there was a main effect of encoding task on d' [$F(1,24) = 68.93$, $MS_e = 1.567$, $p < .0001$; $F(1,24) = 78.26$, $MS_e = 1.093$, $p < .0001$] so that judging pleasantness produced better recognition performance than did judging legibility. Also, as was expected, there was a main effect of delay [$F(1,24) = 6.27$, $MS_e = 1.567$, $p < .05$; $F(1,24) = 7.26$, $MS_e = 1.093$, $p < .05$], so that performance declined with delay. The interaction of these two factors was not reliable ($Fs < 1$). We again demonstrated a perceptual match effect in an explicit memory task: The subjects performed best if the font at test matched the font of the word during encoding [$F(1,24) = 10.87$, $MS_e = 0.199$, $p < .01$; $F(2,48) = 2.91$, $MS_e = 0.203$, $p = .06$]. Also replicating Experiment 1, the effect of matching font did not interact with encoding task ($Fs < 1$).

The font fan manipulation did not produce a main effect, nor did it interact with encoding task (all $Fs < 2$). Importantly, we did find the predicted interaction between the font fan and the font-matching conditions [$F(1,24) = 4.34$, $MS_e = 0.235$, $p < .05$]: A low-fan font facilitated performance more than did a high-fan font when the font was the original encoding font. This prediction derives from the assumption that there exists a link from the original font node to the episodic node that encodes the study event. Activation will spread from a font node to the episodic node only if the font at test is the same as the one used during the encoding phase. How much will spread from the font node to any episodic node depends on the number (fan) of competing associations—that is, how many other words have been seen with that font.

Although there is no link from a swapped font to the episode node, it is worth noting that there was a suggestion of a greater tendency to say *old* for both new and old words in the swapped font condition when the font was high fan. We believe that the greater familiarity of the high-fan fonts causes this spurious feeling of familiarity. Indeed, recently we have replicated that finding (M. J. Peterson & Reder, 2002).

Just as we had failed to find an interaction of the font-matching condition and orienting task, the three-way interaction of font fan, font-matching condition, and orienting task (on d') also was not reliable [$F(1,24) = 2.67$, $MS_e = 0.235$, $p > .05$]. On the other hand, the three-way interaction of font fan, font-matching condition, and delay was reliable [$F(1,24) = 4.49$, $MS_e = 0.235$, $p < .05$]. Figure 4 shows an interaction of font fan and font matching, but only in the delay condition. Earlier, we speculated that Graf and Ryan (1990) might have failed to find an effect of font matching in the pleasantness judgment condition owing to a ceiling on performance with that orienting task. In this experiment, our subjects performed near ceiling in the immediate pleasantness condition and near floor in the delayed legibility condition. Our failure to find sensible results in those two conditions—namely, no advantage of original over

swapped font for low fan in the immediate condition for the pleasantness task or for high fan in the legibility task at a delay—may have been the result of ceiling and floor effects, respectively.

One might question the robustness of the (significant) interaction of font fan with font matching, given that there was a higher three-way interaction with delay. One of the objectives of the next experiment was to keep performance off both the ceiling and the floor and to clearly establish the result that the font-matching effect is modulated by font fan, so that it is larger in the low font fan condition but exists in the high-fan condition, too.

Stronger test of our theoretical position—According to SAC, the reason why low-fan fonts are at an advantage in the matching font condition is that the link from the font node to the episode node will send more activation when there are not many competing contextual links from the node to siphon off activation (refer to Figure 1). We also postulate that recollections derive from sufficient activation's reaching the node that represents the trace of having studied the word in the experimental context. If there is insufficient activation at the episode node, subjects will guess whether the word was studied on the basis of a familiarity assessment derived from the activation level of the concept node or, in some cases, on the basis of the familiarity of the font node. If our account is correct, not only should there be more old responses when there is a link from the font node to the episode node (i.e., in the match condition) and when that link has fewer competitors (the low font fan condition), but this advantage should be manifested exclusively in the *recollective* responses. Experiment 3 employs the *remember/know* paradigm to test these predictions.

EXPERIMENT 3

Remember/Know Paradigm

The remember/know paradigm (Gardiner, 1988; Tulving, 1985) requires subjects to assess the type of memorial information that led to each *old* judgment. In this paradigm, after a subject responds *old*, the next judgment is to assess whether or not the *old* judgment was based on a recollection of experiencing the item on the study list. If so, the subject is to respond *remember*. When subjects do not specifically recollect the experience of studying the item, but instead feel that the item is sufficiently familiar that it must have been on the study list, the appropriate response is *know*.

