Increasing Word Distinctiveness Eliminates the Picture Superiority Effect in Recognition: Evidence for the Physical Distinctiveness Account

Tyler M. Ensor

Aimée M. Surprenant

Ian Neath

Memorial University of Newfoundland

In Press: Memory & Cognition

© 2018, Psychonomic Society. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors’ permission. The final article will be available, upon publication, via its DOI: 10.3758/s13421-018-0858-9

Correspondence:

Ian Neath
Department of Psychology
Memorial University of Newfoundland
St. John’s, NL A1B 3X9
Canada
email: ineath@mun.ca
voice: 709 864 8495
fax: 709 864 2430
WORD DISTINGUISHEDNESS

Abstract

All else being equal, memory is usually better for pictures than words (Paivio & Csapo, 1969). Researchers have attributed pictures’ mnemonic advantage to dual coding (Paivio, 1971), which holds that pictures are more likely than words to be represented with two codes; to conceptual distinctiveness (Hamilton & Geraci, 2006), which holds that processing pictures leads to more conceptual information than processing words; and to physical distinctiveness (Mintzer & Snodgrass, 1999), which holds that pictures differ on more physical dimensions than words.

Here, we present a novel test of the physical distinctiveness account of picture superiority: If the greater physical variability of pictures relative to words is responsible for their mnemonic benefit, then increasing the physical variability of words should reduce or eliminate the picture superiority effect. In Experiments 1-3, we assessed recognition performance for words and pictures when word distinctiveness was increased by varying font style, font size, color, and capitalization. Additionally, in Experiment 3, we further reduced the distinctiveness of pictures by using only those pictures with similar orientations. In Experiment 4, we assessed memory for the original form by asking subjects to identify the form that each probe took during the study phase. Results were consistent with the distinctiveness prediction and, notably, inconsistent with dual coding.

Keywords: picture superiority effect, pictures and words, distinctiveness, dual coding, recognition memory
WORD DISTINCTIVENESS

Increasing Word Distinctiveness Eliminates the Picture Superiority Effect in Recognition: Evidence for the Physical Distinctiveness Account

Memory for pictures is generally better than memory for words, a finding known as the picture superiority effect. For example, Shepard (1967) tested memory for 540 words and 612 color pictures. At test, he showed two words (or pictures) simultaneously, one of which had been in the study phase and one of which was new. Although subjects correctly identified the old words 88.4% of the time, they correctly identified the old pictures 96.7% of the time. Snodgrass, Volvovitz, and Walfish (1972) reported one of the first studies to use a signal detection analysis on recognition memory for words and line drawings. They replicated Shepard’s result, reporting a $d'$ for words of 1.34 compared to 2.49 for pictures. The picture superiority effect has been found with many tests including free recall (Bevan & Steger, 1971; Bousfield, Esterson, & Whitmarsh, 1957; Paivio & Csapo, 1969, 1973), cued recall (Weldon & Coyote, 1996; Weldon, Roediger, & Challis, 1989), serial recall and reconstruction (Nelson, Reed, & McEvoy, 1977), and paired-associate learning (Nelson & Reed, 1976; Paivio & Yarmey, 1966; Wicker, 1970). In addition to the item recognition procedure of Shepard and Snodgrass et al., the picture superiority effect is also seen with tests of associative recognition (Hockley, 2008; Hockley & Bancroft, 2011).

Despite numerous demonstrations, the cause of the picture superiority effect is still debated. Extant accounts can be roughly divided into those emphasizing dual coding (Paivio, 1971, 1991, 2007) and those emphasizing distinctiveness (Hamilton & Geraci, 2006; McBride & Dosher, 2002; Mintzer & Snodgrass, 1999; Nelson, 1979; Nelson et al., 1977; Stenberg, 2006; Weldon & Coyote, 1996). In the present paper, we first review these competing accounts and
WORD DISTINCTIVENESS
then report four experiments for which dual-coding and distinctiveness accounts make contrasting predictions.

**Accounts of the picture superiority effect**

**Dual Coding**

Paivio’s (1971, 1991, 2007) dual-coding theory was the first explanation of the picture superiority effect. Dual-coding theory posits the existence of two independent pathways in memory: one for verbal representations, called the logogen pathway, and one for imaginal representations, called the imagen pathway. When a stimulus is encoded, it is first stored in the pathway corresponding to its presentation modality. For example, the word “cat” is stored in the logogen pathway, and a picture of a cat is stored in the imagen pathway. Importantly, representations in one pathway can elicit representations in the other pathway, such that the word “cat” can elicit an imaginal representation of a cat in the imagen pathway, and a picture of a cat can elicit a verbal representation in the logogen pathway. Because these two forms of representation are independent, the two codes can have additive effects; that is, memory will tend to be better for items represented in two codes than those represented in a single code.

