

Running head: METACOGNITION & PERCEPTUAL CUES

**Memory Predictions are Influenced by Perceptual Information:
Evidence for Metacognitive Illusions**

Matthew G. Rhodes

Colorado State University

Alan D. Castel

University of California, Los Angeles

IN PRESS

Journal of Experimental Psychology: General

Matthew G. Rhodes
Department of Psychology
Colorado State University
Fort Collins, CO 80523-1876
E-mail: matthew.rhodes@colostate.edu
Office: (970) 491-6624
Fax: (970) 491-1032

Abstract

Although perceptual information can capture attention and is utilized when judging size or depth, little work has investigated whether such information is used to make memory predictions. The present study examined how the font size of to-be-remembered words influences predicted memory performance. Participants studied words for a later free recall test that varied in font size and made judgments of learning (JOLs) for each item. JOLs were influenced by font size, with larger font sizes given higher JOLs, whereas little relationship was evident between font size and recall. The effect was modified when other, more valid, sources of information (e.g., associative strength) were available when making JOLs, and persisted despite experience with multiple study-test sessions, using a forgetting scale to assess predictions, and explicitly warning participants that font size has little effect on memory performance. When ease of reading was manipulated, such that large font size words were made less fluent, the effect was eliminated. Thus, highly accessible perceptual cues can strongly influence JOLs, likely via encoding fluency, leading to metacognitive illusions.

Keywords: memory, metacognition, judgments of learning, metamemory, perceptual cues

The perception of information in our environment is often directly related to physical characteristics such as size, intensity or duration. Some of the most influential laws of psychology (e.g., Weber's law, Stevens' power law) are based on these characteristics, and describe how the physical properties of a stimulus are related to its perceived intensity. Furthermore, the visual system makes use of relative size information as a depth cue, conveying important information regarding the proximity of a stimulus. More applied domains such as advertising exploit these principles by using large pictures and colorful slogans that capture consumers' attention and convey information. Thus, physical characteristics such as size are used in a variety of settings to convey information and presumably influence later access to that information. Similarly, one might expect that the perceptual features of new information might influence the degree to which an observer regards that information as memorable.

Indirect support for this assumption comes from a variety of studies that have documented the impact of perception on memory. Typically, the perceptual properties of information, such as luminance and clarity, can have a strong influence on the ability to remember this information. For example, perceptually clear words (Whittlesea, Jacoby, & Girard, 1990; Whittlesea, 1993) are more likely to be regarded as previously studied than words that have been perceptually degraded. In addition, manipulations that enhance the ease with which an item is identified or processed often increase the probability that an item is judged as previously encountered (e.g., Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989; Johnston, Dark, & Jacoby, 1985). The converse, memory influencing perception, has also been frequently demonstrated. For example, prior exposure to an item diminishes judgments of the intensity of background noise (Goldinger, Kleider, & Shelley, 1999; Jacoby, Allan, Collins, & Larwill, 1988) and gives the impression that items appear for a longer duration (Witherspoon & Allan, 1985).

We ask in the current study whether perception may likewise influence and bias judgments of future memory performance. Previous work has demonstrated that people use a variety of cues when making judgments of frequency, liking, and duration, so it may be the case that people use similarly salient and accessible perceptual information when making predictions of how well information will be remembered (even when these cues may have little impact on memory). Thus, in several experiments we manipulated the font size of words presented at study. If predictions of memorability are sensitive to font size, one might expect participants to regard large words as more memorable than small words, even when there is little relation between font size and memory. Such a finding would suggest that the simplest (or most easily accessible) of perceptual characteristics may influence putatively higher-order cognition and provide further evidence regarding those cues that participants use to judge their own learning.

Perception, Attribution, and Metacognition

As noted previously, several researchers have suggested that the availability of perceptual information can influence a wide variety of judgments (see e.g., Jacoby, Kelley, & Dywan, 1989; Kelley & Rhodes, 2002; Schwarz, 2004, for reviews). Borrowing from prior work indicating that judgments of frequency are based on the ease with which examples come to mind (Kahneman & Tversky, 1973), the general thrust of these arguments is that the relative ease with which a stimulus is processed influences judgment (Jacoby & Dallas, 1981). Relative ease is frequently manipulated by varying some perceptual attribute of the stimulus in question. For example, Kleider and Goldinger (2004) administered a recognition test for previously studied faces in which the faces were presented either clearly or in some degree of perceptual distortion. Results showed that participants were substantially more likely to endorse perceptually clear faces as having been studied (see also Whittlesea, 1993; Whittlesea et al., 1990). Other work has

demonstrated that the ease with which a stimulus is identified is positively related to the probability that it is recognized (Jacoby & Dallas, 1981; Johnston et al., 1985).

These findings suggest that the types of cues available influence participants' subjective experience and are used as a basis for judgment. Work from a variety of other domains is consistent with this idea. For example, Werth and Strack (2003) had participants study questions and answers in formats that were easy (high figure-ground contrast) or difficult (low figure-ground contrast) to read. Participants were then asked to judge the likelihood that they would have known the answer. Easily read items elicited higher ratings that the participant would have known the answer. Reber and Schwarz (1999) likewise reported that participants were more likely to endorse statements such as "Osorno is in Chile" if the statement was presented in an easy-to-read color. Oppenheimer and Frank (2008) have further shown that the ease of reading a word (manipulated by varying legibility) influences categorization, with easily read items classified as more typical members of a category. The perceptual clarity of a stimulus may even influence judgments of seemingly unrelated constructs such as intelligence. For example, Oppenheimer (2006) presented participants with texts in an easy- or difficult-to-read font and asked for judgments about the intelligence of the author. Results showed that participants rated an author as more intelligent when the text was in an easy-to-read font, attributing the ease with which the text was read to the intelligence of the author.

Thus, these data demonstrate that the perceptual features of stimuli influence a variety of memorial and nonmemorial judgments by influencing the subjective ease with which a stimulus is processed. However, the link between perception and metacognition has remained largely unexplored (but see, e.g., Busey, Tunnicliff, Loftus, & Loftus, 2000). An extensive literature has documented the manner in which participants make predictions of future memory performance

(e.g., Koriat, 1997; Nelson & Narens, 1990; for reviews see Koriat, 2007; Metcalfe, 2000). For example, one method of studying metacognition is to ask participants to predict future memory performance either immediately after the presentation of an item or following a delay. Such *judgments of learning* (JOLs) allow researchers to compare predicted with actual memory performance. While participants are often reasonably accurate in their predictions, a number of discrepancies between actual and predicted performance have been documented (e.g., Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin, Bjork, & Schwartz, 1998; Carroll, Nelson, & Kirwan, 1997; Castel, McCabe, & Roediger, 2007; Koriat & Bjork, 2005; Mazzoni & Nelson, 1995). For example, Castel et al. (2007) had participants study related (*table-chair*) and unrelated (*gate-bomb*) pairs of words along with identical pairs (*vine-vine*). Participants made JOLs immediately after the presentation of each pair and were later given the first member of the pair (*vine-*) and instructed to recall the word studied with it. Results showed that JOLs consistently overestimated the probability of recalling items from identical pairs, suggesting that fluent processing and perceptual similarity influenced metacognitive judgments.

Discrepancies of this sort provide some indication of how participants make predictions of future memory performance. Of current concern is how participants use the perceptual features of stimuli to make JOLs. Several authors have demonstrated that enhancing the familiarity of a cue, for example through repetition (e.g., Metcalfe, Schwartz, & Joaquim, 1993; see also Jameson, Narens, Goldfarb, & Nelson, 1990), elevates levels of prospective confidence. In addition, there is also some indication that ease of encoding (Begg et al., 1989; Hertzog, Dunlosky, Robinson, & Kidder, 2003) or the ease with which an item is retrieved before a JOL is made (e.g., Benjamin et al., 1998; Matvey, Dunlosky, & Guttentag, 2001) influences such judgments, with participants regarding easily retrieved or processed items as having a high

probability of later recall. However, the prior evidence is at best indirect and an understanding of the potential influence of perceptual qualities of stimuli on predictions of memory performance requires a direct manipulation of such qualities. We undertake such a manipulation in the current study.

