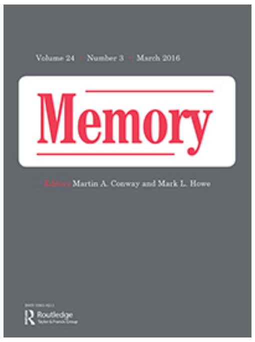

Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: https://www.tandfonline.com/loi/pmem20

Influence of cue word perceptual information on metamemory accuracy in judgement of learning

Xiao Hu, Zhaomin Liu, Tongtong Li & Liang Luo

To cite this article: Xiao Hu, Zhaomin Liu, Tongtong Li & Liang Luo (2016) Influence of cue word perceptual information on metamemory accuracy in judgement of learning, Memory, 24:3, 383-398, DOI: 10.1080/09658211.2015.1009470

To link to this article: https://doi.org/10.1080/09658211.2015.1009470

Published online: 16 Feb 2015.

Memory, 2016
Vol. 24, No. 3, 383–398, http://dx.doi.org/10.1080/09658211.2015.1009470

Influence of cue word perceptual information on metamemory accuracy in judgement of learning

Xiao Hu[1], Zhaomin Liu[2], Tongtong Li[1], and Liang Luo[3]

[1]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China
[2]School of Sociology, China University of Political Science and Law, Beijing, China
[3]State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute for Brain Research and Center for Collaboration and Innovation in Brain and Learning Sciences, Beijing Normal University, Beijing, China

(*Received 2 June 2014; accepted 14 January 2015*)

Previous studies have suggested that perceptual information regarding to-be-remembered words in the study phase affects the accuracy of judgement of learning (JOL). However, few have investigated whether the perceptual information in the JOL phase influences JOL accuracy. This study examined the influence of cue word perceptual information in the JOL phase on immediate and delayed JOL accuracy through changes in cue word font size. In Experiment 1, large-cue word pairs had significantly higher mean JOL magnitude than small-cue word pairs in immediate JOLs and higher relative accuracy than small-cue pairs in delayed JOLs, but font size had no influence on recall performance. Experiment 2 increased the JOL time, and mean JOL magnitude did not reliably differ for large-cue compared with small-cue pairs in immediate JOLs. However, the influence on relative accuracy still existed in delayed JOLs. Experiment 3 increased the familiarity of small-cue words in the delayed JOL phase by adding a lexical decision task. The results indicated that cue word font size no longer affected relative accuracy in delayed JOLs. The three experiments in our study indicated that the perceptual information regarding cue words in the JOL phase affects immediate and delayed JOLs in different ways.

Keywords: Memory; Metamemory; Judgement of learning; Perceptual information.

Metamemory refers to the processes for monitoring and controlling memory activities (Nelson & Narens, 1990). The judgement of learning (JOL) is an important form of metamemory monitoring and refers to judgements of the recall possibilities for memory tests after the study phase (Dunlosky & Metcalfe, 2009). Many studies have shown that JOLs significantly influence subsequent learning processes; people rely on JOLs to adjust subsequent study time allocation (Metcalfe & Kornell, 2005; Nelson, 1993; Son & Metcalfe, 2000) and the selection of study items (Metcalfe & Finn, 2008a; Rhodes & Castel, 2009; Tullis & Benjamin, 2012), for example. Thus, JOL has always been a key focus for metamemory researchers (Bjork, Dunlosky, & Kornell, 2013).

Many prior studies have investigated the cues and heuristics that form the basis of JOLs (Hertzog, Dunlosky, Robinson, & Kidder, 2003; Koriat, 1997; Koriat & Ma'ayan, 2005). One important heuristic that affects JOLs is perceptual information. Research shows that perceptual

Address correspondence to: Zhaomin Liu, School of Sociology, China University of Political Science and Law, Beijing 100088, China. E-mail: zhaominl@cupl.edu.cn; Liang Luo, State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China. E-mail: luoliang@bnu.edu.cn

information has a significant influence on JOLs and that people give higher JOLs to study materials that have larger font sizes or are presented in louder voices (Besken & Mulligan, 2013; Rhodes & Castel, 2008; Rhodes & Castel, 2009). For example, Rhodes and Castel (2008) asked their participants to learn words that were shown in large and small font sizes. Their results showed that participants tended to believe that their memory performance would be greater with larger font sizes compared to smaller font sizes, but there was no significant difference in memory performance between the two categories, which indicated the appearance of a metamemory illusion. However, previous studies that concern the impact of perceptual information on JOLs have focused mainly on the relationship between JOLs and perceptual information about to-be-remembered items in the study phase, and very few have examined perceptual information in the JOL phase. In this study, we examined the impact of cue word perceptual information in the JOL phase on JOL accuracy.

In fact, many studies demonstrate that some factors in the JOL phase can influence memory prediction (Benjamin, 2005; Finn, 2008; Hanczakowski, Zawadzka, Pasek, & Higham, 2013; Todorov, Kornell, Larsson Sundqvist, & Jönsson, 2013). For example, Finn (2008) asked participants to make JOLs within remember and forget frames and found lower mean JOL magnitude and less overconfidence bias in the forget frame than in the remember frame. In addition, Benjamin (2005) showed that JOL time pressure changed the main sources that JOLs were based on, and Hanczakowski et al. (2013) demonstrated that the JOL scale type (probability scale or binary decision) had a significant influence on JOL accuracy. In the computer-based word-pair learning paradigm, cue words in word pairs typically appear on the computer screen when subjects make JOLs (Nelson & Dunlosky, 1991; Son & Metcalfe, 2005). Previous studies have demonstrated that cue word characteristics (such as cue familiarity) have a significant impact on metamemory accuracy (Maki, 1999; Schwartz & Metcalfe, 1992). This study used different font sizes to change perceptual information of cue words to investigate its influence on JOL accuracy.

Researchers divide JOLs into two types based on the interval between the learning and the JOL: immediate JOLs are made immediately after word-pair learning, and delayed JOLs are made after a delay (Nelson & Dunlosky, 1991). Previous studies concerning the impact of the font size of to-be-remembered materials on JOLs have been based mainly on the immediate JOL paradigm (Besken & Mulligan, 2013; Rhodes & Castel, 2008; Susser, Mulligan, & Besken, 2013) and propose that study material font size affects immediate JOLs because of perceptual fluency differences (Rhodes & Castel, 2008) or metamemory beliefs (Mueller, Dunlosky, Tauber, & Rhodes, 2014). In the JOL phase, cue word font size may also affect immediate JOLs through fluency differences or metamemory beliefs, as is suggested as an explanation of the effect of study material font size on JOLs. Thus, in this study, our first purpose was to investigate whether font size in the immediate JOL phase significantly affected JOL magnitude.

Although immediate JOLs are based mainly on heuristics in the study phase (Koriat, 1997), delayed JOLs rely more on long-term memory strength (Rhodes & Tauber, 2011a). Prior studies indicate that delayed JOLs are highly diagnostic of subsequent memory tests and that the relative accuracy of delayed JOLs is much greater than that of immediate JOLs (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991). It is unclear whether cue word font size influences delayed JOL accuracy. Thus, in addition to the relationship between cue word font size and immediate JOLs, the second purpose of this study was to examine the impact of cue word font size on delayed JOLs, based on which we also investigated the mechanisms underlying delayed JOLs.

Researchers propose two different theories on delayed JOLs. Some suggest the retrieval-only theory, which expects that delayed JOLs rely fully on the retrieval process. For example, Nelson and Dunlosky (1991) propose that delayed JOLs are made based on monitoring at the long-term memory level, whereas Spellman and Bjork (1992) believe that delayed JOLs mainly rely on retrieval attempts; JOL is higher if the retrieval attempt succeeds and lower if the attempt fails. Other researchers have suggested the dual-process theory, which proposes that although retrieval is an essential element in delayed JOLs, delayed JOLs do not depend simply on retrieval attempts; instead, delayed JOLs rely on both cue familiarity and retrieval attempts (Metcalfe & Finn, 2008b; Son & Metcalfe, 2005). Cue familiarity is the recognisability of cue words (Schwartz & Metcalfe, 1992). Cue familiarity is high if participants can successfully recognise cue words. According to dual-process theory, making delayed JOLs is a two-stage process (see Figure 1). People firstly recognise the cue word, which is based on cue familiarity. If the cue word is recognised, they proceed to make subsequent retrieval attempts. If the cue word is not recognised, then there is no retrieval attempt, and a very low JOL is given immediately (Metcalfe & Finn, 2008b).

