Examining the effect of list composition on monitoring and control processes in metamemory

Skylar J. Laursen¹ · Chris M. Fiacconi¹

Accepted: 30 September 2020 / Published online: 14 October 2020
© The Psychonomic Society, Inc. 2020

Abstract

According to the widely accepted cue-utilization view, judgments of learning (JOLs) are thought to be comparative in nature, such that they are sensitive to the relative differences between stimuli. Here, we report the results of three experiments that both support and extend this tenet of the cue-utilization view by examining the impact of relative differences on metacognitive control strategies, including study-time allocation and re-study selection. By presenting word pairs of medium-difficulty intermixed with either easy or difficult word pairs we manipulated list composition to assess the impact of the relative difference between items on individuals' JOLs (Experiments 1 and 2a), study-time allocation (Experiment 1), and re-study selection (Experiments 2a and 2b). First, our manipulation of list composition demonstrated that stimuli of equal difficulty are judged to be more or less memorable depending on the context in which they are presented, thereby confirming previous findings that JOLs are sensitive to the relative differences among items. Second, with regard to metacognitive control strategies, our results indicated that list composition may not impact all control strategies in the same fashion. Specifically, the relative differences between items did not appear to influence the amount of study time allocated to a given item, but did affect which items were selected for re-study. These findings have important applied implications, and may assist in the development of more effective guidance on how to best engage in self-regulated learning.

Keywords Metamemory · Metacognitive monitoring · Metacognitive control · Judgments of learning

Introduction

The ability to effectively monitor and control one's learning is an essential ingredient in academic success. While the former refers to an individual's ability to evaluate and assess ongoing learning, the latter speaks to the ability to implement and appropriately adjust learning activities to optimize eventual performance. The link between monitoring and control strategies has been of considerable interest in the meta-memory literature (Dunlosky & Connor, 1997; Fiechter, Benjamin, & Unsworth, 2016; Finn, 2008; Kornell & Metcalfe, 2006; Metcalfe, 2009; Metcalfe & Finn, 2008; Nelson, Dunlosky, Graf, & Narens, 1994; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999). This link is often studied by examining the correspondence between judgments of learning (JOLs),

which are prospective metacognitive predictions of future memory performance (Rhodes, 2016; Schwartz, 1994), and two types of control strategies, namely, study-time allocation (Dunlosky & Connor, 1997; Thiede & Dunlosky, 1999) and re-study selection (Finn, 2008; Kornell & Metcalfe, 2006; Metcalfe & Finn, 2008; Nelson et al., 1994; Thiede & Dunlosky, 1999).

This research on the interplay between monitoring and control led to the development of both the discrepancy-reduction (Dunlosky & Thiede, 1998; Thiede & Dunlosky, 1999) and the region of proximal learning models (Metcalfe, 2002; Metcalfe, 2009), both of which are designed to capture the relationship between monitoring and control. Although there are some important distinctions between these two frameworks, they both predict that when study time is unlimited, individuals will allocate more study time to items judged to be more difficult or less likely to be remembered (Dunlosky & Thiede, 1998; Kornell & Metcalfe, 2006; Metcalfe, 2002; Nelson et al., 1994; Thiede & Dunlosky, 1999). This inverse relationship has not only been found for JOLs and study-time allocation, but also for JOLs and re-study selection, such that items judged to be more difficult or less likely to be

✉ Skylar J. Laursen
slaursen@uoguelph.ca

¹ Department of Psychology, University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada

remembered are also usually more likely to be selected for re-study (Dunlosky & Connor, 1997; Finn, 2008; Metcalfe, 2002; Metcalfe & Finn, 2008; Nelson et al., 1994; Thiede & Dunlosky, 1999), at least when study time is unlimited. By and large, then, there exists strong evidence that some cues that inform monitoring (e.g., item difficulty) can also exert a parallel effect on control. However, it is currently unclear to what extent this pattern is universally applicable to all cues that inform monitoring, nor is it clear whether a given cue always affects different control strategies (e.g., study-time allocation, re-study selection) in the same fashion. The broad aim of the present study is to shed further light on the correspondence between monitoring and control by examining the influence of a specific cue, namely, list composition, on JOLs and metacognitive control functions including study-time allocation and re-study selection.