The Current Study

In the current study we manipulated the font size of words presented words at study. We were specifically interested in whether participants' JOLs would be sensitive to the size of words even when such differences were generally not diagnostic of memory performance (cf. Begg et al., 1989; Mazzoni & Nelson, 1995). The font size of an item should have relatively little impact on memory performance, as the efficacy of encoding should largely reflect the degree to which participants consider the meaning of each item (cf. Craik & Lockhart, 1972). In the present series of experiments, however, words presented in a large font may be processed in a manner that is subjectively more fluent than the processing of words presented in a smaller font (cf. Reber & Schwarz, 1999; Werth & Strack, 2003). Thus, participants may regard size as an important cue for memory and take it into account when making JOLs even when memory performance is not influenced by font size. Such data would suggest that predictions of future memory performance are influenced by the perceptual features of stimuli—even when those features have little bearing on subsequent performance.

One prior study (Busey et al., 2000) has reported an effect of perceptual information on recognition memory. In particular, Busey et al. had participants study faces at different levels of luminance, such that faces were studied in high to low levels of contrast. Following the presentation of each face participants were instructed to judge their confidence that they could recognize the face on a later test. Results showed that prospective and retrospective confidence

judgments (made at test) were influenced by luminance, as was actual recognition performance. However, whereas luminance had an impact on recognition performance, we do not anticipate that font size will have an impact on recall. Thus, the current study will assess whether participants' predictions use perceptual information even when such information is not diagnostic of future performance, potentially leading to metacognitive illusions.

Experiment 1

Participants in Experiment 1 studied words presented in large (48 point) and small (18 point) fonts. Immediately following the presentation of each word participants made a JOL regarding the probability that the word would be recalled on a later test. If participants regard large items as more memorable than small items then JOLs should be positively related to font size.

Method

Participants. Twenty Colorado State University psychology students participated for partial course credit. Participants were tested individually or in groups of up to eight participants.

Materials. Materials consisted of 42 nouns taken from the Kucera and Francis (1967) norms. These were randomly divided into two sets of 18 items that were presented equally often in 18 or 48 pt Arial font. The two sets were equated for frequency ($M = 46.37$), number of letters, ($M = 5.85$), and number of syllables ($M = 1.97$). The remaining six items served either as primacy or recency buffers, presented equally often in 18 or 48 pt font, and were excluded from all analyses reported.

Procedure. All study and test stimuli were presented in white, lowercase letters in the center of a black background on a computer screen. After providing informed consent

participants were instructed that they would study items presented in different font sizes. Immediately following each item participants were instructed to rate their confidence that they would later be able to recall that item on a scale from 0 (*not confident at all*) to 100 (*very confident*). Participants were encouraged to use the entire range of the scale. Following each item, participants were given four sec to record their JOL on a sheet of paper provided. Study items were presented in a fixed random order (with the condition that no more than three items of the same font size were presented consecutively) at a 5-sec rate. Immediately following the study list participants engaged in a filler task for approximately 5 minutes that required them to write down the states of the United States. Finally, participants were instructed to recall as many of the items as they could remember from the study list on a blank sheet of paper provided. Participants were given 4 minutes for the free recall task.

Results and Discussion.

Predicted and actual recall data are presented as percentages in Figure 1, and show that predictions, but not recall, were influenced by font size. In particular, whereas JOLs varied substantially between items presented in 18 ($M = 48.63$) compared to 48 ($M = 60.81$) pt font, $F(1, 19) = 15.27$, $\eta^2_p = .45$, recall did not differ between the two conditions, $F < 1$. (The alpha level for all statistical tests was set to .05.)

We also examined the relation between font size, JOLs and recall by calculating a nonparametric measure of association, the Kruskal-Goodman gamma correlation, for each participant (Nelson, 1984). If font size is predictive of JOLs, such that higher JOLs are given for items presented in a large font size, a positive correlation should exist between font size and JOLs. Results were consistent with this, as the mean correlation between font size and JOLs ($\gamma = .39$) reliably differed from zero, $t(19) = 4.19$. Conversely, the mean correlation between font size

and recall ($\gamma = .03$), did not differ from zero, $t < 1$.

Taken together, data from Experiment 1 suggest that participants regarded items presented in a large font as more memorable than items presented in a small font. Font size thus had a strong influence on predictions but was unrelated to subsequent recall. However, one might assume that prior experience with a study-test trial would provide sufficient information for participants to learn that the size of items presented for study has little impact on recall. This possibility was examined in Experiment 2 by providing participants with a study-test opportunity for two unique lists of items.

Experiment 2

Participants in Experiment 2 first studied a list of items, half in a small and half in a large font, followed by a test of free recall. They then completed the same procedure a second time with a different list of words. If participants can learn to ignore the perceptual characteristics of items presented for study and instead attend to other potential cues, then the influence of font size should not persist to a second study-test opportunity. Consistent with this idea, Koriat and Bjork (2006) observed enhancements in calibration when participants were exposed to multiple study-test cycles (see also Koriat, 1997). Likewise, Castel (2008) demonstrated that memory predictions became sensitive to serial position effects when participants had prior experience with the availability of serial position information at retrieval. However, the influence of font size on JOLs may be sufficiently strong that participants continue to use it as a basis for judgment even after a prior study-test opportunity. Thus, Experiment 2 attempted to determine whether a prior test experience (cf. Finn & Metcalfe, 2007) would guide JOLs or whether participants would continue to rely on cues available at study (cf. Koriat, 1997).

Method

Participants. Twenty-eight Colorado State University psychology students participated for partial course credit. Participants were tested individually or in groups of up to 8 participants.

Materials. Materials consisted of 84 nouns taken from the Kucera and Francis (1967) norms. These were randomly divided into 4 sets of 18 items that were presented equally often in 18 or 48 pt Arial font and in the first or second study list. The sets were equated for frequency ($M = 46.59$), number of letters, ($M = 5.84$), and number of syllables ($M = 1.88$). The remaining 12 items served either as primacy or recency buffers, presented equally often in 18 or 48 pt font, and were excluded from all analyses reported.

Procedure. The procedure was identical to Experiment 1 with one exception. Specifically, immediately after finishing the free recall test, participants studied another (unique) list of items and repeated the same procedure a second time.

Results and Discussion.

We first analyze JOLs and then examine recall performance and gamma correlations as a function of font size. As these analyses show, font size influenced JOLs on both study-test trials.

Predictions. Mean JOLs by font size and time are shown in the dark bars of Figure 2. These data were analyzed in a 2 (Time: Time 1, Time 2) x 2 (Font Size: 18, 48) repeated-measures analysis of variance (ANOVA). Results showed that participants' JOLs were higher at Time 1 ($M = 46.97$) than at Time 2 ($M = 31.98$), $F(1, 27) = 30.65$, $\eta^2_p = .53$. In addition, JOLs were reliably higher for items presented in a large ($M = 42.67$) compared to small ($M = 36.28$) font, $F(1, 27) = 27.32$, $\eta^2_p = .50$. Font Size did not reliably interact with Time, $F(1, 27) = 1.91$, $p = .18$, $\eta^2_p = .07$. Thus, the influence of font size did not wane with a second study-test

opportunity. Planned comparisons confirmed that at both Time 1, $F(1, 27) = 17.45$, $\eta^2_p = .39$, and Time 2, $F(1, 27) = 15.86$, $\eta^2_p = .37$, JOLs were reliably higher for words presented in a large compared to small font, with essentially no diminution in the magnitude of the effect (as measured by effect size).

Recall. The mean percentage of items recalled is shown in the light bars of Figure 2 and were analyzed using the same factors as the analysis of JOLs. These data showed that recall was marginally lower at Time 1 ($M = 19.54$) than at Time 2 ($M = 22.82$), $F(1, 27) = 4.17$, $p = .05$, $\eta^2_p = .13$. However, recall did not differ for words presented in a large ($M = 21.23$) versus small ($M = 21.13$) font, $F < 1$, nor did Time interact with Font Size, $F(1, 27) = 1.80$, $p = .19$, $\eta^2_p = .06$.

Gamma Correlations. As in Experiment 1, the relation between font size, JOLs and recall was examined by calculating gamma correlations for each participant. These data showed that the correlation between font size and JOLs was reliably different from zero at both Time 1 ($\gamma = .26$), $t(27) = 3.27$, and Time 2 ($\gamma = .23$), $t(27) = 5.02$. Conversely, the correlation between font size and recall did not differ from zero at either Time 1 ($\gamma = -.03$) or Time 2 ($\gamma = .02$), $ts < 1$.