From the perspective of dual-coding theory, the picture superiority effect occurs because pictures are more likely to generate representations in the logogen pathway than words are to generate representations in the imagen pathway. In effect, dual-coding theory posits that subjects are more likely to name a picture than they are to imagine a word’s referent. A number of results support dual-coding theory. For example, if the picture superiority effect occurs because pictures are more likely to be represented with two codes than words, inducing subjects to form an image of the word should eliminate the picture superiority effect. Paivio and Csapo (1973) confirmed this prediction. As a second example, the picture superiority effect should be observed with
Distinctiveness Accounts

Distinctiveness accounts of the picture superiority effect can be divided into those that emphasize conceptual distinctiveness and those that emphasize physical distinctiveness. According to conceptual distinctiveness accounts, processing pictures involves greater semantic elaboration than processing words, thereby producing a levels-of-processing effect (Craik & Lockhart, 1972). Support for the conceptual distinctiveness account comes from the finding that semantic-orienting tasks at study reduce or eliminate the picture superiority effect (D’Agostino, O’Neill, & Paivio, 1977; Durso & Johnson, 1980). However, this result is also predicted by the dual-coding account. As D’Agostino et al. note, semantic-orienting tasks such as “how large is the object in the real world?” can lead to generating an image. This leads to equivalent performance for deeply processed words and pictures because both have two codes (see also Paivio, 1976). Nelson et al. (1977) found that conceptual similarity among items disrupted serial
WORD DISTINCTIVENESS

recall of pictures but not words. Again, this supports both the conceptual distinctiveness view and the dual-coding view. For the former, conceptual similarity will affect pictures more than words because pictures entail more conceptual processing; for the latter, on average, only pictures will have the imagen code.

For present purposes, we are more concerned with physical distinctiveness accounts than conceptual distinctiveness accounts. According to physical distinctiveness accounts, pictures are better remembered than words because there is more physical variability from picture to picture than from word to word. The fact that pictures are more variable than words is not controversial: As Nelson (1979) pointed out, the letters, phonemes, and orthographic conventions of a language place constraints on the degree to which words can vary; however, no such constraints exist for pictures. The question, then, is whether the increased physical variability of pictures relative to words is responsible for, or merely incidental to, the picture superiority effect.

One way to test the physical distinctiveness account is to manipulate the physical similarity of pictures. Surprisingly, few studies have taken this approach. Among the exceptions, Nelson, Reed, and Walling (1976) manipulated the physical similarity of picture cues in a paired-associate learning experiment. In one condition, subjects learned picture-word pairs; in another condition, subjects learned word-word pairs. Critically, cues could be physically similar in picture form (e.g., pencil, screwdriver, nail, etc.) or physically dissimilar in picture form (e.g., sheep, banjo, peach, etc.). In the condition with dissimilar picture cues, the standard picture superiority effect was observed. The key data come from the two conditions that used similar

---

1 These accounts are not mutually exclusive; for example, Nelson’s (1979; Nelson et al., 1977) sensory-semantic model includes both physical and conceptual distinctiveness.
WORD DISTINCTIVENESS
pictures and that also manipulated presentation rate, either fast (1100 ms) or slow (2100 ms). In the fast rate condition, the picture superiority effect was reversed and in the slow-rate condition, pictures and words yielded equivalent performance. Making the pictures perceptually more similar eliminated (slow rate) or even reversed (fast rate) the picture superiority effect.

Other support for the physical distinctiveness account is less direct. A number of studies have used the change-form paradigm (Jenkins, Neale, & Deno, 1967). In this paradigm, subjects study a list of pictures and words, followed by an old/new recognition test on which targets can appear in the same form as study or can be changed to the other form. Subjects are instructed to respond “old” regardless of form; for example, if the word “cat” was seen at study but a picture of a cat was shown at test, the correct response is still “old”. These studies typically emphasize discrimination cost, which is the decrease in performance when the forms are changed than when they remain the same. Mintzer and Snodgrass (1999) observed a greater discrimination cost when pictures were changed to words than when words were changed to pictures, which they interpreted as consistent with the physical distinctiveness account. However, another plausible interpretation is that the task of comparing a word probe to a stored picture is simply more difficult than comparing a picture probe to a stored word. For example, a given picture typically has fewer possible labels compared to the number of possible pictures that could be generated from a label. This asymmetry could have made it more difficult to match a word probe to a stored picture than to match a picture probe to a stored word. An additional problem with change-form experiments with recognition tests is determining the appropriate false alarm rate (see Mintzer & Snodgrass for a discussion).

Despite support from experiments such as those of Nelson et al. (1976) and the indirect support from the change-form paradigm, other experiments have yielded mixed support for the
WORD DISTINCTIVENESS

physical distinctiveness account. For example, according to this account, color pictures should be more distinctive than black and white pictures, all else being equal. Although some studies have found better performance with color than black and white pictures (e.g., Borges, Stepnowsky, & Holt, 1977; Bousfield et al., 1957), other studies have found equivalent performance (e.g., Nelson, Metzler, & Reed, 1974; Paivio, Rogers, & Smythe, 1968; Wicker, 1970). Because these studies vary considerably in methodology and stimuli, it is not clear why the results differ. One factor may be ceiling effects such that adding color to a distinctive black and white picture yields little or no additional benefit.

The Present Study

Experimental tests of the physical distinctiveness account have generally manipulated picture distinctiveness by decreasing picture-to-picture variability (Nelson et al., 1977), by using the change-form paradigm (Mintzer & Snodgrass, 1999), or by enhancing the detail of some of the pictures (Bousfield et al., 1957). Another method of testing the physical distinctiveness account, though, is to increase the distinctiveness of the words. Although words are necessarily constrained by a language’s rules (Nelson, 1979), the physical form of the word can be manipulated to increase its distinctiveness. Therefore, we introduced more variability with the words by using multiple fonts, sizes, and font colors rather than presenting the words in a uniform font, size, and font color.