As far as we know, there are only two studies that investigate the dual-process theory (Metcalfe & Finn, 2008b; Son & Metcalfe, 2005). Thus, more studies are needed to investigate the dual-process theory from different perspectives. In this study, we tested the validity of these two delayed-JOL theories by investigating the influence of cue word font size on delayed JOLs. If the retrieval-only theory is correct, then cue word font size should have no impact on delayed JOL accuracy because font size has no influence on recall performance (Kornell, Rhodes, Castel, & Tauber, 2011; Rhodes & Castel, 2008). However, studies show that perceptual information affects recognition memory. When words had low perceptual fluency in recognition tests, participants had the subjective experience of lower familiarity, which made them more likely to classify these words as new words (Johnston, Hawley, & Elliott, 1991; Verfaellie & Cermak, 1999). If the dual-process theory is correct, then people could often be more unfamiliar with small-cue words, giving low-JOL magnitude based on only low cue familiarity without using any retrieval processes. According to Metcalfe and Finn (2008b), relative JOL accuracy should be poorer if the JOLs are based mainly on cue familiarity without retrieval attempts. Thus, the relative delayed-JOL accuracy of small-cue pairs should be reliably lower than that of large-cue pairs.

To summarise, the present study investigated the impact of cue word font size on both immediate and delayed JOLs. We examined these effects by performing three experiments.

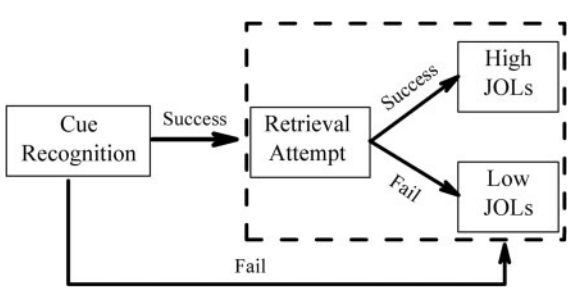

Figure 1. The processes underlying delayed judgement of learning (JOL) according to the dual-process theory. The processes in the dashed box represent the model of retrieval-only theory.

EXPERIMENT 1

In Experiment 1, the participants were required to study a series of word pairs. JOLs in the experiment were made immediately for half of the studied items and were made at a delay for the remaining half. In our computer-based experiment, the computer screen presented the cue words in each word pair during JOLs. Half of the cue words were presented in a large font and the other half were presented in a small font.

We expected that for immediate JOLs, the mean JOL magnitude of large-cue word pairs might be higher than that of small-cue pairs. For delayed JOLs, if the retrieval-only theory is correct, then the cue word font size should have no influence on JOLs. However, if the dual-process theory is correct, then the relative JOL accuracy of small-cue pairs should be reliably lower than that of large-cue pairs.

Participants

The participants were 28 students from Beijing Normal University (11 men, 17 women). Each participant was tested individually, and each participant received 20 RMB (Renminbi, the currency unit of China) as a reward after the experiment.

Materials

The materials consisted of 52 Chinese word pairs. All of the words were two-character words that were from the Chinese word database by Cai and Brysbaert (2010). The word frequency was between .006 and 46.2 per million words. Before the experiment, 200 raters used a four-point rating scale to evaluate the semantic associations of all the word pairs. On the four-point rating scale, 1 represented "very unrelated," and 4 represented "very related." The semantic associations of the word pairs were between 1.3 and 1.9.

Procedure

The experiment consisted of four phases: study (in which immediate JOLs were made), delayed JOLs, a distractor task and recall. In the study phase, the participants had to study all 52 word pairs in 33-pt font with a 4 s presentation time for each word pair. The first and last two word pairs were buffering pairs and were not included in the JOL and recall phases. From the remaining 48 word pairs, half were randomly selected for immediate JOLs. After these word pairs disappeared from the computer screen, the cue words were presented again, and participants were given 4 s for each word pair to report aloud a number from 0 to 100 that represented the probability of recalling the target words on a later memory test. Among the 24 word pairs, half of the cue words were presented in 70-pt font, and the other half were presented in 9-pt.[1] Word pairs were presented in a fixed random order in which no more than three immediate-JOL pairs were presented consecutively.

The participants then made delayed JOLs regarding the remaining 24 word pairs, which were not judged in the immediate JOL phase. The screen displayed one cue word at a time, and participants were required to give aloud a number from 0 to 100 that represented the probability of target word recall on a later memory test with a judgement time of 4 s. As with the immediate JOLs, half of the cue words were shown in 70-pt font and the other half in 9-pt. A fixed random sequence was used in which no more than three cue words of the same font size were presented consecutively, and the time interval for each word pair between learning and delayed JOL was at least 70 s.

After the delayed JOL phase, the participants engaged in a one-min arithmetic distractor task. They were then given a recall test. The screen showed one cue word at a time in 33-pt, and participants were instructed to report the target word aloud. The presentation time for each cue word was 8 s, and the next cue word was shown when the time was up, with a beep sound to remind the participants. In the recall test, the answer to a cue word was coded as "correct" only when participants correctly recalled the corresponding target word.

Data analysis

We analysed JOL magnitude, recall performance and relative accuracy in two steps. First, we conducted 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) analysis of variances (ANOVA) on mean JOL magnitude, recall performance and relative accuracy. We then conducted simple effect analyses no matter whether the interactions were significant or not. We tested simple effects regardless of the significance of the ANOVA interactions for two reasons. First, researchers have noted that simple effect analysis has no necessary relationship with interaction analysis: simple effect analysis addresses the question of whether the effect of one independent variable on the dependent variable is significant at each level of another independent variable, while interaction analysis focuses on whether the effect of one independent variable on the dependent variable reliably differ between each level of another independent variable (Jaccard, 1998; Marascuilo & Levin, 1970). In addition, simple effects can be significant with non-significant interaction (Tybout et al., 2001; Umesh, Peterson, McCann-Nelson, & Vaidyanathan, 1996). Second, according to Umesh et al. (1996), if we are interested in only the main simple effects, then the significance or non-significance of the interaction is irrelevant, and the results of simple effect tests are defensible. In our analyses of mean JOL magnitude, recall performance and gamma correlation, we are interested in whether there is significant difference between large-cue and small-cue pairs in both JOL types (i.e., simple effect). Thus, we tested simple effects in these analyses regardless of whether the interactions were significant or not.

In addition, to further investigate the relative accuracy of delayed JOLs, we analysed the data of delayed-JOL word pairs from two different perspectives. First, we compared mean delayed JOL magnitude of recalled and forgotten word pairs on the final test with different cue word font

[1] Unlike previous studies (e.g., Rhodes & Castel, 2008; Susser et al., 2013), we did not use the 48-pt and 18-pt as the large and small font size because in another study using the paradigm of Rhodes and Castel (2008), we found that mean JOL magnitude did not reliably differ when to-be-remembered Chinese words were shown in 48-pt or 18-pt ($p > .7$). However, mean JOL magnitude was higher when to-be-remembered Chinese words were shown in 70-pt than 9-pt ($p < .01$). To investigate whether participants could see clearly the 9-pt words, we conducted an experiment (35 participants) in which participants completed a lexical decision task (similar to Experiment 3) in which words and pseudowords were shown in either 70-pt or 9-pt. The results showed that the accuracy of lexical decision did not differ ($p > .7$) between 70-pt ($M = .953$) and 9-pt ($M = .956$) and that there was no significant difference ($p > .8$) between reaction time to 70-pt ($M = 888.63$ ms) and 9-pt ($M = 892.50$ ms) words. In addition, all of the participants told us that they could see the 9-pt words clearly.

sizes. The retrieval-only theory suggests that in delayed JOL phase, participants give JOLs based on only the results of their retrieval attempts, which should discriminate well between the word pairs that are remembered or forgotten regardless of cue word font size (Spellman & Bjork, 1992). Thus, cue word font size should have no impact on mean JOL magnitude for either the recalled or forgotten pairs. However, based on the dual-process theory, because retrieval attempts are more often omitted when cue words are shown in small fonts, delayed JOLs should be better at discriminating recalled and forgotten large-cue pairs than small-cue pairs.