The widely accepted cue-utilization view that describes the basis of JOLs identifies the importance of relative differences between items as a key factor in guiding such judgments (Koriat, 1997). In perhaps its strongest form, this view holds that it is not an item's absolute magnitude on a given dimension that is the predominant force in guiding JOLs, but rather it is the *contrast* between different items in a list that is of greatest importance. This general relativity principle has not only been identified as critical within the meta-memory literature, but is a well-known factor that affects behaviour in other domains as well, including episodic memory (McDaniel & Bugg, 2008). Based on their review of previous literature, McDaniel and Bugg (2008) reported that many well-established memory phenomena depend on the contrast between different classes of stimuli. For example, the perceptual interference effect, the finding that perceptually masked words are better recalled than unmasked words (Hirshman & Mulligan, 1991), is typically observed only when masked and unmasked words are presented together in a mixed-list design (Mulligan, 1999).

In the meta-memory literature, this relativity principle has garnered support from studies demonstrating that JOLs can indeed exhibit sensitivity to the contrast between different classes of stimuli (Dunlosky & Matvey, 2001; Susser, Mulligan, & Besken, 2013; Zawadzka & Higham, 2016). In one of the earliest studies examining this issue, Dunlosky and Matvey (2001) investigated how the relative differences in difficulty between word pairs can impact individuals' JOLs. Participants were required to make JOL ratings for related and unrelated word pairs presented in two separate lists. They found that unrelated word pairs were assigned lower JOLs when they followed an initial list of related pairs, and similarly, that related word pairs were assigned higher JOLs when they followed an initial list of unrelated pairs (Dunlosky & Matvey, 2001, Experiments 1 and 2). To account for these results, they speculated that the difference in JOL ratings was due to anchoring effects, such that after being presented with one class of stimuli (i.e., related or unrelated word pairs), JOL ratings for the other class were influenced by the previous (Dunlosky & Matvey,

2001, Experiments 1 and 2). Zawadzka and Higham (2016) recently replicated this basic finding using a mixed-list design. In their study, participants were presented with an initial list of unrelated word pairs and were asked to provide item-by-item JOLs for all pairs. This list served as a baseline measure of JOLs for these items. The initial list was then followed by a second list for which participants again provided item-by-item JOLs. In one condition, the second list contained items from the initial list along with relatively more difficult word pairs, whereas in another condition, the second list contained items from the initial list along with relatively easier word pairs. Critically, JOLs for repeated items in the second list increased relative to baseline when mixed together with more difficult items, whereas JOLs for repeated items in the second list decreased relative to baseline when mixed together with easier items (Zawadzka & Higham, 2016). Extending the importance of relative differences between items beyond manipulations of item-difficulty, Susser et al. (2013) demonstrated that the font-size effect (Rhodes & Castel, 2008) – the finding that words presented in larger font sizes are typically judged to be more memorable – hinges on a mixed-list design. They reasoned that mixed-list designs allow individuals to more easily exploit the relative differences between items when making their JOLs, whereas such comparisons are less likely to occur in pure-list designs (Susser et al., 2013). Together, these results underscore the comparative nature of JOLs, and provide additional evidence that list composition can affect how individuals evaluate their learning.

However, it should be noted that not all previous studies have corroborated these findings. In fact, Richards and Nelson (2004) found no evidence to support the idea that JOLs are sensitive to relative differences in item difficulty. In their experiment, participants were first presented with a list of either easy or difficult Swahili–English word pairs, and were subsequently presented with a list of medium-difficulty Swahili–English word pairs. They reasoned that if the JOLs for the second class of stimuli reflect a comparison between those and the first class presented, medium-difficulty word pairs should be given higher JOL ratings when preceded by difficult word pairs than when preceded by easy word pairs (Richards & Nelson, 2004). However, this result was not observed. Richards and Nelson (2004) found no difference in JOL ratings given for medium-difficulty word pairs when they were preceded by either easy or difficult word pairs. Therefore, although the bulk of the evidence supports the notion that JOLs draw heavily on relative differences between items, such findings are not ubiquitous.