Discussion. Results from Experiment 2 demonstrated that participants continued to use font size as a basis for JOLs even after completing a study-test phase which should have indicated that font size was unrelated to subsequent recall. The stability of this effect leaves some question as to whether participants' predictions would be less sensitive to font size if other, more diagnostic cues were available at study. We test this possibility in Experiment 3.

Experiment 3

Experiments 1 and 2 showed that items presented in a larger font were regarded as more memorable than items presented in a smaller font, despite the fact that memory performance was unrelated to font size. One potential explanation is that participants must make JOLs for a

relatively homogenous set of items in which the only source of variability is the font size in which an item is presented. In Experiment 3 we ask whether introducing another source of variability between items would influence JOLs such that participants would no longer attend to font size. Specifically, participants studied pairs of words, half of which were related (*lips kiss*) and half of which were unrelated (*gate bomb*), with this manipulation of relatedness crossed with font size. Prior work has shown that memory predictions are highly sensitive to semantic relatedness (e.g., Koriat, 1997; Koriat, Bjork, Sheffer, & Bar, 2004), suggesting that relatedness may be a cue of sufficient strength to mitigate any influence of font size. Introducing a more diagnostic source of variability may thus entirely eliminate the influence of font size on JOLs. However, it may be the case that font size is a sufficiently salient cue that continues to influence JOLs even when other (more diagnostic) information is present.

Method

Participants. Twenty-eight Colorado State University psychology students participated for partial course credit. Participants were tested individually or in groups of up to 8 individuals.

Materials. Materials consisted of 48 target words that were paired equally often with a related (e.g., *injury-hurt*) or unrelated (e.g., *cellar-hurt*) cue word. Thus, targets were the same for all participants and only the type of cue was varied. Cue-target pairs were chosen based on the free association norms reported by Nelson, McEvoy, and Schreiber (1999). Forward association values for related pairs ranged from .41 to .75, with an average value of .55. The corresponding value for unrelated pairs was .00. The set of 48 targets were divided into two sets of 24 items, equated for frequency ($M = 66.92$), number of letters, ($M = 4.63$), and number of syllables ($M = 1.19$). Related and unrelated cues used for each target were equated in terms of frequency ($M = 37.32$), number of letters ($M = 5.45$), and number of syllables ($M = 1.71$). Each

pair was presented equally often in a large or small font. An additional set of 6 buffer pairs (3 related, 3 unrelated) was also constructed. Half of the pairs were presented at the beginning of the list and half were presented at the end of the list, equally often in a large or small font, to serve as primacy and recency buffers. These pairs were excluded from the analyses reported.

Procedure. The procedure was identical to Experiment 1 with several important exceptions. First, participants studied related and unrelated pairs of words (instead of single words as in the previous experiments). The change in materials also brought about a small change in instructions for making JOLs. Participants were told to assess the probability, using the scale described previously, of later recalling the target word given the cue. Finally, during the test phase, cue words (e.g., *cellar*?) were shown in a random order (different from that employed at study) and participants were instructed to write down the word that had been paired with the cue at study. Participants were given 10 sec for recall prior to advancing to the next test item. Two pairs, one taken from the primacy buffer and one from the recency buffer, were presented at the beginning of the test in order to familiarize participants with the test procedure. Responses to these pairs were excluded from all analyses.

Results and Discussion.

Predicted and actual recall data are presented in Figure 3 and show that, while JOLs were largely driven by cue-target relatedness, font size continued to have an impact on JOLs.

Predictions. Mean JOLs were examined in a 2 (Font Size: 18, 48) x 2 (Type of Pair: related, unrelated) repeated-measures ANOVA. Results showed that JOLs were substantially higher for related ($M = 74.90$) compared to unrelated ($M = 25.66$) pairs, $F(1, 27) = 163.99$, $\eta_p^2 = .86$, consistent with prior work (e.g., Arbuckle & Cuddy, 1969; Koriat, 1997). Most importantly, JOLs were reliably higher for items in a large ($M = 52.02$) compared to small ($M = 48.55$) font,

$F(1, 27) = 5.58, \eta_p^2 = .17$. Font Size did not interact with the Type of Pair, $F < 1$. Thus, an effect of font size was still apparent even when a more diagnostic basis for JOLs, relatedness, was present.

Recall. Recall data were examined using the same factors as the analysis of JOLs. These data showed that participants recalled significantly more items presented in related ($M = 85.57$) than unrelated ($M = 27.53$) pairs, $F(1, 27) = 543.51, \eta_p^2 = .95$. All other main effects and interactions were not reliable, $F < 1$.

Gamma Correlations. We also examined the relation between font size, JOLs and recall by calculating gamma correlations for each participant. Results showed that font size was more strongly related to JOLs than to recall. In particular, whereas the mean correlation ($\gamma = .11$) between font size and JOLs differed reliably from zero, $t(27) = 4.11$, the mean correlation between font size and recall ($\gamma = -.01$) did not differ from zero, $t(27) < 1$.

Discussion. Experiment 3 showed that font size exerted some degree of influence over JOLs even when other, more diagnostic, cues were present. We do note that, in comparison with the prior experiments, the magnitude of the influence of font size was diminished and considerably smaller than the influence of relatedness. This suggests that while other cues (e.g., relatedness) may inform JOLs, the influence of font size is sufficiently strong to continue to influence JOLs even when other information is available. The fact that font size was still used as a basis for JOLs is also inconsistent with accounts of its influence based on task demands. That is, in prior experiments in which only size varied among items, one could argue that participants were “led” to believe that this source of variability was predictive of future performance. However, the fact that font size continued to influence JOLs even when a more diagnostic source of information was available (one that all participants used) suggests that participants truly

regarded font size as predictive of future performance. Indeed, given the durability of the influence of font size evident in Experiment 3, one wonders whether font size can be ignored at all when making predictions. We attempted a stronger test of this proposition in Experiment 4 by specifically warning participants to ignore font size when making JOLs.

Experiment 4

Experiment 3 showed that providing another, more diagnostic source of variability at encoding was not sufficient to eliminate the influence of font size on JOLs. We used a more overt manipulation of participants' bases for JOLs in Experiment 4 by warning participants that the size of words presented for study was unrelated to future memory performance. However, it is not certain that warning participants to ignore font size will render them completely indifferent to size when making JOLs. In particular, several studies have demonstrated that it is difficult for participants to ignore manipulations of the perceptual features of stimuli or manipulations of the ease with which a stimulus is processed. For example, Lindsay and Kelley (1996) manipulated the ease with which a fragmentary cue could be completed on a test of memory. They observed that participants were more likely to claim that easily completed words had been studied, an effect that persisted even when participants were warned about the nature of the cues (see also Kelley & Lindsay, 1993). Likewise, participants continue to regard background noise as softer when judged in the context of previously presented items, even when informed of the effect (Jacoby et al., 1988). However, other studies have reported changes in performance when participants become aware (Jacoby & Whitehouse, 1989) or are informed (e.g., Kelley & Jacoby, 1996) about manipulations that influence the ease with which an item is processed (see also Alter, Oppenheimer, Epley, & Eyre, 2007; Novemsky, Dahr, Schwarz, & Simonson, 2007).

To summarize, participants in Experiment 4 were specifically warned that the font size of items presented for study would be unrelated to memory performance. If such a warning allows participants to discount font size, then font size should no longer be related to JOLs. However, font size may be a sufficiently salient cue that participants continue to use it as a basis for JOLs even when forewarned.

Method

Participants. Fifty-four Colorado State University psychology students participated for partial course credit. Participants were tested individually or in groups of up to 8 participants.

Materials and Procedure. The materials used and procedure were identical to Experiment 1 with two exceptions. First, immediately prior to the study phase, participants were warned that font size was unrelated to memory performance. The orienting instructions read, in part:

In the following task you will see words presented one-at-a-time in the center of the screen in different font sizes. The size of the word should not affect what you will be able to remember later so do not pay attention to size. However, please do pay attention to each word as your memory for these words will be tested later

As well, after discussing the scale to be used for JOLs, participants were told that:

It has been shown (in previous research) that the size of the word DOES NOT influence how well you will ACTUALLY remember the word, so please keep this in mind when making your predictions.