Second, we compared recall performance of large-cue and small-cue delayed-JOL word pairs with different JOL magnitudes. If the retrieval-only theory is correct, then cue word font size should not affect recall performance regardless of what JOL magnitude is. However, if the dual-process theory is correct, then participants would more easily give a low-JOL magnitude to small-cue than large-cue pairs based on only low cue familiarity without a retrieval attempt, and some of these low-JOL pairs might be recalled in the memory test. Thus, recall performance of small-cue pairs should be reliably higher than that of large-cue pairs when JOL magnitude is low.

Results

Mean JOL magnitude. The mean JOL magnitude is shown in Table 1. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on mean JOL magnitude was conducted. Overall, mean JOL magnitude did not differ between large-cue and small-cue word pairs, $F(1, 27) = 1.37$, $p = .253$, $\eta_p^2 = .05$. Similarly, mean JOL magnitude did not differ between immediate and delayed JOL pairs, $F(1, 27) = 1.53$, $p = .227$, $\eta_p^2 = .05$. The font size did not interact with JOL type, $F(1, 27) = 2.40$, $p = .133$, $\eta_p^2 = .08$. Although the interaction did not reach the significance level, we proceeded to compare the mean JOL magnitude of the large-cue and small-cue pairs in both JOL types. For immediate JOLs, the large-cue word pairs had a significantly higher mean JOL magnitude than small-cue pairs, $F(1, 27) = 7.88$, $p = .009$, $\eta_p^2 = .23$. However, for delayed JOLs, there was no significant difference in mean JOL magnitude between the large-cue and small-cue pairs, $F(1, 27) < .01$, $p = .964$, $\eta_p^2 < .01$.

Recall performance. The recall performance is shown in Table 1. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on recall performance was conducted. Overall, recall performance did not differ between the large-cue and small-cue word pairs, $F(1, 27) = .52$, $p = .475$, $\eta_p^2 = .02$. Similarly, recall performance did not differ between the immediate and delayed JOL pairs, $F(1, 27) = .03$, $p = .859$, $\eta_p^2 < .01$. The font size did not interact with the JOL type, $F(1, 27) = 1.67$, $p = .207$, $\eta_p^2 = .06$. We also compared recall performance of large-cue and small-cue pairs in both of the JOL types. For both immediate-JOL and delayed-JOL pairs, recall performance did not differ between the large-cue and small-cue word pairs, $F_{\text{immediate}}(1, 27) = 1.75$, $p = .197$, $\eta_p^2 = .06$; $F_{\text{delayed}}(1, 27) = .42$, $p = .524$, $\eta_p^2 = .02$.

TABLE 1
Mean JOL magnitude, recall performance and gamma correlations from Experiments 1–3

	Mean JOL magnitude		*Recall performance*		*Gamma*	
	Large	*Small*	*Large*	*Small*	*Large*	*Small*
Experiment 1						
Immediate	53.06 (23.19)	49.38 (21.60)	48.51 (25.67)	43.15 (23.14)	.20 (.51)	.25 (.41)
Delayed	47.02 (17.36)	47.13 (18.82)	45.24 (22.95)	47.32 (23.68)	.94 (.12)	.78 (.25)
Experiment 2						
Immediate	53.34 (23.79)	51.11 (22.90)	46.81 (22.85)	46.81 (22.75)	.37 (.47)	.33 (.43)
Delayed	41.40 (22.44)	41.63 (20.37)	39.84 (25.28)	41.25 (24.44)	.93 (.12)	.79 (.33)
Experiment 3						
Immediate	62.70 (22.43)	60.74 (23.10)	43.67 (23.97)	47.00 (20.82)	.18 (.44)	.12 (.44)
Delayed	40.64 (21.97)	41.13 (20.52)	34.00 (22.56)	37.00 (20.14)	.90 (.16)	.88 (.17)

Standard deviations are reported in parentheses. Recall performance refers to the percentage of words recalled for Experiments 1–3.

Relative accuracy. We used gamma correlation to examine the relative accuracy of the JOLs (Nelson, 1984). Gamma is a non-parametric index of association that ranges from −1.0 to 1.0 and quantifies the association between JOLs and memory performance. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on gamma correlation was conducted. Overall, gamma correlation did not differ between large-cue and small-cue word pairs, $F(1, 24) = 1.22$, $p = .281$, $\eta_p^2 = .05$. However, gamma correlation was reliably higher for delayed-JOL pairs compared with immediate-JOL pairs, $F(1, 24) = 61.35$, $p < .001$, $\eta_p^2 = .72$. The font size did not interact with JOL type, $F(1, 24) = 1.45$, $p = .240$, $\eta_p^2 = .06$. We also analysed gamma correlation of large-cue and small-cue pairs separately in both JOL types. For immediate JOLs, the gamma correlations of the large-cue word pairs were marginally higher than 0, $t(24) = 1.92$, $p = .067$, and those of the small-cue pairs were significantly higher than 0, $t(27) = 3.21$, $p = .003$. In addition, there was no significant difference between the gamma correlations for the immediate JOLs of the large- and small-cue pairs, $F(1, 24) = .07$, $p = .794$, $\eta_p^2 < .01$. For the delayed JOLs, the gamma correlations for both word types were significantly higher than 0, $t_{\text{large}}(27) = 42.47$, $p < .001$; $t_{\text{small}}(26) = 16.31$, $p < .001$. Moreover, relative JOL accuracy with the large-cue pairs was significantly higher than with the small-cue pairs, $F(1, 26) = 8.75$, $p = .007$, $\eta_p^2 = .25$ (see Table 1).[2]

As stated in the data analysis section, to further investigate the relative accuracy of delayed JOLs, we analysed the data of the delayed-JOL word pairs from two different perspectives. First, we conducted a 2 (font size: large vs. small) × 2 (recall status: recalled vs. forgotten) ANOVA on mean JOL magnitude of delayed JOLs. Overall, mean JOL magnitude did not differ between the large-cue and small-cue pairs, $F(1, 26) = .26$, $p = .616$, $\eta_p^2 = .01$. However, mean JOL magnitude was reliably higher for the recalled pairs compared with the forgotten pairs, $F(1, 26) = 216.39$, $p < .001$, $\eta_p^2 = .89$. More importantly, these main effects were qualified by a significant font size × recall status interaction, $F(1, 26) = 14.65$, $p = .001$, $\eta_p^2 = .36$ (see Figure 2). For both the large-cue and small-cue word pairs, the mean JOL magnitude of the recalled pairs was significantly higher than were those of the forgotten pairs, $F_{\text{large}}(1, 26) = 213.29$, $p < .001$, $\eta_p^2 = .89$; $F_{\text{small}}(1, 26) = 127.41$, $p < .001$, $\eta_p^2 = .83$. However, in Figure 2, we can see that the differences between the mean JOL magnitude of the recalled and forgotten pairs were larger for the large-cue compared with the small-cue pairs.

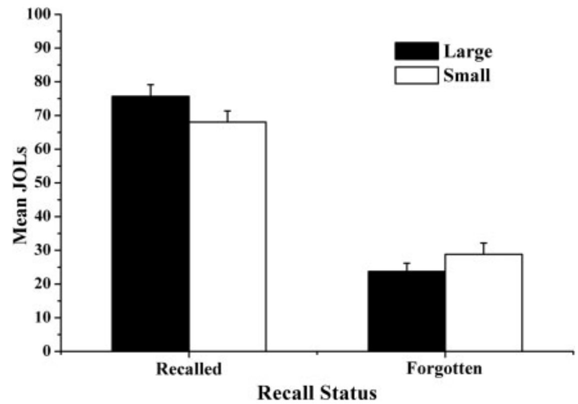

Figure 2. Mean JOL magnitude for delayed-JOL word pairs as a function of cue word font size and recall status in Experiment 1. Error bars represent standard errors.