Overview of current study

From the literature reviewed thus far, it is apparent that there is evidence to support the idea that JOLs are sensitive to relative differences between classes of items as revealed by manipulations of list composition (Susser et al., 2013; Zawadzka & Higham, 2016). However, to our knowledge, it is yet unknown whether

such relative differences also regulate adjustments in control strategies such as study-time allocation and re-study selection. This question is of importance given evidence that many variables known to impact JOLs (e.g., item difficulty) are also known to exert a parallel influence on these control functions (Dunlosky & Connor, 1997; Finn, 2008; Kornell & Metcalfe, 2006; Metcalfe & Finn, 2008; Nelson et al., 1994; Thiede & Dunlosky, 1999). The purpose of the current study is twofold: first, we sought to replicate previous findings that JOLs are sensitive to relative differences in item difficulty (Dunlosky & Matvey, 2001; Zawadzka & Higham, 2016) by manipulating list composition, and second, we sought to examine whether such manipulations of list composition would have a corresponding impact on the amount of study time individuals allocate to a given item (Experiment 1) and the likelihood with which individuals would select an item for re-study (Experiments 2a and 2b).

Experiment 1

Experiment 1 represents our initial attempt at replicating the effect of list composition on JOLs, and at investigating whether list composition has a corresponding influence on metacognitive control. Specifically, we compared JOLs and self-paced study-time allocation to word pairs of medium-difficulty when presented intermixed with either easy (easy context) or difficult (difficult context) word pairs. Following the results of Zawadzka and Higham (2016), we expected that medium items presented together with easy items (easy context) would be assigned lower JOL ratings than medium items presented with difficult items (difficult context). Moreover, to the extent that individuals allocate more study time to items perceived as more difficult (Dunlosky & Thiede, 1998; Thiede & Dunlosky, 1999), particularly when study time is unlimited (Metcalfe, 2002; Metcalfe, 2009; Son & Metcalfe, 2000), we anticipated that the amount of time participants spent studying medium word pairs would be greater in the easy context as compared to the difficult context.

Method

Participants One-hundred and three undergraduate students from the University of Guelph psychology participant pool participated in this study in exchange for course credit (mean age = 18.76 years, $SD = 1.31$, 82 female). Participants were eligible for the study if they had English proficiency, were between 18 and 35 years of age, and had normal or corrected-to-normal vision. The sample size was chosen based on a power analysis using G*Power software (Faul, Erdfelder, Lang, & Buchner, 2007), using an estimated effect size of Cohen's $d = 0.30$, and the power to detect an effect of 0.90. All experimental procedures were approved by the Research Ethics Board (REB) at the University of Guelph.

Materials and design The experiment was conducted using PsychoPy 3.1.5 (Peirce et al., 2019) running on an ASUS desktop computer with a $2,560 \times 1,440$ monitor set to a refresh rate of 144 Hz. Stimuli consisted of 96 word pairs, categorized as Easy, Medium, or Difficult (see Appendix 1 Table 1). To create the stimuli a list of words was assembled from the MRC Psycholinguistic Database using the following parameters: Number of letters: 3–8; Kucera–Francis Written Frequency: 10–300; Concreteness Rating: 300–700; Imageability Rating: 300–700. From this list, word pairs were manually generated by the experimenter. Twenty-four *easy* word pairs were formed such that the two words were highly related (e.g., brother – sister). Forty-eight *medium* word pairs were formed such that the two words could be related using a connector (e.g., sky – (blue) – ocean). Finally, 24 *difficult* word pairs were formed using a random number generator and selecting the words corresponding to the numbers generated (e.g., blanket – tube). The 48 medium word pairs were further separated into two sub-lists, A and B. Using the LSAfun package for R (Günther, Dudschig, & Kaup, 2015) the latent semantic analysis (LSA) for each word pair was calculated to assess how often the two words appeared together in the English language. Independent samples *t*-tests were conducted to ensure LSA values for easy word pairs were larger than those for medium and difficult word pairs, LSA values for medium word pairs were larger than those for difficult word pairs, and LSA values for the two medium sub-lists were equivalent (see Appendix 1 Table 2 for statistical output). Furthermore, the forward strength association (FSA) and backward strength association (BSA) for each word pair was taken from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) to ensure that easy word pairs were more freely associated than medium and difficult word pairs. All stimuli were validated in a pilot study (see Appendix 2 for procedure and results).