In addition to this change in instructions, we also administered a 2-item questionnaire to each participant immediately following the recall task. Specifically, each participant was asked whether font size influenced their JOLs. If the answer to that question was “yes”, participants were further asked whether they regarded the large or small words as more memorable.

Results and Discussion.

Mean JOLs and the mean percentage of items recalled are presented in Figure 4. These data show that font size continued to influence JOLs even when participants had been warned to ignore font size. Specifically, JOLs were reliably higher for large ($M = 48.01$) compared to small ($M = 44.40$) font sizes, $F(1, 53) = 7.78$, $\eta_p^2 = .13$. However, the percentage of words recalled did not differ for large ($M = 17.39$) versus small ($M = 16.98$) font sizes, $F < 1$. Gamma correlations were consistent with these data. In particular, whereas the correlation between font size and JOLs ($\gamma = .11$) reliably differed from zero, $t(53) = 3.09$, the correlation between font size and recall ($\gamma = .03$) did not differ from zero, $t < 1$.

Questionnaire Data. Seventeen out of 54 participants (31%) claimed that font size influenced their JOLs with 16 of the 17 participants (94%) who claimed an influence of font size reporting higher JOLs for large compared to small items. Interestingly, the pattern of JOLs did not differ between participants who did or did not claim that large words were more memorable than small words. Specifically, for participants who claimed that words in a large font were more memorable than words in a small font, JOLs averaged 42.67 and 46.99, for small and large words, respectively. The comparable values for participants who claimed no influence of font size was 44.91 (small words) and 48.16 (large words). Thus, the influence of font size was similar for participants who did or did not claim an influence of font size¹.

Discussion. Results from Experiment 4 indicated that the influence of font size on JOLs persisted even when participants were specifically told that font size was unrelated to recall (cf. Lindsay & Kelley 1996). We do note that in comparison to prior experiments the influence of font size was clearly diminished in Experiment 4. For example, the effect size for the difference between JOLs for large versus small font sizes was smaller in Experiment 4 (*Cohen's d* = .20)

than in Experiment 1 (*Cohen's d* = .85). Thus, it may be more accurate to suggest that the warning diminished but did not eliminate the influence of font size.

The pressing question that remains is why participants regarded large items as more memorable than small items despite prior experience (Experiment 2), a more diagnostic source of information available (Experiment 3), and a specific warning to ignore font size when making JOLs (Experiment 4). We suggest that the subjective experience of large items was significantly more fluent than that for small items, leading participants to use size as a cue for predicting subsequent recall. However, before examining a fluency-based explanation of these data, we consider a potential account based on the scale used for JOLs.

Experiment 5

One possible explanation for the influence of font size on JOLs is that participants simply mapped the dimension manipulated at study (i.e., size) onto the scale used for JOLs. Robertson and Gomez (1980; see also Gomez & Robertson, 1979) have made a similar argument based on their studies of duration estimates. In particular, they presented participants with small and large shapes for very brief durations (ranging from 15-45 msec) and asked for ratings of duration. Robertson and Gomez (1980) reported that size was positively related to judgments of duration, such that larger shapes were given longer judgments of duration. Among other explanations, they considered a strategy hypothesis in which variability from a nontemporal dimension (object size) was used as a basis for judgments on a temporal dimension. That is,

...larger stimuli will be judged longer in temporal extent because “large” and “long” are “more” of the respective dimensions, while “small” and “short” are “less” of the dimensions. (p. 114)

In a similar vein, large words may be regarded as more congruent with portions of the scale used for JOLs that represent larger values, resulting in higher JOLs for large items. We tested this explanation in Experiment 5 by requiring participants to make a *judgment of forgetting* (JOF). Specifically, instead of judging the probability that an item would be *remembered* on a future test, participants in Experiment 5 judged the probability that an item would be *forgotten* on a future test². If participants are mapping font size onto the scale used for judgments, then JOFs should be higher for large words than small words. However, if participants regard large words as being more memorable (i.e., less likely to be forgotten) than small words then JOFs should be smaller for large words.

Method

Participants. Eighteen Colorado State University psychology students participated for partial course credit. Participants were tested individually or in groups of up to 4 participants.

Materials and Procedure. The materials used and procedure were identical to Experiment 1 with one key exception. Specifically, immediately after each word was presented, participants judged their confidence that they would later forget this word on a scale from 0 (*absolutely unlikely to be forgotten*) to 100 (*absolutely likely to be forgotten*).

Results and Discussion.

Mean JOFs and recall performance is presented in Figure 5. These data showed that participants continued to regard large words as more memorable (i.e., less likely to be forgotten) than small words, even when a different scale was used. In particular, participants made higher JOFs for words presented in a small ($M = 55.20$) compared to large ($M = 41.20$) font, $F(1, 17) = 9.47$, $\eta_p^2 = .36$. However, recall did not differ between font sizes, $F < 1$.

Gamma correlations were also calculated to examine the relation between font size, JOFs, and recall. These data showed that whereas the mean correlation ($\gamma = -.39$) between font size and JOLs differed reliably from zero, $t(17) = 3.42$, the mean correlation between font size and recall ($\gamma = .06$), did not differ from zero, $t < 1$. Thus, a reliable negative relationship was apparent between JOFs and font size (i.e., lower predictions of forgetting associated with a larger font size), whereas no relationship was apparent between font size and recall.

Discussion. Data from Experiment 5 demonstrated that the impact of font size on predictions of memory performance was not the product of mapping one dimension (size) onto the scale used for JOLs (cf. Robertson & Gomez, 1980). Specifically, even when predicting forgetting, participants exhibited lower JOFs for items in a large font than items in a small font. The effect obtained was similar to that evident in Experiment 1, which solicited conventional predictions of remembering (i.e., JOLs). For example, if JOFs are transformed to JOLs by subtracting each observation from 100, the resulting means (58.80 and 44.80 for the large and small font sizes, respectively) are almost identical to the JOLs reported in Experiment 1. More formal analyses using these transformed data and examining JOLs as a function of Font Size (small, large) and Experiment (1, 4) confirmed that JOLs did not differ between experiments, $F < 1$. Thus, the influence of large fonts on predictions is not an artifact of the scale used and instead may be the result of another factor, which we explore in Experiment 6.

Experiment 6

Results from the prior experiments have all converged on a similar pattern: participants regard large items as more memorable than small items despite the fact that size has little bearing on subsequent recall performance. We suggest that this occurred because participants regarded large items as subjectively more fluent, and thus more memorable, than small items. Other work

has shown that predictions of future performance can be sensitive to fluency (Begg et al. 1989; Benjamin et al., 1998; Hertzog et al., 2003; Koriat & Ma'ayan, 2005; Matvey, Dunlosky, & Guttentag, 2001; see also Kelley & Lindsay, 1993). For example, Begg et al. suggested that "...[participants] predict success for items that are easiest to process in the manner demanded by the task" (p. 610). As support for this contention, Begg et al. (Experiment 1) demonstrated that memorability and ease-of-studying judgments were sensitive to the same variables (e.g., item concreteness, item frequency). However, none of these prior studies have examined the potential role of fluency in JOLs via a manipulation of the perceptual features of to-be-remembered stimuli.

As a prelude to the present experiment, we had participants make ratings of the ease with which they perceived items in a large or small font. Specifically, participants simply rated the ease with which items could be read. Results showed that ease-of-reading judgments were reliably higher for large words than small words, suggesting that large words were perceived in a manner that was subjectively more fluent than the processing of small words.³ However, such data are largely indirect and do not necessarily ascribe a causal role for fluency in the current study. Thus, we attempted a more direct manipulation of fluency in Experiment 6 by varying the ease with which items could be read. In particular, participants studied items in which each letter was presented in lowercase and in which items were presented in an altered format, with uppercase letters alternating with lowercase letters (e.g., PiAnO). This manipulation of fluency was a within-list variable, such that half of the items presented (both in the normal and alternating condition) were in a large font and half in a small font with these items intermixed within the full list.