Although it might seem surprising that subjects can actually make remember versus know discriminations, the procedure has been used many times with considerable success in terms of separating these two types of processes (Gardiner, 1988; Gardiner & Java, 1990; Gardiner & Parkin, 1990; Rajaram, 1993; Tulving, 1985). It is worth noting that there exist signal detection accounts of remember/know judgments (e.g., Donaldson, 1996; Hirshman & Henzler, 1998; Hirshman & Master, 1997; Inoue & Bellezza, 1998), suggesting that the different responses merely reflect a shift in criterion. However, there is ample evidence that such an account is insufficient to explain the data (see Reder et al., 2000, for a more detailed discussion).

Given the representational assumptions of the SAC model, it has been useful to employ this procedure also because *remember* maps naturally onto recollections. Subjects should give more *remember* responses to studied words tested in the font used at encoding and more if that font had been studied with only one word, rather than with a dozen.

Method

Subjects—The subjects were 64 Carnegie Mellon Undergraduates enrolled in psychology courses, who participated as a partial fulfillment of their course requirements.

Procedure, Design, and Materials—The materials, procedure, and design of Experiment 3 differed little from those in Experiment 2. Given that the orienting task factor did not interact with the factors of interest in either Experiment 1 or 2 and seemed only to produce large differences in performance level, this factor was dropped. Instead, all the subjects were asked to judge the *appropriateness of the font used to present the word*, thereby encouraging them to attend to both the meaning and the perceptual aspects of the words presented. The appropriateness of the font to the word's meaning was rated on a Likert scale with four values.

The other factor that was no longer varied was delay between encoding and test. Instead, all the subjects performed a distractor task (arbitrarily, one of two psychometric tests, one measuring nonverbal fluency and the other focal visual attention; neither used words or the type of fonts used in the study list) between the encoding task and the recognition test. The filler test was administered for exactly 10 min and was included to minimize ceiling effects. We chose a different type of filler task than the one used in Experiment 2, to ensure that our results were not dependent on a filler task that involved similar types of materials.

The other difference between this experiment and the previous ones was that we asked the subjects to judge whether their *old* responses were *remember* responses or *know* responses. After completing the filler task, the subjects were instructed to respond in one of three ways. When the subjects believed that the displayed word had *not* been presented earlier, they were to press a key labeled "N"; when the subjects had a *conscious recollection* of having seen the displayed word in the study phase, they were instructed to press "R"; when the subjects *knew* that the displayed word had been presented earlier but did not have a conscious recollection of having seen it, they were instructed to press "K." Before the subjects were allowed to begin the recognition phase of the experiment, they were carefully instructed in how to distinguish *remember* from *know* memories, were given everyday examples, and were asked to provide examples of their own that illustrated that they understood the distinction between responding *remember* and responding *know*.

The same fonts were used as those in Experiment 2. Likewise, the pool of words used for this experiment was sampled from the same database as that used in Experiment 2. Stimuli were randomly assigned to condition for each subject.

Results and Discussion

Table 3 presents the mean proportions of hits and false alarms and mean d' scores as a function of font-matching condition and font fan. The proportion of *remember* responses is listed parenthetically beside the proportion of hits. The *remember* values are absolute proportions, and thus, the number will always be smaller than the corresponding hit proportion. (The difference between the two values represents the proportion of *know* responses for that condition).

As in Experiments 1 and 2, we conducted ANOVAs on d' performance with and without the novel font condition, placing the statistics that include the novel font condition second. Again, there was a main effect of font-matching condition [$F(1,63) = 38.62$, $MS_e = 0.13$, $p < .001$; $F(2,124) = 20.56$, $MS_e = 0.10$, $p < .001$], so that recognition accuracy was best when the fonts at encoding and test matched. There was also a significant effect of font fan [$F(1,63) = 9.09$, $MS_e = 0.35$, $p < .005$], so that recognition accuracy was better when the test font was low fan. As was anticipated, the interaction between font-matching condition and font fan was statistically reliable [$F(1,63) = 7.27$, $MS_e = 0.11$, $p < .01$]. The left panel of Figure 5 illustrates this interaction between font fan and whether the font matched between encoding and test (novel fonts are undefined for fan font and, hence, are excluded from the figure).