In addition, we separated the delayed-JOL pairs into two groups according to their JOL magnitude, which was 0–50 for the first group and 50–100 for the second.[3] Then a 2 (font size: large vs. small) × 2 (JOL magnitude: high vs. low) ANOVA on recall performance was conducted. Overall, recall performance did not differ between large-cue and small-cue pairs, $F(1, 26) = 1.39$, $p = .249$, $\eta_p^2 = .05$. However, recall performance was reliably higher for high-JOL pairs compared with low-JOL pairs, $F(1, 26) = 309.31$, $p < .001$, $\eta_p^2 = .92$. More importantly, these main effects were qualified by a significant Font size × JOL magnitude interaction, $F(1, 26) = 7.63$, $p = .010$, $\eta_p^2 = .23$ (see Figure 3). For the high-JOL pairs, recall performance did not differ between the large-cue and

[2]Some of participants did not report the target word for either immediate- or delayed-JOL pairs. These participants were excluded from the gamma correlation analyses, which are reflected in the variations in the degrees of freedom that are reported for the statistical tests in each of the experiments. The participants who recalled no delayed-JOL targets were also excluded from additional ANOVAs on mean JOL magnitude and recall performance of delayed-JOL pairs.

[3]Participants in our study made JOLs by using a percentage scale containing 101 levels (0–100), with the middle value of 50, 50 levels lower than 50 (0–49), and 50 levels higher than 50 (51–100). Placing word pairs with a JOL magnitude of exactly 50 in the high-JOL or low-JOL group would be inappropriate. Thus, if the JOL magnitude of a word pair was exactly 50, then it was regarded as two half-items in both the high-JOL and low-JOL groups.

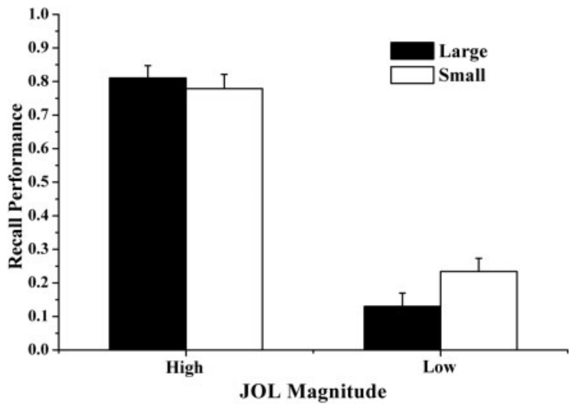

Figure 3. Recall performance for delayed-JOL word pairs as a function of cue word font size and JOL magnitude in Experiment 1. Error bars represent standard errors.

small-cue pairs, $F(1, 26) = .62$, $p = .438$, $\eta_p^2 = .02$. However, for the low-JOL pairs, recall performance was reliably higher for the small-cue pairs compared with the large-cue pairs, $F(1, 26) = 7.14$, $p = .013$, $\eta_p^2 = .22$.

Discussion

The results from Experiment 1 showed that the mean JOL magnitude of immediate JOLs was significantly higher for large-cue word pairs compared with small-cue pairs. However, cue word font size did not lead to significant difference in recall performance. For delayed-JOL word pairs, cue word font size had no impact on mean JOL magnitude. However, the gamma correlations of the delayed JOLs were significantly higher for large-cue pairs compared with small-cue pairs. The analysis of the delayed-JOL pairs also revealed a greater mean JOL discrepancy between the recalled and forgotten large-cue pairs than that found with the small-cue pairs, and it showed that recall performance was higher for small-cue pairs than large-cue pairs when JOL magnitude was low.

One possible explanation for these results is that the time given to make JOLs may have influenced JOL accuracy in Experiment 1. Previous studies have indicated that when people make judgements under time pressure, some information will be underestimated or even ignored, which harms the quality of the judgements (Maule & Edland, 1997). Thus, the information about to-be-remembered items (such as item difficulty or retrieval fluency; see Koriat, 1997) could be underestimated when immediate JOL time is limited, and immediate JOLs could be mainly based on cue word font size, which could lead to a metamemory illusion. If JOL time had been increased, then people might have considered more of the study material information, which could have reduced or even eliminated the metacognitive illusion in Experiment 1. In addition, studies have demonstrated that processing words in small fonts is more difficult than processing them in large fonts (Rhodes & Castel, 2008). Thus, when JOL time is short, retrieval attempt time should be more limited for small-cue than for large-cue pairs after participants successfully perceive the cue words. More of the remembered small-cue pairs could not be successfully retrieved in the delayed JOLs, which could have led to reduced relative accuracy. If the delayed JOL time had been increased, the difference in relative accuracy might have disappeared.

EXPERIMENT 2

In Experiment 2, the immediate and delayed JOL times were increased from 4 s to 8 s. For immediate JOLs, if people consider more study material information with increased JOL time, then the metamemory illusion of immediate-JOL word pairs should reduce or disappear. For delayed JOLs, we expected that there might be no difference in relative accuracy between large-cue and small-cue pairs when participants had a sufficient retrieval attempt time.

Participants

The participants were 34 students from Beijing Normal University (11 men, 23 women). Each participant was tested individually, and each was given 20 RMB as a reward after the experiment.

Materials

The materials were the same as those that were used in Experiment 1.

Procedure

The procedure was the same as that in Experiment 1 except that Experiment 2 increased the immediate and delayed JOL times from 4 s to 8 s.

Data analysis

Similar to Experiment 1, in Experiment 2 we conducted 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVAs on mean JOL magnitude, recall performance and relative accuracy and simple effect tests regardless of whether the interaction in these two-way ANOVAs was significant or not. In addition, as in Experiment 1, to further investigate the relative accuracy of delayed JOLs, we analysed mean JOL magnitude of delayed JOLs of the recalled and forgotten word pairs on the final test with different cue word font sizes, and we compared the recall performance of the large-cue and small-cue word pairs that had been given high or low delayed JOLs.

Results

Mean JOL magnitude. The mean JOL magnitude is shown in Table 1. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on mean JOL magnitude was conducted. Overall, mean JOL magnitude did not differ between the large-cue and small-cue word pairs, $F(1, 33) = .83$, $p = .369$, $\eta_p^2 = .03$. However, mean JOL magnitude was reliably higher for immediate-JOL pairs compared with delayed-JOL pairs, $F(1, 33) = 7.18$, $p = .011$, $\eta_p^2 = .18$. The font size did not interact with JOL type, $F(1, 33) = .94$, $p = .338$, $\eta_p^2 = .03$. We also compared mean JOL magnitude of the large-cue and small-cue pairs in both of the JOL types. The mean JOL magnitude of immediate JOLs had no significant difference between large-cue and small-cue pairs, although the mean JOL magnitude for the large-cue pairs was higher, $F(1, 33) = 2.77$, $p = .106$, $\eta_p^2 = .08$. Similarly, for the delayed-JOL word pairs, there was no significant difference in mean JOL magnitude between the large-cue and small-cue pairs, $F(1, 33) = .01$, $p = .908$, $\eta_p^2 < .01$.

Recall performance. The recall performance is shown in Table 1. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on recall performance was conducted. Overall, recall performance did not differ between the large-cue and small-cue word pairs, $F(1, 33) = .14$, $p = .713$, $\eta_p^2 < .01$. However, recall performance was reliably higher for immediate-JOL pairs compared with delayed-JOL pairs, $F(1, 33) = 6.79$, $p = .014$, $\eta_p^2 = .17$. The font size did not interact with JOL type, $F(1, 33) = .09$, $p = .761$, $\eta_p^2 < .01$. We also compared recall performance of the large-cue and small-cue pairs in both of the JOL types. For both immediate-JOL and delayed-JOL pairs, recall performance did not differ between the large-cue and small-cue word pairs, $F_{\text{immediate}}(1, 33) < .01$, $p > .999$, $\eta_p^2 < .01$; $F_{\text{delayed}}(1, 33) = .24$, $p = .630$, $\eta_p^2 = .01$.

Relative accuracy. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on gamma correlation was conducted. Overall, gamma correlation did not differ between large-cue and small-cue word pairs, $F(1, 28) = 2.04$, $p = .165$, $\eta_p^2 = .07$. However, gamma correlation was reliably higher for delayed-JOL pairs compared with immediate-JOL pairs, $F(1, 28) = 63.27$, $p < .001$, $\eta_p^2 = .69$. The font size did not interact with JOL type, $F(1, 28) = .60$, $p = .444$, $\eta_p^2 = .02$. We also analysed gamma correlation of the large-cue and small-cue pairs separately for both of the JOL types. For immediate JOLs, the gamma correlations for word pairs in both of the cue word font sizes were significantly higher than 0, $t_{\text{large}}(31) = 4.49$, $p < .001$; $t_{\text{small}}(32) = 4.38$, $p < .001$. In addition, as in Experiment 1, there was no significant difference in the gamma correlations between large-cue and small-cue word pairs, $F(1, 31) < .01$, $p = .959$, $\eta_p^2 < .01$. For delayed JOLs, the gamma correlations for word pairs in both of the cue word font sizes were also significantly higher than 0, $t_{\text{large}}(30) = 41.43$, $p < .001$; $t_{\text{small}}(32) = 13.66$, $p < .001$. Moreover, there were significant gamma correlation differences between the two word types with delayed JOLs, and the large-cue pairs had significantly higher gamma correlations than did the small-cue pairs, $F(1, 30) = 4.74$, $p = .037$, $\eta_p^2 = .14$ (see Table 1).