To our knowledge, the only previous picture superiority effect study to manipulate the distinctiveness of words was carried out by Paivio et al. (1968). In a free recall experiment, Paivio et al. manipulated whether subjects studied black words, color words, black and white pictures, or color pictures. They found a standard picture superiority effect, but no effect of distinctiveness. There are a number of reasons why manipulating the distinctiveness of the words
WORD DISTINCTIVENESS

may not have had an effect in this study. First, only color varied; the words were still uniform on other dimensions, such as font and font size. Second, the test was free recall, which requires the subjects to write down words. In our experiments, we addressed both of these factors. In addition to manipulating word color, we also varied font type, font size, and capitalization. This produced two types of words: those presented in standard, black font with uniform capitalization, size, and color, and those presented in distinctive, colored fonts. We also included two types of pictures, black and white as well as color. Rather than use free recall, we used recognition, in which the distinctiveness of the words should be apparent at both study and test (we expand on this point in the General Discussion).

Both the dual-coding and physical distinctiveness accounts predict that recognition performance will be best for color pictures and worst for black words. The two accounts differ in their predictions of performance on the color words and black and white pictures. From the perspective of dual-coding theory, there should still be a memory advantage for black and white pictures over color words because the manipulation does nothing to remove the additional code available to pictures. In contrast, the physical distinctiveness account predicts that performance should follow relative distinctiveness rather than probability of forming dual codes. Memory for black and white pictures and for color words should be in between that of color pictures and black words, and to the extent that the color words are sufficiently distinctive and the black and white pictures are more uniform, the picture superiority effect should be attenuated or abolished.

Experiment 1

In Experiment 1, subjects saw four kinds of stimuli: color pictures, black and white pictures, color words, and black words. They then received an old/new recognition test. Dual-coding theory predicts an advantage for both kinds of pictures over both kinds of words, because
WORD DISTINCTIVENESS

the pictures are more likely to have two codes regardless of whether they are presented in color or in black and white form. However, the physical distinctiveness account predicts little or no difference in performance when comparing the less distinct black and white pictures to the more distinct color words.

Method

Subjects. Thirty volunteers from Prolific.AC participated, and each was paid the equivalent of £8.00 per hour (prorated). For this and all subsequent experiments, the following inclusion criteria were used: (1) Native speaker of English; (2) approval rating of at least 90% on prior submissions at Prolific.AC; and (3) age between 19 and 39. The mean age was 30.37 (SD = 5.96, range 19-39), and 19 self-identified as female and 11 self-identified as male. The sample size was based on a pilot study.

Materials. The pictures were 167 color line drawings from Rossion and Pourtois (2004), which is a revised version of the Snodgrass and Vanderwart (1980) stimuli. Pictures were excluded if they were best described by a two-word name (e.g., baby carriage). A grayscale version was made of each picture, and additional pictures were discarded if the color and grayscale versions were too similar (e.g., cloud, key, needle). The words were the names of the pictures. The black words were drawn in lowercase 72 point Helvetica Neue and then an image was made. An image was shown rather than drawing text on the browser window to ensure that the exact same font and size was seen by all subjects. When displayed as an image, the text was approximately 32 point in size. The distinctive words were generated using cooltext.com; each word could vary in font style, font size, color, and capitalization. Figure 1 shows some
**WORD DISTINCTIVENESS**

examples.²

**Procedure.** Subjects first answered questions about their age and sex. After reading instructions on a computer screen, subjects clicked on a button to start. Twenty pictures (10 color and 10 black and white) and 20 words (10 black and 10 distinctive) were shown one at a time for 1 s (with a 2 s stimulus onset asynchrony) in the middle of the computer screen either in their actual size (pictures) or reduced such that the black words appeared to be approximately 32 point. All stimuli were shown against a light gray background (#EEEEEE). After all 40 stimuli had been shown, a message appeared informing the subjects that they would now see 40 pictures and 40 words. They were asked to answer the question, “Was this item shown in the list?” They responded by clicking on either the “Yes” button if they believed the stimulus had been shown in the original list, or the “No” button otherwise. Assignment of stimuli to conditions was randomized for each subject, and the order of the study and test items was also randomized for each subject.

**Results and Discussion**

For this and all subsequent experiments, statistical analyses were carried out using R version 3.4.1 (R Core Team, 2017). Generalized eta squared \( \eta_g^2 \); Olejnik & Algina, 2003) is the effect size reported for \( F \) tests and Cohen’s \( d \) (Cohen, 1992) is the effect size reported for \( t \) tests.

Table 1 shows means and standard deviations for hit and false-alarm rates, \( d' \), and \( C \).³

---

² The stimuli are available at https://memory.psych.mun.ca/research/stimuli/j75-pse.shtml or from that last author.

³ When calculating \( d' \) and \( C \), hit and false alarm rates of 1 and 0 were changed to 0.99 and 0.01, respectively.
WORD DISTINCTIVENESS

Overall, the hit rate was higher for pictures ($M = 0.879$, $SD = 0.120$) than words ($M = 0.777$, $SD = 0.124$), but the false-alarm rate was equivalent for pictures ($M = 0.120$, $SD = 0.188$) and words ($M = 0.120$, $SD = 0.157$). A picture superiority effect was observed in discrimination, with $d'$ higher for pictures ($M = 3.139$, $SD = 1.137$) than words ($M = 2.562$, $SD = 0.940$).