Remember judgments—The unique predictions for Experiment 3 involve the proportion of *remember* responses in the various conditions. As was predicted, there was a main effect of font-matching condition [$F(1,63) = 87.18$, $MS_e = 0.02$, $p < .0001$; $F(2,124) = 65.29$, $MS_e = 0.01$, $p < .0001$], so that there were more *remember* hits when the fonts matched from encoding to test. Also, there was a main effect of font fan [$F(1,63) = 20.41$, $MS_e = 0.017$, $p < .0001$], so that there were more *remember* hits when the font was unique than when the font had been studied with a dozen words. This greater tendency to give a *remember* response for low- than for high-fan fonts was twice as large in the original font condition than in the swapped font condition (12% vs. 6%, respectively), and this interaction between fan and font-matching condition was reliable [$F(1,63) = 4.05$, $MS_e = 0.01$, $p < .05$]. This interaction is illustrated in the right panel of Figure 5.

GENERAL DISCUSSION

In this article our goal has been to explore the mechanisms that drive the perceptual match effect in recognition memory. In particular, we have sought to test an account for these effects that does not posit different mechanisms than one would posit to explain semantic (i.e., conceptual) memory effects. We have modeled the effects described in this paper with a simulation that uses the same parameter values (e.g., for spread of activation, decay, strengthening) for the perceptual information as for the semantic information (Cary & Reder, 2000). These parameter values were taken from earlier modeling efforts in other domains (e.g., Reder et al., 2000).

Although a *levels of processing* manipulation affected memory performance, as has traditionally been found, this manipulation did not interact with the matching font effect. The variable that modulated the matching font effect was the *fan of the font*. A matching font that was unique produced a greater memory benefit and more recollection responses than did a matching font that had been associated with a dozen words. Our results do not rule out the possibility that a portion of the effect is modulated by a match in *processing*, as is proposed by the TAP framework; however, the perceptual matching effects seem more dependent on a memory representation and retrieval mechanism—specifically, the font fan effect as specified in SAC: the degree of interference from the variable number of associations competing for activation from the font. Although it might be possible to develop an explanation for these effects within the TAP framework, it is important to emphasize that these effects were *predicted* by SAC, not *postdicted*.

The perceptual matching effects in our experiment can be explained without positing separate memory systems, such as the ones offered in the PRS of Schacter and Tulving (1994). According to that view, perceptual memory effects are due to an independent processing system that presumably operates in a qualitatively different way than does conceptual memory processes. However, the nature of these results seems quite consistent with an account that treats perceptual information exactly the same as conceptual information. Not only have we demonstrated that perceptual matching effects can be obtained for explicit memories, we have also demonstrated these same font-matching effects for implicit memory tasks (Cary & Reder, 2000). Here, too, we used the same parameter values, as well as assuming the same memory representation for both implicit and explicit memory, for both perceptual and semantic information.

Conclusion

In summary, we have offered a partial explanation for when and why one obtains perceptual similarity effects in memory. We suspect that the modality effects are a generalization of these findings and that those effects in audition (e.g., Clarke & Morton, 1983; Graf, Shimamura, & Squire, 1985; Jacoby & Witherspoon, 1982; Kirsner, Milech, & Standen, 1983; Light et al.,

1992; Roediger & Blaxton, 1987a, 1987b; Schacter & Graf, 1989) would have been stronger if a different voice had been used for each stimulus item. Our interpretation also implies that the memory advantage associated with pictures over words might be attributable in part to the greater distinctiveness of the pictures. After all, any given picture, like an unusual font, has been seen in relatively few contexts and will have less contextual fan to interfere with episodic recognition. In fact, in recent work in our laboratory (Arndt & Reder, 2001), we have found that distinctiveness ratings for fonts are affected by the randomly determined contextual variability—that is, the number of words with which each font was paired during the study phase.

It is not necessary to assume a multiple systems view (i.e., that perceptual and conceptual information are stored in and processed by qualitatively different systems) to explain our results. Likewise, these data do not require the assumptions of the multiple processes view, that perceptual memory effects are somehow separate from conceptual memory effects and that these effects are only invoked when an encoding task orients processing in one way or the other. This is not to say that the way people process information does not have an impact on their subsequent processing of the same material (Franks, Bilbrey, Lien, & McNamara, 2000). Our data suggest that performance depends on more than recreating the same processing at test that occurred during encoding. Performance appears to be affected by a representation and processing pair in such a way that, if the representation has a greater number of competing associations, performance will be degraded.