Participants were presented with two lists consisting of 48 word pairs each across two study-test cycles. Each list was comprised of either 24 easy or difficult items, and one of the medium sub-lists. Pairing of each medium sub-list with the easy and difficult items was counterbalanced, as was the order in which the medium-easy (easy context) and medium-difficult (difficult context) lists were presented. Furthermore, the order of the word pairs was randomized within each list. This design allowed us to use within-subject comparisons between the medium word pairs across the two lists to analyze the effect of list composition.

Procedure Participants were brought into the lab three at a time and seated in front of separate testing computers. After obtaining informed consent, and a general overview of the experiment was given, participants were shown the first list of word pairs. Each trial began with a 500-ms fixation dot, followed by the word pair presented in the center of the screen. Participants were given an unlimited amount of time to study

the word pair and were required to indicate when they had finished studying by pressing the space bar. After the space bar was pressed a box appeared below the word pair, which indicated the participant was to make a JOL rating. Participants typed in a rating from 0 to 100 of how likely they would be to recall the target word when presented with the cue word at a later time, using the keyboard. Each JOL was self-paced with no experimenter-imposed time limit. Once participants made their JOL they pressed the ENTER key to begin the next trial. After all word pairs in the first list had been presented, participants completed a 5-min math (distractor) task. Following the distractor task participants completed a self-paced cued-recall test. Participants were presented with all cue words from the first list one at a time, and were instructed to recall each target word as accurately as possible.

Following completion of the cued-recall test, participants had completed one study-test cycle. Participants then completed a second study-test cycle, which was identical to the first with the exception that the word pairs and cues presented came from the second list.

Results

Three participants' data were excluded, two due to technical issues and one because they experienced a nose bleed during the experiment. All exclusions were performed blind to whether data conformed to experimental predictions. All statistical analyses were performed using *R* software (R Core Team, 2017), and a significance level of .05 was adopted. However, traditional null hypothesis significance testing (NHST) does not allow researchers to make conclusions regarding null results (i.e., a non-significant *p*-value does not equate to no difference between groups, it only means that there was no detectable difference), whereas Bayesian analyses can provide direct evidence that the null hypothesis is more probable than the alternative (Kruschke, 2013). Therefore, null NHST results are supported by Bayes factors (BF_{10}), calculated using the *BayesFactor* package for *R* (Morey & Rouder, 2018). BF_{10} values > 1 indicate support for the alternative hypothesis with benchmarks as follows: 1–3 indicates anecdotal evidence, 3–10 moderate evidence, and > 10 strong evidence. BF_{10} values < 1 indicate support for the null hypothesis with benchmarks as follows: 0.33–1 indicates anecdotal evidence, 0.1–0.33 moderate evidence, and < 0.1 strong evidence. Results for comparisons between all three word pair difficulties and list composition are presented separately.

Pair difficulty To further ensure the validity of our stimuli, mean JOL ratings, study time, and cued-recall performance for easy, medium, and difficult word pairs were calculated, see Fig. 1 (panels 1–3, respectively), and analysed using linear trend analyses. The decision to use linear trend analyses was