A one-way analysis of variance (ANOVA) performed on hit rates produced a main effect of stimulus type, $F(3, 87) = 10.784$, $MS_e = 0.017$, $p < 0.001$, $\eta^2_g = 0.170$. Inspection of Table 1 shows that the hit rate increased from black words to distinctive words to black and white pictures to color pictures. Planned contrasts showed hit-rate advantages for black and white pictures over black words, $t(29) = 3.817$, $p < 0.01$, $d = 0.697$, and for color pictures over distinctive words, $t(29) = 2.176$, $p < 0.05$, $d = 0.397$. However, the difference between black and white pictures and distinctive words was not significant, $t(29) = 0.646$, $p > 0.5$, $d = 0.118$.

A one-way ANOVA performed on false-alarm rates did not yield a significant effect of stimulus type, $F(3, 87) = 1.376$, $MS_e = 0.014$, $p > .25$, $\eta^2_g = 0.013$.

A one-way ANOVA performed on $d'$ produced a main effect of stimulus type, $F(3, 87) = 11.277$, $MS_e = 0.688$, $p < 0.001$, $\eta^2_g = 0.125$. Inspection of Table 1 shows that $d'$ increased from black words to distinctive words to black and white pictures to color pictures. Planned contrasts showed a discrimination advantage for black and white pictures over black words, $t(29) = 3.328$, $p < 0.01$, $d = 0.608$, and no significant discrimination difference between distinctive words and black and white pictures, $t(29) = 0.372$, $p > 0.700$, $d = 0.068$. The difference in discrimination between color pictures over distinctive words did not reach the adopted significance level, $t(29) = 1.881$, $p = 0.070$, $d = 0.343$.

A one-way ANOVA performed on $C$ produced a main effect of stimulus type, $F(3, 87) = 3.845$, $MS_e = 0.234$, $p = .012$, $\eta^2_g = .063$. Inspection of Table 1 shows that responses became
WORD DISTINCTIVENESS

increasingly conservative from color pictures to black and white pictures to distinctive words to black words.

The most notable result from Experiment 1 is that discrimination of the distinctive words and black and white pictures was equivalent by $d'$; that is, there was no picture superiority effect when the words were made more distinctive and the pictures were made less distinctive. Also noteworthy is that the picture superiority effect was larger when comparing black words to black and white pictures than when comparing distinctive words to color pictures, suggesting that an increase in the distinctiveness of the words without a decrease in the distinctiveness of the pictures attenuates the picture superiority effect.

**Experiment 2**

The $d'$ values obtained in Experiment 1 were fairly large, particularly for the most distinctive stimuli, the color pictures. In Experiment 2, we performed a replication of Experiment 1 but we doubled the number of targets and distractors to remove any possibility that ceiling effects were involved in producing the results observed in Experiment 1.

**Method**

**Subjects.** Thirty different volunteers from Prolific.AC participated, and each was paid £8.00 per hour (prorated). The mean age was 28.50 ($SD = 5.96$, range 19-39) and 20 self-identified as female and 10 self-identified as male.

**Materials.** The stimuli were the same as in Experiment 1.

**Procedure.** The procedure was identical to Experiment 1 except that the number of study trials and test trials were each doubled.

**Results and Discussion**

Table 2 shows means and standard deviations for hit and false-alarm rates, $d'$, and C.
WORD DISTINCTIVENESS

Overall, discrimination was better for pictures \((M = 1.990, SD = 0.805)\) than words \((M = 1.481, SD = 0.507)\).

A one-way ANOVA performed on hit rates produced a main effect of stimulus type, \(F(3, 87) = 8.355, MS_e = 0.014, p < 0.001, \eta_g^2 = 0.080\). Table 2 shows that the hit rate increased from black words to color pictures to black and white pictures to distinctive words. Planned contrasts showed a hit-rate advantage for black and white pictures over black words, \(t(29) = 3.301, p < 0.01, d = 0.603\), but no significant differences between distinctive words and color pictures, \(t(29) = 1.143, p > 0.25, d = 0.209\), or distinctive words and black and white pictures, \(t(29) = 1.003, p < 0.30, d = 0.183\).

A one-way ANOVA performed on false-alarm rates produced a main effect of stimulus type, \(F(3, 87) = 5.957, MS_e = 0.017, p < 0.01, \eta_g^2 = 0.075\). Table 2 shows that false-alarm rates increased from color pictures to black and white pictures to distinctive words to black words. Planned contrasts showed significantly higher false-alarm rates to black words than black and white pictures, \(t(29) = 2.967, p < 0.01, d = 0.542\), and to distinctive words than color pictures, \(t(29) = 2.304, p < 0.05, d = 0.421\). However, the false-alarm rate did not significantly differ between distinctive words and black and white pictures, \(t(29) = 1.066, p > 0.25, d = 0.195\).

A one-way ANOVA performed on \(d'\) produced a main effect of stimulus type, \(F(3, 87) = 19.904, MS_e = 0.306, p < 0.001, \eta_g^2 = 0.212\). Table 2 shows that \(d'\) increased from black words to distinctive words to black and white pictures to color pictures. Planned contrasts showed a discrimination advantage for black and white pictures over black words, \(t(29) = 5.567, p < 0.001, d = 1.016\). However, discrimination did not significantly differ between color pictures and distinctive words, \(t(29) = 1.328, p > 0.19, d = 0.242\), or black and white pictures and distinctive words, \(t(29) = 0.001, p > 0.99, d < 0.001\).
WORD DISTINCTIVENESS

A one-way ANOVA performed on C produced a nonsignificant main effect of stimulus type, $F(3, 87) = 2.017, MS_e = 0.193, p = .117, \eta_g^2 = .025$.