References

Anderson JR. Retrieval of propositional information from long-term memory. Cognitive Psychology 1974;6:451–474.

Arndt J, Reder LM. The role of contextual competition on judgments of distinctiveness. 2001Manuscript in preparation

Blaxton TA. Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. Journal of Experimental Psychology: Learning, Memory, & Cognition 1989;15:657–668.

Cantor J, Engle RW. Working-memory capacity as long-term memory activation: An individual-differences approach. Journal of Experimental Psychology: Learning, Memory, & Cognition 1993;19:1101–1114.

Carr TH, Brown JS, Charalambous A. Repetition and reading: Perceptual encoding mechanisms are very abstract but not very interactive. Journal of Experimental Psychology: Learning, Memory, & Cognition 1989;15:763–778.

Cary, M.; Reder, LM. Modeling the role of perceptual features in word recognition and fragment completion; 2000, November; Paper presented at the 41st Annual Meeting of the Psychonomic Society; New Orleans.

Cave CB, Bost PR, Cobb RE. Effects of color and pattern on implicit and explicit picture memory. Journal of Experimental Psychology: Learning, Memory, & Cognition 1996;22:639–653.

Clarke R, Morton J. Cross modality facilitation in tachistoscopic word recognition. Quarterly Journal of Experimental Psychology 1983;35A:79–96.

Craik FIM, Lockhart RS. Levels of processing: A framework for memory research. Journal of Verbal Learning & Verbal Behavior 1972;11:671–684.

Craik FIM, Moscovitch M, McDowd JM. Contributions of surface and conceptual information to performance on implicit and explicit memory tasks. Journal of Experimental Psychology: Learning, Memory, & Cognition 1994;20:864–875.

Curran T, Schacter DL, Bessenoff G. Visual specificity effects on memory: Beyond transfer appropriate processing? Canadian Journal of Experimental Psychology 1996;50:22–33. [PubMed: 8653095]

Donaldson W. The role of decision processes in remembering and knowing. Memory & Cognition 1996;24:523–533.

Durgunođlu AY, Roediger HL III. Test differences in accessing bilingual memory. Journal of Memory & Language 1987;26:377–391.

Feustel TC, Shiffrin RM, Salasoo A. Episodic and lexical contributions to the repetition effect in word identification. Journal of Experimental Psychology: General 1983;112:309–346. [PubMed: 6225826]

Franks JJ, Bilbrey CW, Lien KG, McNamara TP. Transfer-appropriate processing (TAP) and repetition priming. Memory & Cognition 2000;28:1140–1151.

Gardiner JM. Functional aspects of recollective experience. Memory & Cognition 1988;16:309–313.

Gardiner JM, Java RI. Recollective experience in word and nonword recognition. Memory & Cognition 1990;18:23–30.

Gardiner JM, Parkin AJ. Attention and recollective experience in recognition memory. Memory & Cognition 1990;18:579–583.

Graf P, Levy BA. Reading and remembering: Conceptual and perceptual processing involved in reading rotated passages. Journal of Verbal Learning & Verbal Behavior 1984;23:405–424.

Graf P, Ryan L. Transfer-appropriate processing for implicit and explicit memory. Journal of Experimental Psychology: Learning, Memory, & Cognition 1990;16:978–992.

Graf P, Shimamura AP, Squire LR. Priming across modalities and priming across category levels: Extending the domain of preserved function in amnesia. Journal of Experimental Psychology: Learning, Memory, & Cognition 1985;11:386–396.

Hirshman E, Henzler A. The role of decision processes in conscious recollection. Psychological Science 1998;9:61–65.

Hirshman E, Master S. Modeling the conscious correlates of recognition memory: Reflections on the remember–know paradigm. Memory & Cognition 1997;25:345–351.

Hofstadter, D. Metafont, metamathematics, and metaphysics. In: Hofstadter, D., editor. Metamagical themas: Questing for the essence of mind and pattern. New York: Basic Books; 1985. p. 260-296.

Hofstadter, D.; McGraw, G. the Fluid Analogies Research Group. Letter spirit: Esthetic perception and creative play in the rich microcosm of the Roman alphabet. In: Hofstadter, D., editor. Fluid concepts and creative analogies. New York: Basic Books; 1995. p. 405-466.