Presentation in an alternating format should diminish the fluency with which a word is read. Therefore, if large items are given higher JOLs because they are perceived in a manner that is subjectively more fluent than small items, presenting words in an alternating format should eliminate the difference in JOLs between large and small items. That is, the alternating format should make the perception of large items markedly less fluent, reducing JOLs for large items presented in an alternating format compared to large items presented in the standard format. Such data would suggest that fluency plays a causal role in the higher JOLs given for large compared to small font sizes in the previous experiments.

Method

Participants. Twenty Colorado State University psychology students participated for partial course credit. Participants were tested individually or in groups of up to 8 participants.

Materials. Materials consisted of 54 nouns taken from the Kucera and Francis (1967) norms. These were randomly divided into 4 sets of 12 items that were presented equally often in 18 or 48 pt Arial font and in a *standard format* (i.e., each letter in lowercase) or in an *alternating format* by alternating uppercase with lowercase letters (e.g., PiAnO). The sets were equated for frequency ($M = 46.42$), number of letters, ($M = 5.90$), and number of syllables ($M = 1.79$). The remaining 6 items served either as primacy or recency buffers, presented equally often in 18 or 48 pt font, and were excluded from all analyses reported.

Procedure. The procedure was identical to Experiment 1 with the exception that half of the items were presented in an alternating format and half in a standard format. Items were presented in a fixed random order with the condition that no more than 3 items of the same format or size were presented consecutively.

Results and Discussion.

Predicted and actual recall data are presented in Figure 6. As can be seen, while JOLs were higher in the standard condition for large compared to small items, no such difference was evident between large and small items in the alternating format condition.

Predictions. Mean JOLs were analyzed in a 2 (Font Size: 18 pt, 48 pt) x 2 (Font Format: standard, alternating) repeated measures ANOVA. Results showed that participants made reliably higher JOLs for words presented in a large ($M = 57.16$) compared to small ($M = 48.94$) font, $F(1, 19) = 15.21$, $\eta_p^2 = .45$. In addition, JOLs were higher for items presented in a standard ($M = 57.28$) compared to alternating ($M = 48.81$) format, $F(1, 19) = 6.41$, $\eta_p^2 = .25$. More importantly, a reliable Font Size x Font Format interaction was present, $F(1, 19) = 16.26$, $\eta_p^2 = .46$. This reflects the fact that, in the standard font condition, items in a large font were given reliably higher JOLs than items in a small font, $F(1, 19) = 29.18$, $\eta_p^2 = .61$. Conversely, in the alternating font condition, JOLs did not differ by font size, $F < 1$. Thus, font size only influenced JOLs for items in a standard font, suggesting that font size had its effect by influencing the ease with which items were read. When words were presented in an alternating format reading became uniformly difficult (dysfluent), rendering font size irrelevant to JOLs.

Recall Data. Mean recall performance was analyzed in a 2 (Font Size: 18 pt, 48 pt) x 2 (Font Format: standard, alternating) repeated-measures ANOVA. Results showed that recall did not differ for items presented in a large ($M = 14.58$) compared to small ($M = 12.92$) font, $F(1, 19) = 1.20$, $p = .29$, $\eta_p^2 = .06$, nor did recall differ by font format, $F < 1$. In addition, Font Size did not interact with Font Format, $F < 1$.

Gamma Correlations. Overall, the correlation between font size and JOLs ($\gamma = .26$) reliably differed from zero, $t(19) = 4.17$, whereas the correlation between font size and recall ($\gamma = .06$) did not differ from zero, $t < 1$, replicating prior experiments. We also examined correlations

for each font format. Results for the standard format condition indicated that the correlation between font size and JOLs ($\gamma = .48$) reliably differed from zero, $t(19) = 5.67$, whereas the correlation between font size and recall ($\gamma = .16$) did not differ from zero, $t(18) = 1.19$. For the alternating format condition, neither the correlation between font size and JOLs ($\gamma = .11$), $t(19) = 1.31$, nor the correlation between font size and recall ($\gamma = -.12$), $t < 1$, differed from zero. Thus, font size only influenced predictions when participants were presented with items in a standard font. As further evidence of this, a follow-up analysis confirmed that the correlation between font size and JOLs was reliably greater for the standard compared to the alternating format condition, $F(1, 19) = 11.29$, $\eta_p^2 = .37$.

Discussion. Results from the standard format condition were consistent with the prior experiments. That is, in the standard format condition, participants used the size of items presented for study as a basis for JOLs, with large items given higher JOLs than small items and a reliable relationship ($\gamma = .48$) evident between font size and JOLs, as has been shown in the previous experiments. However, critically for the present arguments, a very different pattern was apparent in the alternating format condition. In that condition, JOLs did not differ by font size nor was the correlation between font size and JOLs reliable⁴. Thus, diminishing the fluency with which items were read (as occurred in the alternating format condition) eliminated the influence of font size. We suggest that such data implicate perceived fluency as the mechanism by which large items are regarded as more memorable than small items. We consider the implications of this and other issues in the General Discussion.

General Discussion

The current study examined whether participants' predictions of future memory performance would be influenced by the perceptual qualities (in the present examination, font

size) of words presented for study, even when such perceptual features were largely unrelated to memory performance. The primary results can be summarized as follows:

- 1.) Participants' JOLs were strongly influenced by font size even though actual recall performance was unrelated to font size (e.g., Experiment 1).
- 2.) The influence of font size persisted across an additional study-test trial (Experiment 2), when more diagnostic bases for judgment were present (Experiment 3), when participants were explicitly warned to disregard font size (Experiment 4), and when judgments of forgetting were elicited (Experiment 5).
- 3.) Large items were perceived in a subjectively more fluent manner than small items (Experiment 6), leading participants give higher JOLs for large compared to small items.

Taken together, these data indicate that participants made predictions of future memory performance based on salient perceptual information available at encoding and not on the processing that occurred after initial perception (cf. Shaw & Craik, 1989). Why might participants regard larger stimuli as being more memorable than smaller stimuli? We suggest that participants misinterpreted the subjective ease of processing engendered by large items as predictive of future ease of retrieval (cf. Benjamin et al., 1998), even though size was unrelated to recall. For example, in Experiment 6, manipulating the ease with which items could be read by presenting items an alternating format reduced the fluency with which items were perceived and negated the influence of font size. That is, making perception dysfluent eliminated the difference in JOLs for large and small items, providing evidence that fluency played a causal role in the higher JOLs given for large items in the current study.

In some ways, it is perhaps not surprising that participants attended to size in their JOLs, although this led to metacognitive illusions (cf. Rhodes & Jacoby, 2007). In particular, given

that studied words were relatively similar on several dimensions (e.g., frequency, length), the size of a word likely became a salient source of information for predictions, even when such information was not diagnostic of future memory performance (cf. Kelley & Rhodes, 2002). Such a conclusion is consistent with Koriat's (1997) cue-utilization framework which suggests that JOLs are the product of inferences made about future memory performance based on the cues available to the learner. Specifically, Koriat suggests that participants often make JOLs based on *intrinsic cues*, defined as "...those characteristics of the study items that are perceived to disclose the items' a priori ease or difficulty of learning" (p. 350) at the expense of other bases for JOLs, such as the conditions of learning (*extrinsic cues*) or information derived from one's own memory performance (*mnemonic cues*). As indicative of this approach, Koriat et al. (2004) had participants study related and unrelated words pairs and make immediate JOLs concerning the probability of recalling the second word of the pair either immediately, after one day, or after one week. Results showed that whereas recall was far better when tested immediately rather than one day or one week later, JOLs were sensitive to relatedness but entirely insensitive to the retention interval. Thus, participants made JOLs based on the nature of the items presented (an *intrinsic cue*) at the expense of other, more diagnostic information.

The current study extends this concept to suggest that the perceptual features of stimuli belong in the class of intrinsic cues that are attended to at the expense of other, more diagnostic information that is not readily available from perception. Interestingly, the influence of such perceptual information persisted even when other dimensions of to-be-remembered stimuli that are predictive of performance (e.g., relatedness, as in Experiment 3) were present or when participants were specifically instructed to ignore size (Experiment 4). While this serves to make an account of these data based on demand characteristics largely untenable it also begs the

question of the extent to which perceptual information will be attended to at the expense of other information available at study. For example, will any variation in the perceptual features of stimuli (e.g., differences in color) be used as a basis for JOLs? In fact, we do not suggest that any manipulation of perceptual information available at study will impact predictions of learning. Rather, we suggest that only perceptual information which influences the subjective experience of fluency will impact predictions of learning. Thus, manipulating the color of stimuli without affecting clarity would have no impact on JOLs, whereas presenting items in differing degrees of perceptual distortion (e.g., Whittlesea et al., 1990) would be predicted to impact predictions of learning. Busey et al.'s (2000) finding that faces presented at high levels of contrast were regarded as more memorable than faces presented at low levels of contrast is consistent with such a conclusion (see also Experiment 6 in the current study).