As in Experiment 1, we conducted a 2 (font size: large vs. small) × 2 (Recall status: recalled vs. forgotten) ANOVA on mean JOL magnitude of delayed JOLs. Overall, mean JOL magnitude did not differ between the large-cue and small-cue pairs, $F(1, 30) = .43$, $p = .515$, $\eta_p^2 = .01$. However, mean JOL magnitude was reliably higher for recalled pairs compared with forgotten pairs, $F(1, 30) = 238.93$, $p < .001$, $\eta_p^2 = .89$. Unlike in Experiment 1, the recall status × font size interaction was not significant, although there was a trend, $F(1, 30) = 2.38$, $p = .133$, $\eta_p^2 = .07$ (see Figure 4). We also compared the mean JOL magnitude of recalled and forgotten pairs in both cue word font size. For both the large-cue

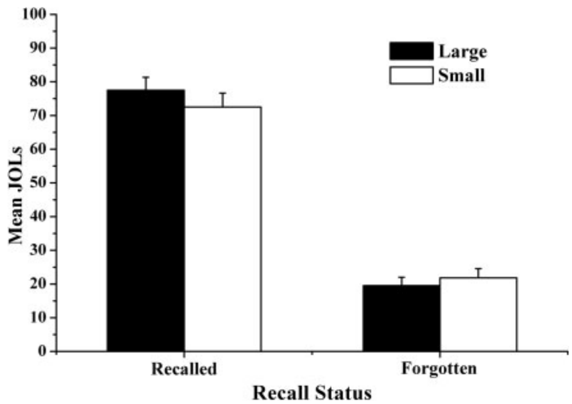

Figure 4. Mean JOL magnitude for delayed-JOL word pairs as a function of cue word font size and recall status in Experiment 2. Error bars represent standard errors.

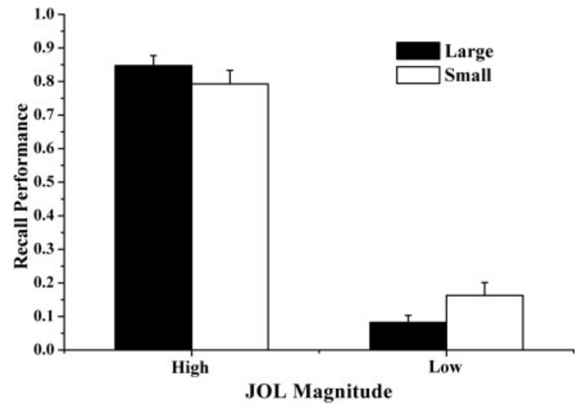

Figure 5. Recall performance for delayed-JOL word pairs as a function of cue word font size and JOL magnitude in Experiment 2. Error bars represent standard errors.

and small-cue word pairs, the mean JOL magnitude of the recalled pairs was significantly higher than were those of the forgotten pairs, $F_{\text{large}}(1, 30) = 237.45$, $p < .001$, $\eta_p^2 = .89$; $F_{\text{small}}(1, 30) = 117.69$, $p < .001$, $\eta_p^2 = .80$.

We also separated the delayed-JOL pairs into two groups according to their JOL magnitude, which was 0–50 for the first group and 50–100 for the second. Then a 2 (font size: large vs. small) × 2 (JOL magnitude: high vs. low) ANOVA on recall performance was conducted. Overall, recall performance did not differ between the large-cue and small-cue pairs, $F(1, 30) = .16$, $p = .696$, $\eta_p^2 = .01$. However, recall performance was reliably higher for high-JOL pairs compared with low-JOL pairs, $F(1, 30) = 488.13$, $p < .001$, $\eta_p^2 = .94$. More importantly, these main effects were qualified by a significant font size × JOL magnitude interaction, $F(1, 30) = 4.90$, $p = .035$, $\eta_p^2 = .14$ (see Figure 5). For the high-JOL pairs, recall performance did not differ between large-cue and small-cue pairs, $F(1, 30) = 1.55$, $p = .222$, $\eta_p^2 = .05$. However, for the low-JOL pairs, recall performance was marginally higher for small-cue pairs compared with large-cue pairs, $F(1, 30) = 3.00$, $p = .094$, $\eta_p^2 = .09$.

Discussion

In Experiment 2, the mean JOL magnitude of immediate JOLs did not significantly differ between the large-cue and small-cue word pairs, which supported our hypothesis that the metamemory illusion of immediate-JOL pairs was reduced with increased JOL time. For delayed JOLs, although the ANOVA results for the mean JOL magnitude of recalled and forgotten pairs with different cue word font sizes revealed no recall status × font size interaction, significant differences in gamma correlations between large-cue and small-cue pairs were still observed. The small-cue pairs had reliably lower relative JOL accuracy than did the large-cue pairs even when delayed JOL time was increased. In addition, recall performance was marginally higher for small-cue than large-cue pairs when JOL magnitude was low. One possible explanation is that retrieval attempts with the delayed JOLs of small-cue pairs were more often avoided because of the small fonts, which participants interpreted as low cue familiarity and which reduced the relative accuracy of the small-cue pairs, as proposed by the dual-process theory (Metcalfe & Finn, 2008b). To directly examine the relationship between cue familiarity and delayed JOLs, we increased the cue familiarity of the small-font cue words in Experiment 3.

EXPERIMENT 3

The procedure in Experiment 3 was similar to that in Experiment 2. The only difference was that Experiment 3 added a lexical decision task between the study and delayed JOL phases. During that task, a word or pseudoword appeared on the screen, and participants determined whether the stimulus on the screen was a word or not by pressing a button. In fact, the words in the lexical decision task included all of the small-font cue words in the subsequent delayed-JOL phase but not those in the large font. The actual purpose of the lexical decision task in Experiment

3 was to increase the participants' familiarity with the small-font cue words. This method was similar to the method in the experiments that were conducted by Schwartz and Metcalfe (1992). Schwartz and Metcalfe asked participants to make pleasantness judgements on some of the to-be-remembered words before the memory task to increase cue familiarity. We expected that in Experiment 3, cue word font size would not have a significant influence on the relative accuracy of the delayed JOLs.

As in Experiment 2, Experiment 3 also investigated the influence of cue word font size on immediate JOLs. Our hypothesis was the same as that of Experiment 2: the cue word font size should have no significant influence on mean JOL magnitude of immediate JOLs.

Participants

The participants were 25 students from Beijing Normal University (5 men and 20 women). Each participant was tested individually, and each received 20 RMB as a reward after the experiment.

Materials

The materials were similar to those in Experiments 1 and 2 except that we added 24 two-character words and 36 pseudowords for the lexical decision task. All of the pseudowords came from normal two-character words in which the sequence of two characters was exchanged so that they became pseudowords. All 60 two-character words (including all of the original words that formed the pseudowords) were from the Chinese word database by Cai and Brysbaert (2010), with word frequencies between .03 and 35.71 per million words.

Procedure

The procedure was in essence the same as in Experiment 2. The only difference was the addition of the lexical decision task between the study and delayed-JOL phases. The lexical decision task contained 36 words and 36 pseudowords; the 36 words consisted of 24 new words and 12 studied words, and the 12 studied words were the same as the 12 cue words shown in 9-pt font in the subsequent delayed-JOL phase. In the lexical decision task, one word (or pseudoword) was shown on the screen each time in 33-pt font. The participants were required to press a button to determine whether the screen showed a word or a pseudoword.

Data analysis

As in Experiment 1 and 2, in Experiment 3 we conducted 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVAs on mean JOL magnitude, recall performance and relative accuracy and simple effect tests regardless of whether the interaction in these two-way ANOVAs was significant or not. In addition, as in Experiment 1 and 2, to further investigate the relative accuracy of delayed JOLs, we analysed mean JOL magnitude of delayed JOLs of the recalled and forgotten word pairs on the final test with different cue word font sizes, and we compared recall performance of large-cue and small-cue word pairs that had been given high or low delayed JOLs.