Inoue C, Bellezza FS. The detection model of recognition using know and remember judgments. Memory & Cognition 1998;26:299–308.

Jacoby LL. Remembering the data: Analyzing interactive processes in reading. Journal of Verbal Learning & Verbal Behavior 1983;22:485–508.

Jacoby LL, Dallas M. On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General 1981;110:306–340. [PubMed: 6457080]

Jacoby LL, Hayman CA. Specific visual transfer in word identification. Journal of Experimental Psychology: Learning, Memory, & Cognition 1987;13:456–463.

Jacoby LL, Witherspoon D. Remembering without awareness. Canadian Journal of Psychology 1982;36:300–324.

Kimberg DY. Executive functions, working memory, and frontal lobe function. 1994Unpublished doctoral dissertation, Carnegie Mellon University

Kirsner K, Milech D, Standen P. Common and modality-specific processes in the mental lexicon. Memory & Cognition 1983;11:621–630.

Kolers PA. Specificity of operations in sentence recognition. Cognitive Psychology 1975;7:289–306.

Kolers PA. Reading and knowing. Canadian Journal of Psychology 1979;33:106–117.

Kolers PA, Ostry DJ. Time course of loss of information regarding pattern analyzing operations. Journal of Verbal Learning & Verbal Behavior 1974;13:599–612.

Kuèera, H.; Francis, WN. Computational analysis of present-day American English. Providence, RI: Brown University Press; 1967.

Lewis CH, Anderson JR. Interference with real world knowledge. Cognitive Psychology 1976;8:311–335.

Light LL, LaVoie D, Valencia-Laver D, Owens SA, Mead G. Direct and indirect measures of memory for modality in young and older adults. Journal of Experimental Psychology: Learning, Memory, & Cognition 1992;18:1284–1297.

Mandler G. Recognizing: The judgment of previous occurrence. Psychological Review 1980;87:252–271.

Murdock BB, Walker KD. Modality effects in free recall. Journal of Verbal Learning & Verbal Behavior 1969;8:665–676.

Peterson MJ, Reder LM. The role of spurious feature familiarity in recognition memory. 2002Manuscript in preparation

Peterson SB, Potts GR. Global and specific components of information interpretation. Journal of Verbal Learning & Verbal Behavior 1982;21:403–420.

Rajaram S. Remembering and knowing: Two means of access to the personal past. Memory & Cognition 1993;21:89–102.

Rajaram S, Roediger HL III. Direct comparison of four implicit memory tests. Journal of Experimental Psychology: Learning, Memory, & Cognition 1993;19:765–776.

Reder LM, Angstadt P, Cary M, Ayers MA, Erickson MA. A reexamination of stimulus-frequency effects in recognition: Two mirrors for low- and high-frequency pseudowords. Journal of Experimental Psychology: Learning, Memory, & Cognition 2002;28:138–152.

Reder LM, Nhouyvanisvong A, Schunn CD, Ayers MS, Angstadt P, Hiraki K. 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, & Cognition 2000;26:294–320.

Reder LM, Ross BH. Integrated knowledge of different tasks: The role of retrieval strategy on fan effects. Journal of Experimental Psychology: Learning, Memory, & Cognition 1983;9:55–72.

Reder, LM.; Schunn, CD. Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In: Reder, LM., editor. Implicit memory and metacognition. Mahwah, NJ: Erlbaum; 1996. p. 45-77.

Reder LM, Wible C. Strategy use in question-answering: Memory strength and task constraints on fan effects. Memory & Cognition 1984;12:411–419.

Richardson-Klavehn A, Bjork RA. Measures of memory. Annual Review of Psychology 1998;36:475–543.

Roediger HL III, Blaxton TA. Effects of varying modality, surface features, and retention interval on priming in word-fragment completion. Memory & Cognition 1987a;15:379–388.

Roediger, HL., III; Blaxton, TA. Retrieval modes produce dissociations in memory for surface information. In: Gorfein, DS.; Hoffman, RR., editors. The Ebbinghaus Centennial Conference. Hillsdale, NJ: Erlbaum; 1987b. p. 349-379.

Roediger, HL., III; McDermott, KB. Implicit memory in normal human subjects. In: Spinnler, H.; Boller, F., editors. Handbook of neuropsycholog y. Amsterdam: Elsevier; 1993. p. 64-131.