made because word-pair difficulty can be thought of as a continuous variable, and therefore, if difficulty influences JOLs, study time, and memory performance, we should see differences in JOLs, study time, and cued-recall performance that follow this continuous trend. JOL ratings were found to follow a significant negative linear trend, such that as word-pair difficulty increased, JOL ratings decreased, $t(99) = -22.69, p < .001, R^2_{\text{alerting}} = 0.99$.¹ Study time was found to follow a significant positive linear trend, such that as word-pair difficulty increased, so did study time, $t(99) = 5.40, p < .001, R^2_{\text{alerting}} = 0.94$. Additionally, it was found that cued-recall performance followed a significant negative linear trend, such that as word-pair difficulty increased the proportion of targets correctly recalled decreased, $t(99) = -32.99, p < .001, R^2_{\text{alerting}} = 1.0$. Comparable to the results of the pilot study (Appendix 2), these results demonstrate that our manipulation of word-pair difficulty was valid. As well, in keeping with the previous literature, we found that JOLs were predictive of actual memory performance as we found a significant overall gamma correlation ($M = 0.58, SD = 0.14$), $t(99) = 42.03, p < .001$, Cohen's $d = 4.20$.

List composition

JOL ratings: In line with our hypothesis, a paired-samples *t*-test found that list composition had a significant effect on JOL ratings given for medium word pairs. More specifically, medium word pairs were given higher JOL ratings when presented in a difficult ($M = 58.75, SD = 9.37$), as compared to an easy ($M = 53.38, SD = 9.37$) context, $t(99) = 3.95, p < .001$, Cohen's $d = 0.32$, 95% CI [0.15, 0.48]; means are presented in Fig. 2 (panel 1).

Study-time allocation: Contrary to our hypothesis, however, a paired-samples *t*-test revealed that list composition had no effect on the amount of time participants spent studying medium word pairs. More specifically, there was no difference in the amount of time participants spent studying medium word pairs when presented in a difficult ($M = 3.77$ sec., $SD = 2.11$) as compared to an easy ($M = 3.64$ sec., $SD = 2.11$) context, $t(99) = 0.42, p = .67$, Cohen's $d = 0.05$, 95% CI [-0.18, 0.28], $BF_{10} = 0.12$; means are presented in Fig. 2 (panel 2).²

Cued-recall: To analyse the cued-recall data, we calculated the mean proportion of medium word pair targets correctly recalled when presented in an easy and difficult context (means are presented in Fig. 2, panel 3). Using a paired-samples *t*-test it was found that list composition did not have a significant effect on the proportion of targets recalled. More specifically, there was no

¹ R^2_{alerting} is the proportion of variability among condition means that can be attributed to the specific contrast being tested, and can range from 0 to 1.

² We also conducted a paired-samples *t*-test using the median study-time allocation for medium difficulty word pairs. Similarly, we found no difference in the amount of study time allocated to medium items in the easy ($M = 2.91, SD = 1.74$) and difficult ($M = 3.07, SD = 1.74$) contexts, $t(99) = 0.65, p = .52$, Cohen's $d = 0.07$, 95% CI [-0.14, 0.28], $BF_{10} = 0.14$.

Fig. 1 Word-pair difficulty results from Experiment 1. Panel 1 depicts mean JOL ratings given for easy, medium, and difficult word pairs. Panel 2 shows the mean study time for easy, medium, and difficult word pairs.

Panel 3 depicts the mean proportion of easy, medium, and difficult targets recalled. Error-bars represent 95% CIs corrected for within-subject comparisons (Morey, 2008)

difference in the proportion of targets recalled when medium word pairs were presented in an easy ($M = 0.56$, $SD = 0.12$) as compared to a difficult ($M = 0.55$, $SD = 0.12$) context, $t(99) = 0.46$, $p = .64$, Cohen's $d = 0.036$, 95% CI $[-0.11, 0.19]$, $BF_{10} = 0.12$.

Discussion

The results of Experiment 1 demonstrate that list composition can impact how individuals evaluate their learning, but that it does not seem to impact the control strategies implemented, at