Results

Lexical decision. The participants' mean accuracy rate reached .971 on the lexical decision task. The participants' response times for pseudoword judgement (M = 1043.22 ms, SE = 53.77 ms) were significantly higher than those for word judgement (M = 895.51 ms, SE = 41.18 ms), $F(1, 24) = 41.65$, $p < .001$, $\eta_p^2 = .63$. The response times did not differ as to whether the words were presented in the study phase (M = 876.23 ms, SE = 49.91 ms) or not (M = 905.15 ms, SE = 41.86 ms), $F(1, 24) = .61$, $p = .441$, $\eta_p^2 = .03$.

Mean JOL magnitude. The mean JOL results are shown in Table 1. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on mean JOL magnitude was conducted. Overall, mean JOL magnitude did not differ between large-cue and small-cue word pairs, $F(1, 24) = .22$, $p = .642$, $\eta_p^2 = .01$. However, mean JOL magnitude was reliably higher for immediate-JOL pairs compared with delayed-JOL pairs, $F(1, 24) = 16.48$, $p < .001$, $\eta_p^2 = .41$. The font size did not interact with JOL type, $F(1, 24) = .80$, $p = .380$, $\eta_p^2 = .03$. We also compared mean JOL magnitude of large-cue and small-cue pairs in both JOL types. For both immediate-JOL and delayed-JOL pairs, mean JOL magnitude did not differ between large-cue and small-cue word pairs, $F_{\text{immediate}}(1, 24) = 1.31$, $p = .263$, $\eta_p^2 = .05$; $F_{\text{delayed}}(1, 24) = .04$, $p = .839$, $\eta_p^2 < .01$.

Recall performance. The recall performance is shown in Table 1. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on recall performance was conducted. Overall, recall performance did not differ between the large-cue and small-cue word pairs, $F(1, 24) = 2.21$, $p = .150$, $\eta_p^2 = .08$. However, recall performance was reliably higher for immediate-JOL pairs compared with delayed-JOL pairs, $F(1, 24) = 16.09$, $p = .001$, $\eta_p^2 = .40$. The font size did not interact with JOL type, $F(1, 24) < .01$, $p = .948$, $\eta_p^2 < .01$. We also compared recall performance of large-cue and small-cue pairs in both of the JOL types. For both immediate-JOL and delayed-JOL pairs, recall performance did not differ between large-cue and small-cue word pairs, $F_{\text{immediate}}(1, 24) = .80$, $p = .380$, $\eta_p^2 = .03$; $F_{\text{delayed}}(1, 24) = 1.12$, $p = .302$, $\eta_p^2 = .04$.

Relative accuracy. A 2 (font size: large vs. small) × 2 (JOL type: immediate vs. delayed) ANOVA on gamma correlation was conducted. Overall, gamma correlation did not differ between large-cue and small-cue word pairs, $F(1, 20) = .87$, $p = .361$, $\eta_p^2 = .04$. However, gamma correlation was reliably higher for delayed-JOL pairs compared with immediate-JOL pairs, $F(1, 20) = 80.92$, $p < .001$, $\eta_p^2 = .80$. The font size did not interact with JOL type, $F(1, 20) = .20$, $p = .664$, $\eta_p^2 = .01$. We also analysed gamma correlation of large-cue and small-cue pairs separately in both of the JOL types. For immediate JOLs, the gamma correlations for the large-cue pairs were higher than 0 but the difference was marginal, $t(21) = 1.92$, $p = .069$; the gamma correlations for the small-cue pairs were not significantly higher than 0, $t(22) = 1.28$, $p = .214$. In addition, there was no significant difference between the gamma correlations of the large-cue and small-cue pairs, $F(1, 21) = .21$, $p = .654$, $\eta_p^2 = .01$. For delayed JOLs, the gamma correlations for both word types were significantly higher than 0, $t_{\text{large}}(23) = 27.74$, $p < .001$; $t_{\text{small}}(23) = 25.18$, $p < .001$. Moreover, there was no significant difference between the gamma correlations for the large-cue and small-cue delayed-JOL pairs, $F(1, 22) = .44$, $p = .514$, $\eta_p^2 = .02$ (see Table 1).

Similar to Experiments 1 and 2, we conducted a 2 (font size: large vs. small) × 2 (recall status: recalled vs. forgotten) ANOVA on mean JOL magnitude of delayed JOLs. Overall, mean JOL magnitude did not differ between the large-cue and small-cue pairs, $F(1, 22) = 1.63$, $p = .215$, $\eta_p^2 = .07$. However, mean JOL magnitude was reliably higher for the recalled pairs compared with the forgotten pairs, $F(1, 22) = 203.07$, $p < .001$, $\eta_p^2 = .90$. More importantly, the recall status × font size interaction was not significant, $F(1, 22) = .29$, $p = .595$, $\eta_p^2 = .01$ (see Figure 6). We also compared the mean JOL magnitude of recalled and forgotten pairs in both cue word font size. For both the large-cue and small-cue word pairs, the mean JOL magnitude of the recalled pairs was significantly higher than were those of the forgotten pairs, $F_{\text{large}}(1, 22) = 158.27$, $p < .001$, $\eta_p^2 = .88$; $F_{\text{small}}(1, 22) = 173.49$, $p < .001$, $\eta_p^2 = .89$.

We also separated the delayed-JOL pairs into two groups according to their JOL magnitude, which was 0–50 for the first group and 50–100 for the second. Then a 2 (font size: large vs. small) × 2 (JOL magnitude: high vs. low) ANOVA on recall performance was conducted. Overall, recall performance did not differ between large-cue and small-cue pairs, $F(1, 22) = .71$, $p = .407$, $\eta_p^2 = .03$. However, recall performance was reliably higher for high-JOL pairs compared with low-JOL pairs, $F(1, 22) = 560.99$, $p < .001$, $\eta_p^2 = .96$. The font size did not interact with JOL magnitude, $F(1, 22) = .84$, $p = .371$, $\eta_p^2 = .04$ (see Figure 7). For both the high-JOL and low-JOL pairs, recall performance did not differ between large-cue and small-cue pairs, $F_{\text{high}}(1, 22) = .92$, $p = .347$, $\eta_p^2 = .04$; $F_{\text{low}}(1, 22) < .01$, $p = .956$, $\eta_p^2 < .01$.

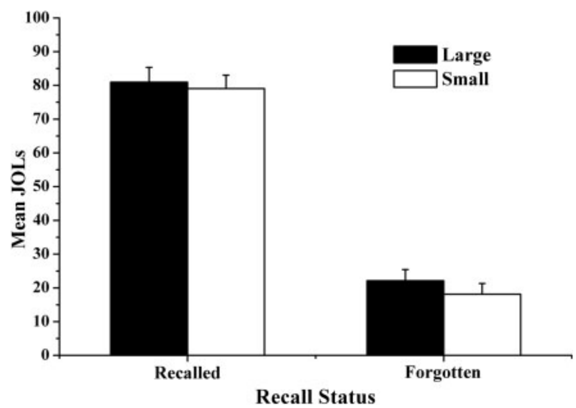

Figure 6. Mean JOL magnitude for delayed-JOL word pairs as a function of cue word font size and recall status in Experiment 3. Error bars represent standard errors.

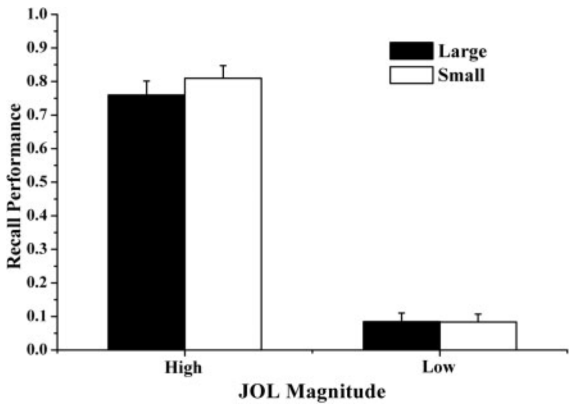

Figure 7. Recall performance for delayed-JOL word pairs as a function of cue word font size and JOL magnitude in Experiment 3. Error bars represent standard errors.