Roediger, HL., III; Srinivas, K.; Weldon, MS. Dissociations between implicit measures of retention. In: Lewandowsky, SS.; Dunn, JC.; Kirsner, K., editors. Implicit memory: Theoretical issues. Hillsdale, NJ: Erlbaum; 1989. p. 67-84.

Schacter DL, Graf P. Modality specificity of implicit memory for new associations. Journal of Experimental Psychology: Learning, Memory, & Cognition 1989;15:3–12.

Schacter, DL.; Tulving, E. What are the memory systems of 1994?. In: Schacter, DL.; Tulving, E., editors. Memory systems. Cambridge, MA: MIT Press; 1994. p. 1-38.

Schunn CD, Reder LM, Nhouyvanisvong A, Richards DR, Stroffolino PJ. To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. Journal of Experimental Psychology: Learning, Memory, & Cognition 1997;23:3–29.

Tulving E. Memory and consciousness. Canadian Psychologist 1985;26:1–12.

Tulving E, Schacter DL. Priming and human memory systems. Science 1990;247:301–306. [PubMed: 2296719]

Weldon MS, Roediger HL III. Altering retrieval demands reverses the picture superiority effect. Memory & Cognition 1987;15:269–280.

Zbrodoff NJ. Why is 9 + 7 harder than 2 + 3? Strength and interference as explanations of the problem-size effect. Memory & Cognition 1995;23:689–700.
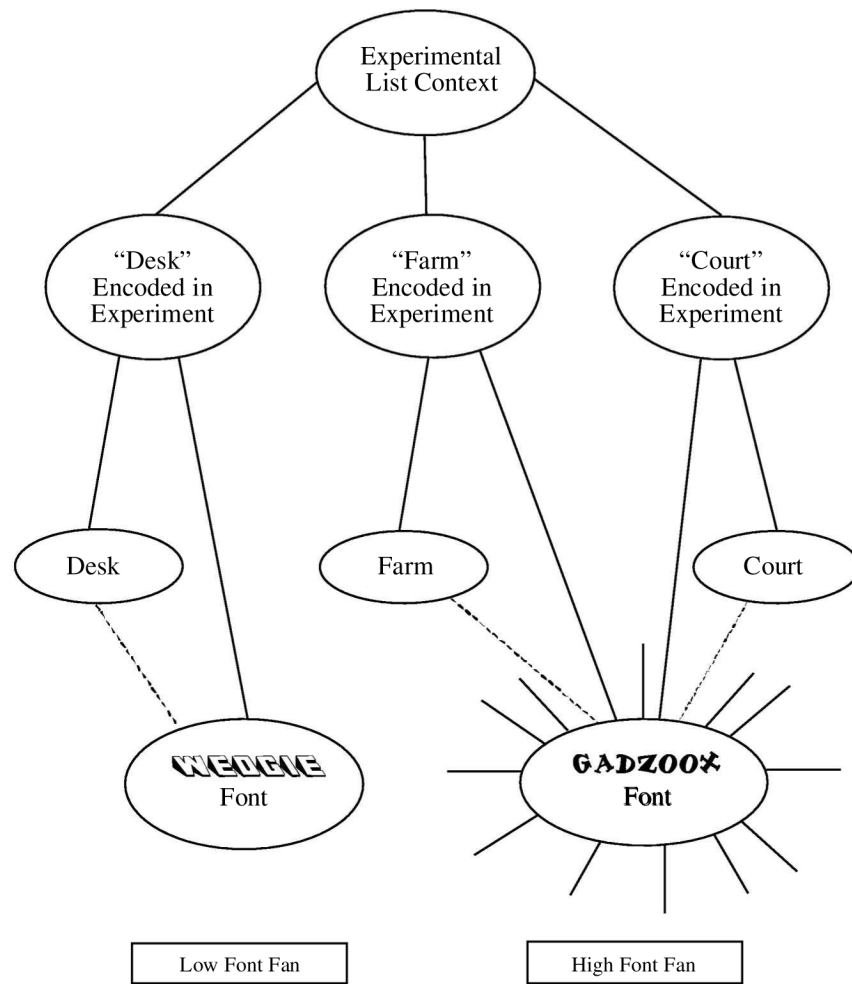
Figure 1.
A schematic illustration of the proposed memory representation of several words in an experiment that varies font fan. The word *desk* is illustrated as being encoded in a unique font (on the left), and two other words as encoded in a high-fan font.