It may be the case that perceived importance (cf. Castel, 2007) also influenced JOLs in the present work. Although our position is that fluency is likely the basis for most JOLs when font size varies, size may also reflect the importance of to-be-learned information, with large items considered to be more important than small items and thus given higher JOLs. This might be especially true if only a small number of words were presented in a larger font size relative to small font words, leading to a possible effect of salience or an isolation effect (cf. Dunlosky, Hunt, & Clark, 2000). However, in the present set of studies, an equal number of small and large words were presented, and participants were instructed that all words were important to remember for the later memory test. Further work is needed to determine how these two variables might interact in other settings (e.g., educational settings, advertising, newspaper headlines, or when fine print is present in contracts, for applied examples). It may be the case

that font size is an unusually accessible and highly available cue when making metacognitive judgments and for this reason it is given priority but only in the absence of other relevant cues.

The findings reported in the current study, suggesting that perceptual cues are used to guide metacognitive judgments, also have applied and pedagogical aspects that merit consideration. For example, if students have difficulty perceiving information (e.g., need glasses, listening to music while studying), then they will likely be aware that learning may be impaired. However, if students integrate highly available perceptual cues (e.g., size, bold, or highlighted text) as a means of assessing learning, they may misattribute any enhancements in perceptual clarity to mastery of material. Thus, if students use perceptual cues that enhance initial processing (but not later memory), subsequent performance may reflect overconfidence.

Summary and Conclusions

Overall, the current data are consistent with other work documenting the link between perception and cognition. Previous work has shown that perceptually clear statements or words are more likely to be regarded as having been known all along (Werth & Strack, 2003), to have a higher truth value (Reber & Schwarz, 1999), to be seen as more typical members of a category (Oppenheimer & Frank, 2008) and to produce higher ratings of the intelligence of the author (Oppenheimer, 2006). However, the current study extends the prior literature on the relation between cognition and perception to show that the size of a stimulus influences its perceived memorability, even when size is largely unrelated to memory performance. We do not suggest that the perceptual features of stimuli will always be unrelated to memory performance. Rather, we suggest that people use fluently processed perceptual information that is highly accessible at encoding when making memory predictions, leading to metacognitive illusions when such information does not contribute to actual memory performance.

References

- Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology*, 81, 126-131.
- Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, 136, 569-576.
- Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language*, 28, 610-632.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metacognitive index. *Journal of Experimental Psychology: General*, 127, 55-69.
- Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts of the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review*, 7, 26-48.
- Carroll, M., Nelson, T. O., & Kirwan, A. (1997). Tradeoff of semantic relatedness and degree of overlearning: Differential effects on metamemory and long-term retention. *Acta Psychologica*, 95, 239-253.
- Castel, A. D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 225-270). London: Academic Press.
- Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition*, 36, 429-437.

- Castel, A. D., McCabe, D. P., & Roediger, H. L. III. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*, 107-111.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.
- Dunlosky, J., Hunt, R. R., & Clark, E. (2000). Is perceptual salience needed in explanations of the isolation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 649-657.
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning Memory and Cognition*.
- Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition*, *27*, 328-338.
- Gomez, L. M., & Robertson, L. C. (1979). The filled-duration illusion: The function of temporal and nontemporal set. *Perception & Psychophysics*, *25*, 432-438.
- Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 22-34.
- Jacoby, L. L., Allan, L. G., Collins, J. C., & Larwill, L. K. (1988). Memory influences subjective experience: Noise judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 240-247.
- Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 306-340.

- Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger and F.I.M. Craik (Eds.), *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, (pp. 391-422). Hillsdale, NJ: Erlbaum.
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118, 126-135.
- Jameson, K. A., Narens, L., Goldfarb, K., & Nelson, T. O. (1990). The influence of near-threshold priming on metamemory and recall. *Acta Psychologica*, 73, 55-68.
- Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 3-11.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237-251.
- Kelley, C. M., & Jacoby, L. L. (1996). Adult egocentrism: Subjective experience versus analytic bases for judgment. *Journal of Memory & Language*, 35, 157-175.
- Kelley, C. M., & Lindsay, S. D. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language*, 32, 1-24.
- Kelley, C. M., & Rhodes, M. G. (2002). Making sense and nonsense of experience: Attributions in memory and judgment. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 293-320). New York: Academic Press.
- Kleider, H. M., & Goldinger, S. D. (2004). Illusions of face memory: Clarity breeds familiarity. *Journal of Memory and Language*, 50, 196-211.

- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349-370.
- Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, and E. Thompson (Eds.), *The Cambridge Handbook of Consciousness* (pp. 289-325). New York: Cambridge UP.
- Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31, 187-194.
- Koriat, A., & Bjork, R. A. (2006). Illusions of competence during study can be remedied by manipulations that enhance learners' sensitivity to retrieval conditions at test. *Memory & Cognition*, 34, 959-972.
- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, 133, 643-656.
- Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, 52, 478-492.
- Kucera, H. & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Lindsay, S. D., & Kelley, C. M. (1996). Creating illusions of familiarity in a cued recall remember/know paradigm. *Journal of Memory and Language*, 35, 197-211.
- Matvey, G., Dunlosky, J., & Guttentag, R. (2001). Fluency of retrieval at study affects judgments of learning (JOLs): An analytic or nonanalytic basis for JOLs? *Memory & Cognition*, 29, 222-233.

- Mazzoni, G., & Nelson, T. O. (1995). Judgments of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 1263-1274.
- Metcalfe, J. (2000). Metamemory: Theory and data. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 197-211). New York: Oxford UP.
- Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 1410-1423.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1999). *The University of South Florida word association, rhyme, and word fragment norms*. Available at <http://w3.usf.edu/FreeAssociation/>.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95, 109-133.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (pp. 125-173). New York: Academic Press.
- Novemsky, N., Dahr, R., Schwarz, N., & Simonson, I. (2007). Preference fluency in choice. *Journal of Marketing Research*, 44, 347-356.
- Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology*, 20, 139-156.
- Oppenheimer, D. M., & Frank, M. C. (2008). A rose in any other font wouldn't smell as sweet: Effects of perceptual fluency on categorization. *Cognition*, 106, 1178-1194.

- Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness & Cognition*, 8, 338-342.
- Rhodes, M. G., & Jacoby, L. L. (2007). Toward analyzing cognitive illusions: Past, present, and future. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 379-393). New York: Psychology Press.
- Robertson, L. M., & Gomez, L. (1980). Figural vs. configural effects in the filled duration illusion. *Perception & Psychophysics*, 27, 111-116.
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14, 332-348.
- Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging*, 4, 131-135.
- Werth, L., & Strack, F. (2003). An inferential approach to the knew-it-all-along phenomenon. *Memory*, 11, 411-419.
- Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 19, 1235-1253.
- Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, 29, 716-732.
- Witherspoon, D., & Allan, L. G. (1985). The effects of a prior presentation on temporal judgments in a perceptual identification task. *Memory & Cognition*, 13, 103-111.

Author Note

We would like to thank Jeffrey Boman, Talyn Olguin, Katie Sammons, Haley Smith, and Torrey Trover for assistance with data collection, and Asher Koriat, Danny Oppenheimer, David McCabe, and Colleen Kelley for useful comments on a previous version of this manuscript. Please address all correspondence to Matthew G. Rhodes, Department of Psychology, Fort Collins, CO, 80523-1876. E-mail: matthew.rhodes@colostate.edu.