Discussion

In Experiment 3, the relative accuracy of the delayed JOLs was not affected by cue word font size. This result indicated that when the familiarity of small-cue words was increased, the relative JOL accuracy did not differ the between large-cue and small-cue pairs.

GENERAL DISCUSSION

The current study examined the influence of cue word perceptual information in the JOL phase on immediate and delayed JOLs through font size changes. The results from three experiments demonstrated that cue word font size affected immediate and delayed JOLs in different ways.

For immediate JOLs, in Experiment 1, mean JOL magnitude was significantly higher for large-cue word pairs compared with small-cue pairs but there was no significant difference in recall performance, which led to a metamemory illusion. These results showed that the effect of varying font size during the JOL phase is similar to the effect of varying font size during the study phase, both of which had significant influence on immediate JOLs (Rhodes & Castel, 2008; Susser et al., 2013). Rhodes and Castel (2008) demonstrated that people interpret the font size of to-be-remembered items as ease of processing words and give higher immediate JOLs to words that are shown in larger sizes. In Experiment 1, although the font sizes changed in the JOL phase rather than in the to-be-remembered materials, the participants might have still interpreted the font size differences as encoding fluency and have given higher JOLs to large-cue pairs. Another possible explanation for the mean JOL difference is that participants might have held the metamemory belief that words that are printed in larger font sizes are more easily remembered (Mueller, Dunlosky, Tauber, & Rhodes, 2014). Thus, the participants could have given higher JOLs to words that are shown in the larger font size even when the font size difference was unrelated to the study materials.

Unlike in Experiment 1, when JOL time was increased in Experiments 2 and 3, the difference between mean JOL magnitude of the large-cue and small-cue pairs was reduced. These results indicate that the influence of cue word font size on immediate JOLs is related to time pressure. Although in this study we did not record how long participants took to make JOLs, our findings in another unpublished study (Hu & Luo, 2014) show that participants took significantly more time to make judgements ($p < .001$) when they were asked to make immediate JOLs in 8 s (M = 2884.96 ms) than 4 s (M = 1786.84 ms). Thus, when asked to make JOLs in a limited amount of time, participants could feel stressed and take less time to make judgements. Previous studies show that people can underestimate certain important information when making judgements under time pressure, which may reduce the quality of their judgements (Maule & Edland, 1997; Svenson, Edland, & Slovic, 1990). For example, Svenson et al. (1990) asked participants to judge which of two students would be more qualified to follow a university programme based on their grades in different high school courses. Their results showed that when participants made judgements under time pressure, they weighed the course with the highest grade more heavily and underestimated the other course grades. Research on metamemory indicates that immediate JOLs are affected by many heuristics (Koriat, 1997). Making JOLs under time pressure can lead to a reduced range of cue utilisation, which could cause metacognitive illusion. In our study, the participants might have underestimated the study material information with the limited JOL time in Experiment 1. When the JOL time was increased in Experiment 2 and 3, the participants gave more consideration to the study material information, which reduced the metacognitive illusion. Further study is needed to investigate in more detail why participants underestimate the information from to-be-remembered items when JOL time is limited.

For delayed JOLs, Experiment 1 showed that relative JOL accuracy was significantly higher for large-cue pairs compared with small-cue pairs. In Experiment 2, we increased JOL time to test whether a limited amount of time for delayed JOLs led to a difference in relative accuracy. Hu and Luo (2014) find that reaction time for making delayed JOLs was reliably longer ($p < .001$) when JOL time was 8 s (M = 3844.91 ms) compared with 4 s (M = 2220.10 ms), which shows that participants took more time for retrieval attempts when the time for delayed JOLs was longer. However, Experiment 2 revealed the same font-size effect on gamma correlation when participants had sufficient time for retrieval attempts. These results support the dual-process theory, which suggests that when people make delayed JOLs, they firstly recognise the cue words and then retrieve the target words (Metcalfe & Finn, 2008b; Son & Metcalfe, 2005). Small-cue pairs are more easily misrecognised as new words (Johnston et al., 1991; Verfaellie & Cermak, 1999) and given low JOLs based on only low cue familiarity, which can reduce relative accuracy. To further verify the dual-process theory, Experiment 3 increased the familiarity of the small-cue words, and there was no difference in gamma correlations between the large-cue and small-cue pairs. The experimental design of Experiment 3 is unbalanced, and in this design we increased only the familiarity of small-cue words in the delayed JOL phase. However, the results were consistent with our prediction that relative accuracy did not differ between large-cue and small-cue delayed-JOL pairs when cue familiarity of small-cue words was increased. Our results show that cue recognition based on cue familiarity and retrieval attempts based on target retrievability are both important processes in delayed JOLs. Previous research demonstrates that cue familiarity and target retrievability interact in a number of metamemory judgements such as feeling of knowing (FOK) and delayed JOLs (Koriat & Levy-Sadot, 2001; Metcalfe & Finn, 2008b). Our results are consistent with the results of these studies and suggest that cue familiarity plays an important role in metamemory monitoring. In addition, according to the dual-process theory, when people make low delayed JOLs based only on cue familiarity, response times are significantly shorter than they are for low JOLs that are based on retrieval attempts (Metcalfe & Finn, 2008b; Son & Metcalfe, 2005). Future research could investigate the differences in delayed JOL response times between large-cue and small-cue word pairs.

In addition, our findings show that recall performance of immediate-JOL pairs was significantly higher than that of delayed-JOL pairs in Experiments 2 and 3, but in Experiment 1 recall performance did not differ between the two types of word pairs. Previous studies have reported different results regarding the influence of JOL type on recall performance. Some studies have found no difference in recall performance between immediate-JOL and delayed-JOL word pairs (e.g., Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991), whereas other studies have shown that the JOL type had a significant influence on recall performance. For example, Van Overschelde and Nelson (2006) found that the recall performance of immediate-JOL pairs is lower than that of word pairs with short-delayed JOLs (5 intervening trials between the study and JOL phase); however, the recall difference disappeared when there were 50 trials between the study and delayed JOL phase. In addition, Finn and Metcalfe (2007) showed that immediate-JOL word pairs led to higher recall performance than did delayed-JOL pairs in their first experiment, in which participants made immediate and delayed JOLs in two different word lists. However, in their second experiment, the recall difference was non-significant when the immediate-JOL and delayed-JOL word pairs were mixed within a single list. The findings of these studies suggest that the influence of JOL type on recall performance may vary depending on the experimental paradigm. One important difference in the experimental paradigm between our study and previous studies is that the JOLs were self-paced in previous studies, in which the JOL phase ended immediately after participants made judgements. However, participants in our study made JOLs during a fixed time period and might still have time to improve their memory strength in the JOL phase after they made JOLs. Given this difference, there are two potential explanations for the influence of JOL type on recall performance in Experiments 2 and 3. The first explanation concerns the different strategies that participants could use to improve their memory performance in the immediate- and delayed-JOL phases. Participants could rehearse the studied cue-target word pairs in the immediate-JOL phase because they had recently seen the complete word pairs, whereas they could not rehearse some of the studied word pairs in the delayed-JOL phase, which might have led to the difference in recall performance. Although the increased delayed JOL time in Experiments 2 and 3 encouraged participants to make delayed JOLs based on their retrieval attempts (Metcalfe & Finn, 2008b), the retrieval attempts did not lead to better recall performance in the memory test several minutes later relative to the word rehearsal (Roediger & Karpicke, 2006). Moreover, in the delayed JOL phase, participants could match the cue words with incorrect target words in their retrieval attempt, which may have led to a false memory in the memory test and may have reduced their recall performance (cf. Rhodes & Tauber, 2011b). However, we found that the proportion of mismatch errors did not differ between the immediate-JOL and delayed-JOL pairs in any of the experiments ($p > .1$), indicating that the recall difference between immediate-JOL and delayed-JOL pairs might not result from mismatch errors. More importantly, the difference in learning strategies between the immediate- and delayed-JOL phases was present in all three experiments, but the influence of JOL type on recall performance was significant only in Experiments 2 and 3. Thus, the difference in learning strategies may not account for the recall difference between immediate-JOL and delayed-JOL pairs.