Figure 2.
Examples of fonts used to present words at encoding and at test.
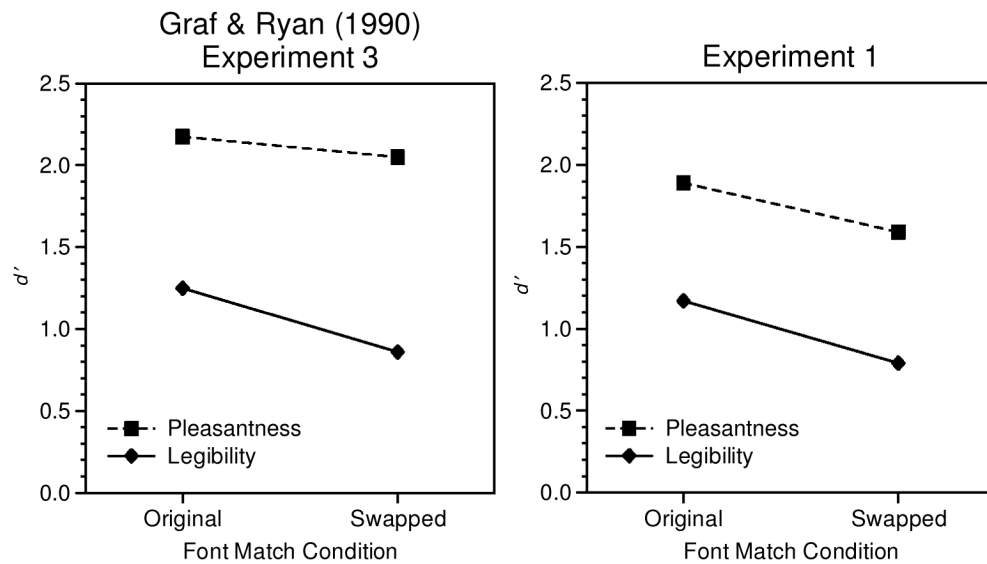
Graf & Ryan (1990)

Figure 3.
Comparison of Graf and Ryan's (1990) data (from their Experiment 3, Table 9 on p. 986) with corresponding data from our Experiment 1. Figures plot mean *d'* scores as a function of encoding task and font-matching condition.
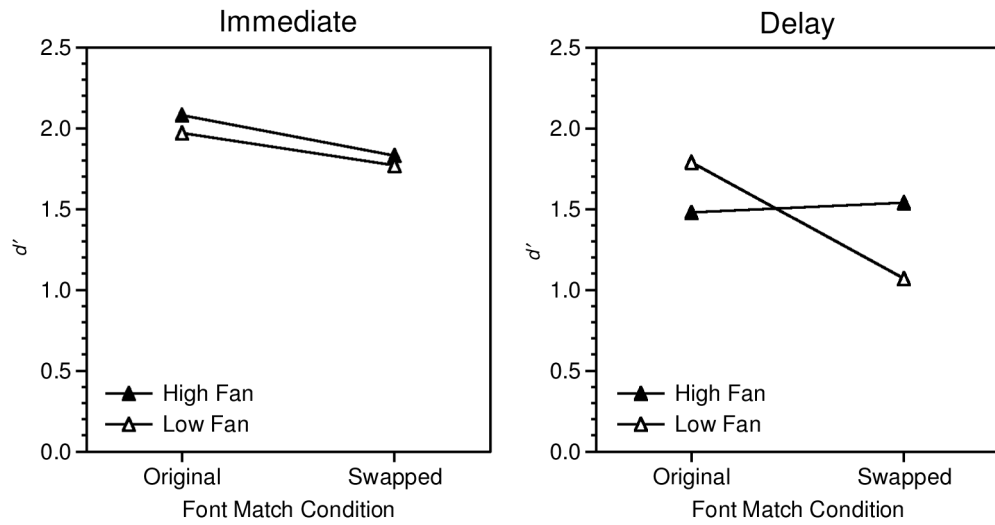
Figure 4.
Mean *d′* scores as a function of font-matching condition and fan font, for the immediate and delayed tests collapsed over encoding task for Experiment 2.
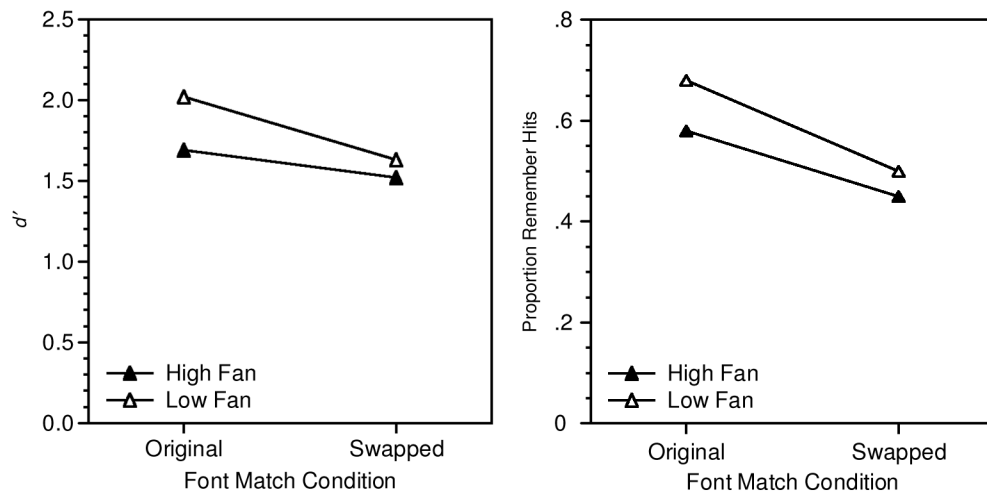
Figure 5.
Mean *d'* scores (left panel) and proportions of *remember* hits (right panel) as a function of font-matching condition and fan font for Experiment 3.