Footnotes

- 1 We also analyzed JOLs for these groups, albeit with large differences in sample size, in a 2 (Font Size: large, small) x 2 (Group: claim uninfluenced, claim influenced) mixed-factor ANOVA. These data showed that JOLs did not differ between groups, $F < 1$, nor did group interact with Font Size, $F < 1$.
- 2 Koriat et al. (2004; Experiment 7) have used a similar manipulation and reported that having participants frame judgments in terms of forgetting improved the calibration of predictions with respect to different retention intervals. However, whereas aggregate judgments of forgetting were solicited by Koriat et al., item-by-item judgments were used in Experiment 5.
- 3 The reader may contact the first author (M. G. Rhodes) for a full description and analysis of this experiment.
- 4 We do note that JOLs for items in the 18 pt, alternating format condition were numerically but not reliably lower than JOLs for items in the 18 pt, standard format condition. One possibility is that this reflects Type II error and with greater power a reliable difference would be detected. However, another possibility is that presenting items in a small font is sufficiently dysfluent that even presentation in an alternating format leaves little possibility to make perception even more dysfluent. This may be contrasted with a presentation in a large font which is typically perceived in a highly fluent manner and thus allows ample opportunity for fluency to be diminished.

Figure Caption

Figure 1. Predicted and actual recall performance by font size in Experiment 1. Error bars reflect standard errors of the mean in all figures.

Figure 2. Predicted and actual recall performance by font size and test time in Experiment 2.

Figure 3. Predicted and actual recall performance by font size and pair type in Experiment 3.

Figure 4. Predicted and actual recall performance by font size in Experiment 4, in which participants were explicitly warned that font size was unrelated to later recall.

Figure 5. Predicted and actual recall performance by font size in Experiment 5. Note that in Experiment 5 participants made a *judgment of forgetting* (JOF), judging for each word the probability that it would be forgotten on a future test.

Figure 6. Predicted and actual recall performance by font size and format in Experiment 6.

Figure 1.

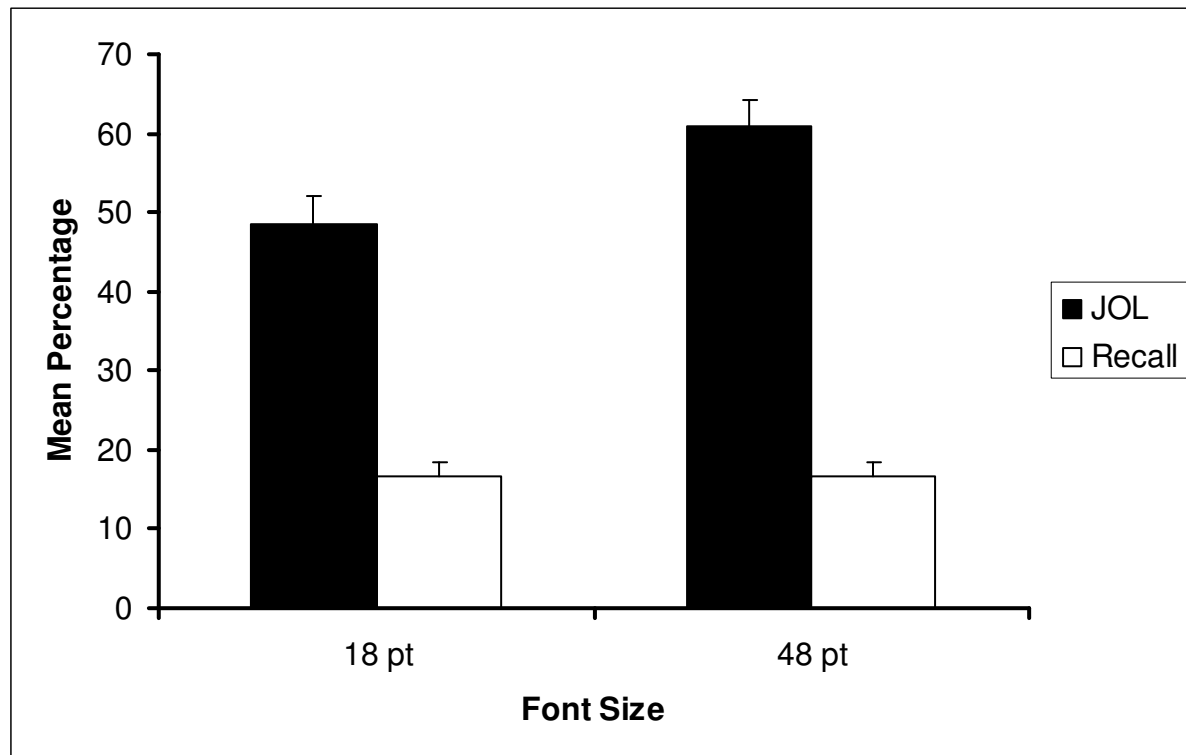


Figure 2.

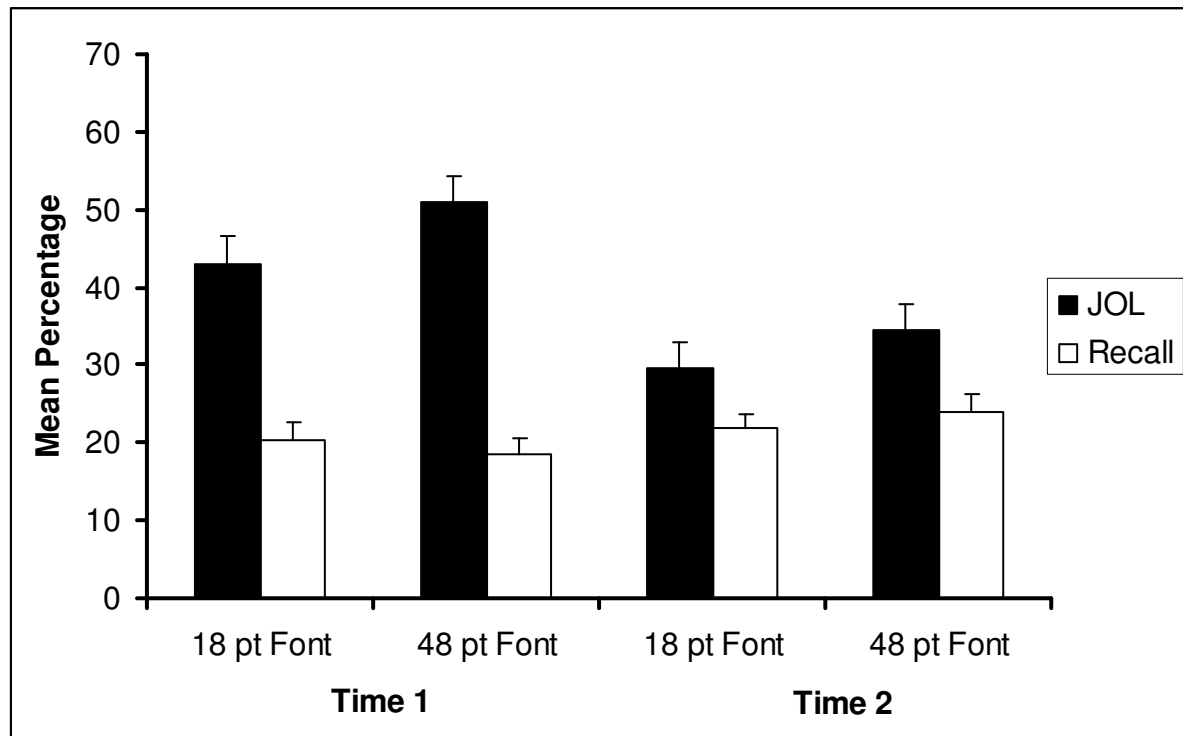


Figure 3.