As in Experiment 1, Experiment 2 produced a reliable picture superiority effect when black words were compared to black and white pictures. However, increasing the distinctiveness of the words eliminated the picture superiority effect, both when compared to black and white and color pictures.

**Experiment 3**

In Experiments 1 and 2, we increased the distinctiveness of the words and rendered them comparable to black and white pictures in terms of memory performance. The purpose of Experiment 3 was to include a manipulation that decreased the distinctiveness of the pictures. Whereas words in Experiments 1 and 2 were always shown in the same orientation, horizontally within a rectangular envelope, the orientation and envelope of the pictures varied considerably. In Experiment 3, the distinctiveness of the pictures was reduced by using only those items that could be displayed horizontally within a (mostly) rectangular envelope. For example, in the original stimuli, pictures such as carrot, comb, pencil, and many others were shown diagonally. These were edited such that the picture was horizontal (examples are shown in Figure 2). The physical distinctiveness account predicts that the picture superiority effect will reverse: distinctive words should be recognized more accurately than black and white pictures with the same orientation.

**Method**

**Subjects.** Thirty different volunteers from Prolific.AC participated, and each was paid £8.00 per hour (prorated). The mean age was 29.70 ($SD = 5.37$, range 20-39) and 22 self-identified as female and 8 self-identified as male.
WORD DISTINCTIVENESS

**Materials.** The pictures were chosen such that they either already had a horizontal orientation (e.g., truck), could be rotated to have a horizontal orientation (e.g., asparagus), or were replaced by new versions that had a horizontal orientation (e.g., bus). All fit within a (mostly) rectangular envelope, although the height of this rectangle varied and included both short (e.g., needle) and tall (e.g., fence) items. Many pictures had not been used previously because the color and black and white versions were too similar (e.g., nail). There were a total of 41 pictures. The distinctive words were the names of these horizontal pictures. For the color pictures, a subset of those used in Experiment 2 were again used, but they were chosen with the restriction that none had a horizontal orientation or fit within a rectangular envelope. The black words were the names of the color pictures.

**Procedure.** The procedure was identical to that of Experiment 1.

**Results and Discussion**

Table 3 shows means and standard deviations for hit and false-alarm rates, $d'$, and $C$. Discrimination was better for pictures ($M = 2.886, SD = 1.022$) than words ($M = 2.472, SD = 0.836$).

A one-way ANOVA performed on hit rates produced a main effect of stimulus type, $F(3, 87) = 14.360, MSe = 0.014, p < 0.001, \eta^2_g = 0.169$. Table 3 shows that the hit rate increased from black words to black and white pictures to distinctive words to color pictures. Planned contrasts showed a hit-rate advantage for black and white pictures over black words, $t(29) = 3.932, p < 0.001, d = 0.718$, and for color pictures over distinctive words, $t(29) = 2.043, p = 0.050, d = 0.373$. However, the hit rate did not significantly differ between distinctive words and black and white pictures, $t(29) = 0.012, p > 0.99, d = 0.002$.

A one-way ANOVA performed on false-alarm rates produced a main effect of stimulus
**WORD DISTINCTIVENESS**

type, \( F(3, 87) = 13.677, MS_e = 0.011, p < 0.001, \eta_g^2 = 0.201 \). Table 3 shows that the false-alarm rate increased from color pictures to distinctive words to black words to black and white pictures. Planned contrasts showed no significant difference in false-alarm rates between color pictures and distinctive words, \( t(29) = 0.402, p = 0.691, d = 0.073 \). The difference in false-alarm rates between black and white pictures and black words just failed to reach the adopted significance level, \( t(29) = 1.819, p = 0.079, d = 0.332 \). However, the false alarm rate was significantly higher to black and white pictures than distinctive words, \( t(29) = 4.693, p < 0.001, d = 0.857 \).

A one-way ANOVA performed on \( d' \) produced a significant main effect of stimulus type, \( F(3, 87) = 23.671, MS_e = 0.540, p < 0.001, \eta_g^2 = 0.226 \). Table 3 shows that \( d' \) increased from black words to black and white pictures to distinctive words to color pictures. Planned contrasts showed that discrimination was significantly better for color pictures than distinctive words, \( t(29) = 2.384, p = 0.024, d = 0.435 \). However, the difference in discrimination between black and white pictures and black words did not reach the adopted significance level, \( t(29) = 1.981, p = 0.057, d = 0.362 \), whereas discrimination was significantly better for distinctive words than black and white pictures, \( t(29) = 3.171, p < 0.01, d = 0.579 \).

A one-way ANOVA performed on \( C \) produced a main effect of stimulus type, \( F(3, 87) = 10.233, MS_e = 0.190, p = .003, \eta_g^2 = .103 \). Inspection of Table 1 shows that responses were more liberal to black and white pictures than the other stimuli.