Table 1

Proportion of Hits and False Alarms and Mean d' Scores, as a Function of Font-Matching Condition and Encoding Task in Experiment 1

Group/DV	Font-Matching Condition		
	Original	Swapped	Novel
Pleasantness			
Hit	.87	.78	.74
FA	NA	.29	.22
d'	1.89	1.59	1.76
Legibility			
Hit	.77	.64	.53
FA	NA	.35	.33
d'	1.17	0.79	0.56

Note—Font-matching indicates what condition the test word was seen in. NA means *not applicable*, since the original font is undefined for nonstudied items. Consult the text for definition of conditions. DV, dependent variable.

Table 2

Proportions of Hits and False Alarms (FAs) and Mean *d′* Scores as a Function of Font Fan, Font Match, Encoding Task, and Test Delay in Experiment 2

	Font-Matching Condition		
Encoding/Fan Condition	**Original**	**Swapped**	**Novel**
Immediate Test Condition			
Pleasantness			
High Fan			
Hits	.94	.87	
FA	NA	.06	
d′	3.43	3.00	
Low Fan			
Hits	.92	.91	
FA	NA	.10	
d′	3.04	3.01	
Novel			
Hits			.86
FA			.05
d′			3.23
Legibility			
High Fan			
Hits	.63	.63	
FA	NA	.33	
d′	0.88	0.85	
Low Fan			
Hits	.72	.58	
FA	NA	.26	
d′	1.37	0.95	
Novel			
Hits			.53
FA			.21
d′			0.94
Delayed Test Condition			
Pleasantness			
High Fan			
Hits	.82	.81	
FA	NA	.13	
d′	2.42	2.44	
Low Fan			
Hits	.83	.71	
FA	NA	.11	
d′	2.62	2.07	
Novel			
Hits			.73
FA			.10
d′			2.33
Legibility			
High Fan			
Hits	.58	.60	
FA	NA	.40	
d′	0.54	0.63	
Low Fan			
Hits	.59	.27	
FA	NA	.29	
d′	0.96	0.06	
Novel			
Hits			.38
FA			.25
d′			0.44

Table 3

Proportion of Hits and False Alarms (FAs) and Mean d' Scores as a Function of Font-Matching and Font Fan Conditions in Experiment 3

Encoding/Fan Condition	Font-Matching Condition		
	Original	Swapped	Novel
High fan			
Hits	.85 (.58)	.80 (.45)	
FA	NA	.33	
d'	1.69	1.52	
Low fan			
Hits	.90 (.68)	.81 (.50)	
FA	NA	.30	
d'	2.02	1.63	
Novel			
Hits			.77 (.44)
FA			.29
d'			1.49

Note—The proportions of *remember* judgments for hits are given in parentheses.