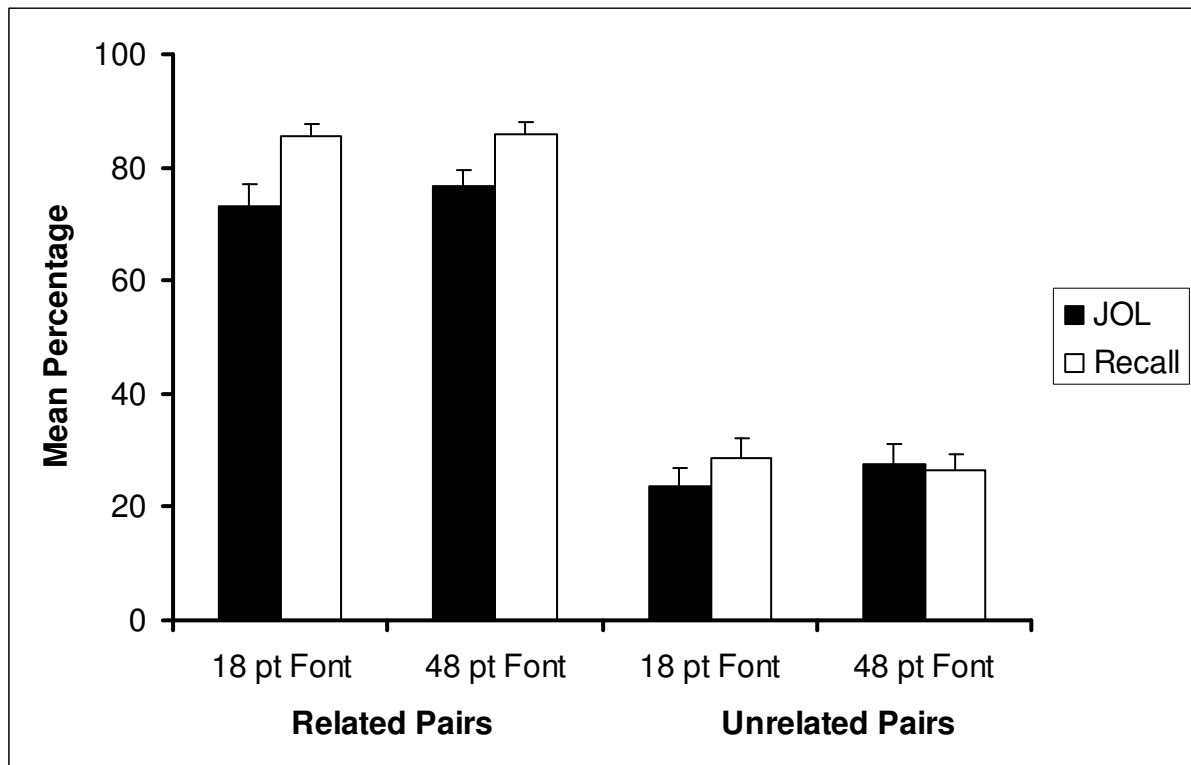


Figure 4.

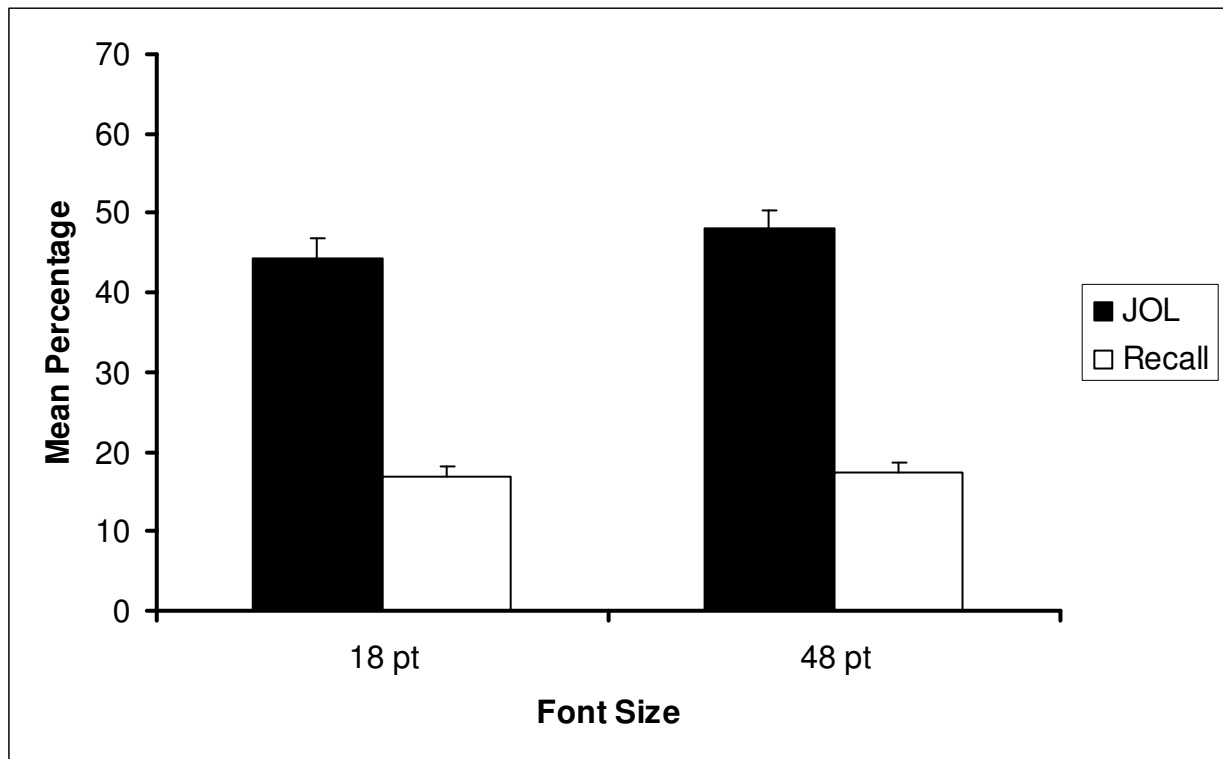


Figure 5.

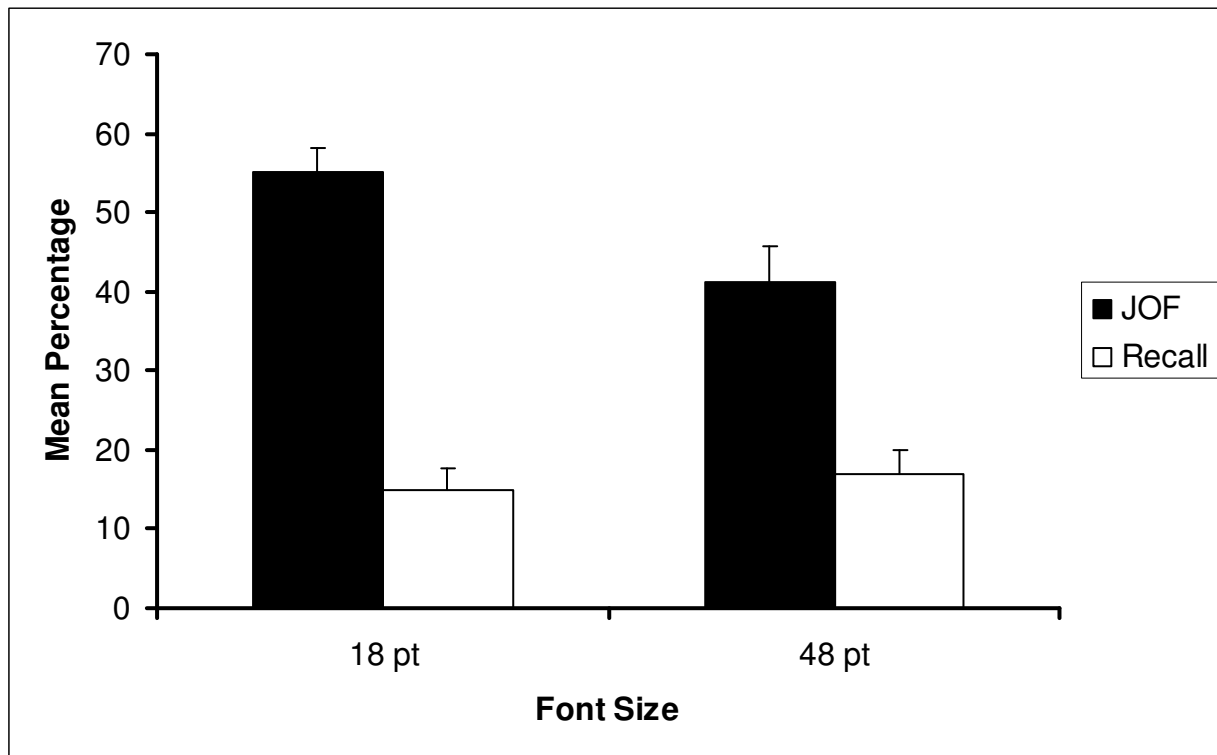


Figure 6.