The second explanation for the influence of JOL type on recall performance is that according to Hu and Luo (2014), the time that participants could use to improve their memory performance after making JOLs was longer in the immediate-JOL phase than in the delayed-JOL phase, as the participants spent more time making delayed JOLs than immediate JOLs. Hu and Luo (2014) also found that the time difference between the delayed and immediate JOLs was marginally greater when the participants were asked to make JOLs in 8 s compared with in 4 s ($p = .06$); thus, when the JOL time was 8 s, participants had much less time to improve their memory performance in the delayed-JOL phase than in the immediate-JOL phase after they made JOLs. This time difference may explain why a difference in recall performance was observed in Experiments 2 and 3 but not in Experiment 1. Further research is needed to test the effect of JOL type on recall performance. Moreover, we found that the mean JOL magnitude was higher for immediate-JOL pairs than for delayed-JOL pairs in Experiments 2 and 3, which suggests that participants were aware of the difference in memory performance.

Many studies demonstrate that perceptual information regarding to-be-remembered items affects memory prediction (Rhodes & Castel, 2008; Rhodes & Castel, 2009). However, few have investigated whether perceptual information in the JOL phase has a significant influence on JOLs. The results from our experiments show that font size in the JOL phase affects both immediate and delayed JOLs. More studies are needed to examine the effects of other types of cue word perceptual information (such as sound volume; see Rhodes & Castel, 2009) or the impact of cue word perceptual information on other types of JOLs, such as global JOLs (cf. Barnes & Dougherty, 2007). Whether cue word perceptual information has a significant impact on subsequent learning could also be investigated. Moreover, our study indicates that relying on cue word font size does not always lead to accurate memory prediction. Thus, future research should also investigate how to reduce or eliminate the negative influence of cue word font size on the JOL accuracy, using theory- and experience-based procedures (Koriat & Bjork, 2006).

DISCLOSURE STATEMENT

No potential conflict of interest was reported by the authors.

FUNDING

This study was supported by the Natural Science Foundation of China [grant number 31100817], [grant number 31000506], Beijing Higher Education Young Elite Teacher Project and the project for young people who are devoted to the researches of humanities and social sciences [grant number 10YJCXLX032] of the Ministry of Education of China.

REFERENCES

Barnes, K. A., & Dougherty, M. R. (2007). The effect of divided attention on global judgment of learning accuracy. *The American Journal of Psychology*, *120*, 347–359.

Benjamin, A. S. (2005). Response speeding mediates the contributions of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin & Review*, *12*, 874–879. doi:10.3758/BF03196779

Besken, M., & Mulligan, N. W. (2013). Easily perceived, easily remembered? Perceptual interference produces a double dissociation between metamemory and memory performance. *Memory & Cognition*, *41*, 897–903. doi:10.3758/s13421-013-0307-8

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. doi:10.1146/annurev-psych-113011-143823

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*, *5*(6), e10729. doi:10.1371/journal.pone.0010729

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition*, *20*, 374–380. doi:10.3758/BF03210921

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, *36*, 813–821. doi:10.3758/MC.36.4.813

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*, *33*, 238–244. doi:10.1037/0278-7393.33.1.238

Hanczakowski, M., Zawadzka, K., Pasek, T., & Higham, P. A. (2013). Calibration of metacognitive judgments: Insights from the underconfidence-with-practice effect. *Journal of Memory and Language*, *69*, 429–444. doi:10.1016/j.jml.2013.05.003

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*(1), 22–34. doi:10.1037/0278-7393.29.1.22

Hu, X., & Luo, L. (2014). *Time pressure leads to inaccurate judgment of learning*. Unpublished manuscript. Beijing Normal University.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. Thousand Oaks, CA: Sage.

Johnston, W. A., Hawley, K. J., & Elliott, J. M. (1991). Contribution of perceptual fluency to recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 210–223. doi:10.1037/0278-7393.17.2.210

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. doi:10.1037/0096-3445.126.4.349

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. doi:10.3758/BF03193244

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(1), 34–53. doi:10.1037/0278-7393.27.1.34

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. doi:10.1016/j.jml.2005.01.001

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease-of-processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science*, *22*, 787–794. doi:10.1177/0956797611407929

Maki, R. H. (1999). The roles of competition, target accessibility, and cue familiarity in metamemory for word pairs. *Journal of Experimental Psychology-Learning, Memory, and Cognition*, *25*, 1011–1023. doi:10.1037/0278-7393.25.4.1011

Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. *American Educational Research Journal*, *7*, 397–421. doi:10.3102/00028312007003397

Maule, A. J., & Edland, A. C. (1997). The effects of time pressure on human judgment and decision making. In R. Ranyard, W. R. Crozier, & O. Svenson (Eds.), *Decision making: Cognitive models and explanations* (pp. 189–204). London: Routledge.

Metcalfe, J., & Finn, B. (2008a). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*, 174–179. doi:10.3758/PBR.15.1.174

Metcalfe, J., & Finn, B. (2008b). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1084–1097. doi:10.1037/a0012580

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*, 463–477. doi:10.1016/j.jml.2004.12.001

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, *70*, 1–12. doi:10.1016/j.jml.2013.09.007

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*(1), 109–133. doi:10.1037/0033-2909.95.1.109

Nelson, T. O. (1993). Judgments of learning and the allocation of study time. *Journal of Experimental Psychology: General*, *122*, 269–273. doi:10.1037/0096-3445.122.2.269

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "Delayed-JOL effect". *Psychological Science*, *2*, 267–270. doi:10.1111/j.1467-9280.1991.tb00147.x

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In H. B. Gordon (Ed.), *Psychology of learning and motivation* (Vol. 26, pp. 125–173). New York, NY: Academic Press.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625. doi:10.1037/a0013684

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, *16*, 550–554. doi:10.3758/PBR.16.3.550

Rhodes, M. G., & Tauber, S. K. (2011a). The influence of delaying judgments of learning on metacognitive accuracy: A meta-analytic review. *Psychological Bulletin*, *137*(1), 131–148. doi:10.1037/a0021705

Rhodes, M. G., & Tauber, S. K. (2011b). Monitoring memory errors: The influence of the veracity of retrieved information on the accuracy of judgements of learning. *Memory*, *19*, 853–870. doi:10.1080/09658211.2011.613841

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1074–1083. doi:10.1037/0278-7393.18.5.1074

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 204–221. doi:10.1037/0278-7393.26.1.204

Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage process. *Memory & Cognition*, *33*, 1116–1129. doi:10.3758/BF03193217

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science*, *3*, 315–316. doi:10.1111/j.1467-9280.1992.tb00680.x

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition*, *41*, 1000–1011. doi:10.3758/s13421-013-0323-8

Svenson, O., Edland, A., & Slovic, P. (1990). Choices and judgments of incompletely described decision alternatives under time pressure. *Acta Psychologica*, *75*, 153–169. doi:10.1016/0001-6918(90)90084-S

Todorov, I., Kornell, N., Larsson Sundqvist, M., & Jönsson, F. U. (2013). Phrasing questions in terms of current (not future) knowledge increases preferences for cue-only judgments of learning. *Archives of Scientific Psychology*, *1*(1), 7–13. doi:10.1037/arc0000002

Tullis, J. G., & Benjamin, A. S. (2012). Consequences of restudy choices in younger and older learners. *Psychonomic Bulletin & Review*, *19*, 743–749. doi:10.3758/s13423-012-0266-2

Tybout, A., Sternthal, B., Keppel, G., Verducci, J., Meyers-Levy, J., Barnes, J., … Maxwell, S. (2001). Analysis of variance. *Journal of Consumer Psychology*, *10*(1–2), 5–35. doi:10.1207/S15327663JCP1001&2_03

Umesh, U. N., Peterson, R. A., McCann-Nelson, M., & Vaidyanathan, R. (1996). Type IV error in marketing research: The investigation of ANOVA interactions. *Journal of the Academy of Marketing Science*, *24*(1), 17–26. doi:10.1007/BF02893934

Van Overschelde, J. P., & Nelson, T. O. (2006). Delayed judgments of learning cause both a decrease in absolute accuracy (calibration) and an increase in relative accuracy (resolution). *Memory & Cognition*, *34*, 1527–1538. doi:10.3758/BF03195916

Verfaellie, M., & Cermak, L. S. (1999). Perceptual fluency as a cue for recognition judgments in amnesia. *Neuropsychology*, *13*, 198–205. doi:10.1037/0894-4105.13.2.198