In Experiment 3, we were successful in reversing the picture superiority effect by using the distinctive words presented in colorful fonts of varying size and also by using black and white pictures that shared the same orientation and envelope shape. As predicted by the physical distinctiveness account, making the words more distinctive (relative to how they are usually shown) and making the pictures less distinctive (relative to how they are usually shown) reversed
Experiment 4

The purpose of Experiment 4 was to test the distinctiveness account in a slightly different paradigm. As in Experiment 2, eighty words and pictures were again shown at study, but the task at test was to determine the form during the study phase: black word, black and white picture, distinctive word, or color picture. Performance should be best for color pictures and worst for black words because these are the most and least distinctive, respectively. Because of the limited number of black and white pictures with the same orientation and envelope shape (only 41) used in Experiment 3, we reverted to the larger set of black and white pictures used in Experiments 1 and 2. Therefore, we predict no advantage in performance for black and white pictures over distinctive words, but also no reversal.

Subjects. Thirty different volunteers from Prolific.AC participated, and each was paid £8.00 per hour (prorated). The mean age was 25.77 (SD = 5.95, range 19-38) and 23 self-identified as female and 7 self-identified as male.

Materials. The stimuli were the same as those used in Experiments 1 and 2.

Procedure. The study phase was identical to Experiment 2. The test phase was different. Rather than an old/new recognition judgment, subjects were asked to identify the form in which the probe had appeared at study. All four versions of a probe were shown on the screen: black text, distinctive text, black and white picture, and color picture. Subjects were asked to click on the form that appeared at study. The assignment of stimuli to format in the study phase, the order of the items, the test order were randomly determined for each subject, and the order of the forms were randomly determined for each test trial.

Results and Discussion
WORD DISTINCTIVENESS

Table 4 shows the proportion of responses as a function of the study phase format. One immediate concern was the very large number of black word responses relative to the other responses. There were 20 words shown in each format during study, but the mean number of black word, black and white picture, distinctive word, and color picture responses were 27.27, 15.43, 15.80, and 21.50, respectively. Clearly, subjects favored clicking on the item shown in black font. Our interpretation of this bias is that when uncertain, black word was likely to be the default response, most likely on metamemory grounds. We suspect that subjects thought that if the word had been shown in one of the other formats, they would have remembered the additional distinctive features and because they did not, it must have been the most “boring” form that had appeared.

To analyze the data so that both discrimination and bias were taken into account, we used the model of Sridharan, Steinmetz, Moore, and Knudsen (2014). First, the model was fit to each subject’s data and we computed $r^2$ as a measure of fit. This varied from a low of 0.621 to a high of 0.990, with a mean of 0.880 ($SD = 0.095$); 24 of the 30 subjects’ data were fit with $r^2 > 0.80$. The fitted model produces a measure for sensitivity, $s$, and for bias, $c$, for each form for each subject.\footnote{Sridharan et al. (2014) use the term $d$ for sensitivity, but we use $s$ to avoid confusion with either Cohen’s $d$ or $d’$, which are used elsewhere in this paper.} The means are shown in Table 5. Overall, discrimination was better for pictures ($M = 1.225, SD = 0.669$) than words ($M = 1.187, SD = 0.841$).

A one-way ANOVA performed on the $s$ values produced a main effect of stimulus type, $F(3, 87) = 5.840, MS_c = 0.400, p < 0.01, \eta_g^2 = 0.075$. Table 5 shows that discrimination was lowest for black words and highest for distinctive words. Planned contrasts showed that
WORD DISTINCTIVENESS

discrimination did not significantly differ between black words and black and white pictures, 
$t(29) = 0.519, p > 0.60, d = 0.095$, or between distinctive words and color pictures, $t(29) = 0.063, 
p > 0.95, d = 0.012$. However, discrimination was significantly better for distinctive words than 
black and white pictures, $t(29) = 2.452, p < 0.05, d = 0.448$.

The bias data did not meet the sphericity assumption of the repeated-measures ANOVA; 
the F statistic was therefore evaluated with respect to Greenhouse-Geisser adjusted degrees of 
freedom (Greenhouse & Geisser, 1959). The main effect of stimulus type was not significant, 
$F(1.440, 41.758) = 1.473, MS_e = 0.245, p = .239, \eta^2_p = .034$.

In Experiment 4, using a different methodology, we found evidence that increasing the 
distinctiveness of the words, by allowing them to vary in font, font size, and color, eliminated the 
picture superiority effect relative to color pictures and reversed it relative to black and white 
pictures.

**General Discussion**

The purpose of the four experiments was to test the physical distinctiveness account of 
the picture superiority effect. In Experiments 1 and 2, we compared recognition performance for 
words displayed in black font; words displayed in varying distinctive, colorful fonts; black and 
white pictures; and color pictures. Although the standard picture superiority effect was observed 
when black words were compared to color pictures, the picture superiority effect was eliminated 
when distinctive words were compared to black and white pictures, and reduced when distinctive 
words were compared to color pictures. In Experiment 3, we further decreased the 
distinctiveness of the black and white pictures by using pictures of similar objects, all of which 
were presented at the same horizontal angle. Critically, this produced a reversal of the picture 
superiority effect: although color pictures were still recognized better than distinctive words,
distinctive words were significantly better recognized than black and white pictures. Finally, in Experiment 4, we extended the old/new recognition results of Experiments 1-3 to a forced-choice procedure in which subjects were asked to identify the form that each probe took during the study phase. Once again, performance followed distinctiveness, and there was a significant advantage for distinctive words over black and white pictures.

The results of the present studies are consistent with the physical distinctiveness account of the picture superiority effect (see, e.g., Mintzer & Snodgrass, 1999), and inconsistent with dual-coding theory (Paivio, 2007). Below, we consider the theoretical implications of our results. First, however, we address the concern that several of our conclusions rely on null results.

The Null-Hypothesis Objection

In frequentist analyses, null results are inherently ambiguous. They do not provide sufficient evidence to reject the null hypothesis, but they also provide no evidence regarding the truth of the null hypothesis. In Experiments 1 and 2, we found nonsignificant differences when recognition of distinctive words was compared to black and white pictures or color pictures. Similarly, the difference between distinctive words and black and white pictures in Experiment 4 was not significant. One might therefore object that our argument rests on null results.

We do not believe that the reliance on null results is problematic for our conclusions. First, it is not our contention that the increase in distinctiveness of the word stimuli completely abolished the picture superiority effect. We remain agnostic on this point. Rather, we argue that the magnitude of the picture superiority effect was substantially reduced when the words were made more distinctive. Whether this fully abolishes the picture superiority effect, reverses it, or simply reduces it is tangential to our argument. Second, even if one ignored all of the null results, there remains positive results that are consistent with and support the above: In two experiments,
WORD DISTINCTIVENESS

we observed a significant reversal of the picture superiority effect.

Theoretical Considerations

According to dual-coding theory (Paivio, 1971, 1991, 2007), the picture superiority effect obtains because pictures are more often named than words are imaged, and storing representations in more pathways leads to superior memory. From the perspective of dual-coding theory, the only reason the picture superiority effect should fail to emerge is if subjects generate imaginal representations of words at the same rate as they label pictures (Paivio & Csapo, 1973), or if changes in instructions or methods preclude pictures from being named, such as presenting the stimuli too quickly for labeling to be feasible (Paivio & Csapo, 1969). Therefore, a dual-coding interpretation of our results would have to make the assumption that the distinctive words were more often imagined than the black words. It is not clear why subjects would more often generate an image of an alligator if they see the word presented in a colorful, distinctive font than if they see it in a standard, uniform black font.

An alternative account might be that subjects are forming an image of the colorful word itself. That is, rather than forming an image of an alligator, they are forming an image of how the word looks. If this is the case, then dual-coding theory predicts that memory for distinctive words and color pictures should be equivalent because each has two codes. We do not see a plausible way for dual-coding theory to account for the dependence on the size of the picture superiority effect on the perceptual distinctiveness of the words.

Our results are, however, fully consistent with accounts of the picture superiority effect that posit a role for physical distinctiveness (Mintzer & Snodgrass, 1999; Nelson, 1979; Nelson et al., 1977; Weldon & Coyote, 1996). The physical distinctiveness account holds that the picture superiority effect results from the greater variability among pictures compared to words. By
WORD DISTINCTIVENESS

increasing the physical variability of word stimuli, we showed that the picture superiority effect is reduced or eliminated.

To the best of our knowledge, the only previous picture superiority effect study to vary the distinctiveness of words was carried out by Paivio et al. (1968). In their study, distinctive words were presented in different color texts. Unlike the present results, Paivio et al. found no attenuation of the picture superiority effect when words were made distinctive. Two possible explanations for the discrepancy between our results and Paivio et al.’s (1968) occur to us. First, it is possible that color alone is not sufficiently distinctive to affect the magnitude of the picture superiority effect. To frame this another way, adding only one more dimension of perceptual variability to words may not be sufficient given the many different dimensions of perceptual variability in color pictures. In the present study, we varied not only color, but a given word could have multiple colors or color gradients. In addition, we also varied font, font size, and capitalization. We think adding variability along more dimensions could readily lead to a more effective manipulation of distinctiveness.

An additional reason for the discrepancy concerns how memory was tested. In the present work, we assessed memory using old/new (Experiments 1-3) or four-alternative forced-choice (Experiment 4) recognition. In both types of tests, the physical form of the item is present at both study and test. In contrast, Paivio et al. (1968) tested memory with a free-recall test. It is conceivable that physical distinctiveness plays a larger role in recognition than in free recall, and that dual coding plays a larger role in free recall than recognition. Indeed, Paivio (1976, p. 123) suggests exactly this, although he emphasizes that this conclusion “is a relative one”. He still appeals to dual-coding theory to explain many phenomena in recognition. We think it likely that because free-recall tests require subjects to verbally report, write, or type their responses,
WORD DISTINCTIVENESS

verbally labeling pictures is a necessary condition for accurate performance. In contrast, verbally
labelling a picture at study is not a necessary condition for accurate performance when the test
consists of showing the same picture again. It is therefore possible that although pictures are
more often labeled than words (consistent with dual-coding theory), the representation stored in
the logogen pathway plays less of a role in recognition than in free recall.

Conclusion

In four experiments, we found an attenuation of the picture superiority effect when words
were made more distinctive relative to pictures, including reversing the effect. These results are
consistent with the physical distinctiveness account of the picture superiority effect but cannot be
accommodated by dual-coding theory. At least when memory is assessed by recognition, dual-
coding theory does not appear to be a tenable explanation of the picture superiority effect.
References


WORD DISTINCTIVENESS


WORD DISTINCTIVENESS


WORD DISTINCTIVENESS


WORD DISTINCTIVENESS

Author Notes

This research was supported, in part, by a postgraduate scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC) to TME, and by grants from NSERC to IN and AMS.
WORD DISTINCTIVENESS

Table 1

<table>
<thead>
<tr>
<th>Measure</th>
<th>Low Distinctiveness</th>
<th>High Distinctiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Words</td>
<td>Pictures</td>
</tr>
<tr>
<td>Hit</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td></td>
<td>0.719</td>
<td>0.175</td>
</tr>
<tr>
<td>FA</td>
<td>0.151</td>
<td>0.185</td>
</tr>
<tr>
<td>$d'$</td>
<td>2.178</td>
<td>1.074</td>
</tr>
<tr>
<td>$C$</td>
<td>0.341</td>
<td>0.638</td>
</tr>
</tbody>
</table>

Recognition performance for low distinctiveness stimuli (black words, black and white pictures), and high distinctiveness stimuli (color words, color pictures) in Experiment 1.
### WORD DISTINCTIVENESS

Table 2

<table>
<thead>
<tr>
<th>Measure</th>
<th>Low Distinctiveness</th>
<th>High Distinctiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Words</td>
<td>Pictures</td>
</tr>
<tr>
<td>Hit</td>
<td>0.632</td>
<td>0.194</td>
</tr>
<tr>
<td></td>
<td>0.747</td>
<td>0.194</td>
</tr>
<tr>
<td>FA</td>
<td>0.292</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>0.195</td>
<td>0.149</td>
</tr>
<tr>
<td>$d'$</td>
<td>1.091</td>
<td>0.623</td>
</tr>
<tr>
<td></td>
<td>1.896</td>
<td>0.843</td>
</tr>
<tr>
<td>$C$</td>
<td>0.144</td>
<td>0.601</td>
</tr>
<tr>
<td></td>
<td>0.106</td>
<td>0.604</td>
</tr>
</tbody>
</table>

Recognition performance for low distinctiveness stimuli (black words, black and white pictures), and high distinctiveness stimuli (color words, color pictures) in Experiment 2.
WORD DISTINCTIVENESS

Table 3

<table>
<thead>
<tr>
<th>Measure</th>
<th>Low Distinctiveness</th>
<th>High Distinctiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Words</td>
<td>Pictures</td>
</tr>
<tr>
<td>Measure</td>
<td>$M$</td>
<td>$SD$</td>
</tr>
<tr>
<td>Hit</td>
<td>0.696</td>
<td>0.203</td>
</tr>
<tr>
<td>FA</td>
<td>0.149</td>
<td>0.145</td>
</tr>
<tr>
<td>$d'$</td>
<td>1.951</td>
<td>1.001</td>
</tr>
<tr>
<td>$C$</td>
<td>0.317</td>
<td>0.567</td>
</tr>
</tbody>
</table>

Recognition performance for low distinctiveness stimuli (black words, black and white pictures), and high distinctiveness stimuli (color words, color pictures) in Experiment 3. Compared to Experiments 1 and 2, the low distinctiveness pictures were made even more similar to one another.
Table 4: Proportion of responses as a function of the study phase format in Experiment 4. The diagonal shows correct responses. Note: B W = black words; BW P = black and white pictures; C W = color words; C P = color pictures.

<table>
<thead>
<tr>
<th>Study Phase Format</th>
<th>Response</th>
<th>B W</th>
<th>BW P</th>
<th>C W</th>
<th>C P</th>
</tr>
</thead>
<tbody>
<tr>
<td>B W</td>
<td></td>
<td>0.637</td>
<td>0.100</td>
<td>0.148</td>
<td>0.115</td>
</tr>
<tr>
<td>BW P</td>
<td></td>
<td>0.228</td>
<td>0.457</td>
<td>0.068</td>
<td>0.247</td>
</tr>
<tr>
<td>C W</td>
<td></td>
<td>0.312</td>
<td>0.087</td>
<td>0.510</td>
<td>0.092</td>
</tr>
<tr>
<td>C P</td>
<td></td>
<td>0.187</td>
<td>0.128</td>
<td>0.063</td>
<td>0.622</td>
</tr>
</tbody>
</table>
WORD DISTINCTIVENESS

Table 5

<table>
<thead>
<tr>
<th></th>
<th>B W</th>
<th>BW P</th>
<th>C W</th>
<th>C P</th>
</tr>
</thead>
<tbody>
<tr>
<td>s (sensitivity)</td>
<td>0.923</td>
<td>1.010</td>
<td>1.451</td>
<td>1.440</td>
</tr>
<tr>
<td>c (bias)</td>
<td>-0.160</td>
<td>-0.386</td>
<td>-0.307</td>
<td>-0.170</td>
</tr>
</tbody>
</table>

Sensitivity and bias estimates from the Sridharan et al. (2014) model for the four stimulus forms in Experiment 4. Note: B W = black words; BW P = black and white pictures; C W = color words; C P = color pictures.
Figure 1: Examples of the stimuli in Experiments 1, 2, and 4.

<table>
<thead>
<tr>
<th>Image 1</th>
<th>Image 2</th>
<th>Image 3</th>
<th>Image 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>alligator</td>
<td>alligator</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>apple</td>
<td>apple</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>arrow</td>
<td>arrow</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>artichoke</td>
<td>artichoke</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>asparagus</td>
<td>asparagus</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>axe</td>
<td>axe</td>
</tr>
<tr>
<td>[Image]</td>
<td>[Image]</td>
<td>ball</td>
<td>ball</td>
</tr>
</tbody>
</table>
WORD DISTINCTIVENESS

Figure 2: Examples of the black and white images used in Experiment 3.