Fig. 2 List composition results from Experiment 1. Panel 1 depicts mean JOL ratings given for medium word pairs when presented in an easy and difficult context. Panel 2 shows mean study time for medium word pairs when presented in an easy and difficult context. Panel 3 depicts the mean

proportion medium targets recalled in an easy and difficult context. Points represent individual subjects, and error-bars represent 95% CIs corrected for within-subject comparisons (Morey, 2008)

least as measured by self-paced study time. More specifically, the JOL results show that medium word pairs were judged to be more likely to be remembered when presented alongside more challenging (difficult context) word pairs than when presented alongside less challenging (easy context) word pairs. In contrast, the amount of time individuals spent studying medium word pairs did not differ when they were presented in either an easy or a difficult context. While the finding that study-time allocation was not sensitive to list composition was somewhat unexpected, it is not unprecedented. Previous research by Jia et al. (2016, Experiment 2a) has demonstrated that some manipulations (i.e., word frequency) can affect JOLs without having a corresponding effect on study-time allocation. Therefore, it appears as though list composition impacts how individuals evaluate their learning, but that it does not influence the amount of study time allocated to medium items.

Thus far, our attempt at unveiling whether list composition impacts the control strategies implemented by individuals has yielded no such evidence. Given that we looked only at one form of metacognitive control, namely study-time allocation, we questioned whether other forms of metacognitive control may be more sensitive to list composition effects. As discussed previously the link between individuals' JOLs and re-study selections is well established (Finn, 2008; Metcalfe & Finn, 2008; Nelson et al., 1994; Thiede & Dunlosky, 1999), and it may be that the relationship between JOLs and re-study selection is stronger than that between JOLs and study-time allocation. Therefore, we reasoned that list composition may have an effect on individuals' re-study decisions.

Experiment 2a

The general structure of Experiment 2a was similar to Experiment 1, with the only difference being the type of control strategy participants implemented. Rather than allowing participants to control the amount of time studying each word pair, we instead presented each word pair for an equal amount of time and asked participants to indicate whether, if given the chance, they would like to re-study the word pair at a later time. To the extent that re-study decisions are sensitive to list composition, we expected that medium word pairs would be selected for re-study more often when presented in a list composed of easy and medium word pairs (easy context) than when presented in a list composed of difficult and medium word pairs (difficult context).

Method

Participants Fifty-four undergraduate students from the University of Guelph psychology participant pool participated in this study in exchange for course credit (mean

age = 18.43 years, $SD = 0.92$, 48 female). The eligibility criteria used was the same as Experiment 1. The sample size was chosen based on a power analysis using the *BUCSS* package for *R* (Anderson & Kelley, 2018), using the *t*-value reported for the effect of list composition on JOLs found in Experiment 1, an assurance of .5, and the power to detect an effect of .80. All experimental procedures were approved by the Research Ethics Board (REB) at the University of Guelph.

Materials and design All materials and counterbalancing procedures were the same as those used in Experiment 1.

Procedure The procedure was the exact same as in Experiment 1 except that participants were not given an unlimited amount of time to study each word pair, instead each word pair was presented for 3.5 s. Additionally, after giving a JOL rating participants were required to indicate whether, if given the chance, they would like to re-study the word pair by selecting yes (1) or no (0). No word pairs were presented for re-study.

Results

Two participants' data were excluded due to technical errors. All exclusions were performed blind to whether data conformed to experimental predictions.

Pair difficulty Similar to Experiment 1, the mean JOL ratings, re-study decisions, and cued-recall performance were calculated for easy, medium, and difficult word pairs (see Fig. 3, panels 1–3, respectively), and analyzed using linear trend analyses. As in Experiment 1, JOL ratings were found to follow a significant negative linear trend, such that as word-pair difficulty increased, JOL ratings decreased, $t(51) = -12.36$, $p < .001$, $R^2_{\text{alerting}} = 0.99$. Re-study decisions were found to follow a significant positive linear trend, such that as word-pair difficulty increased, the proportion of word pairs chosen to re-study also increased, $t(51) = 9.36$, $p < .001$, $R^2_{\text{alerting}} = 0.98$. Also as in Experiment 1, cued-recall performance was found to follow a significant negative linear trend, such that as word-pair difficulty increased the proportion of targets correctly recalled decreased, $t(51) = -29.65$, $p < .001$, $R^2_{\text{alerting}} = 0.99$. Again, these results validate the difficulty of the experimenter generated word pairs. As well, similar to Experiment 1, the overall gamma correlation ($M = 0.54$, $SD = 0.21$) revealed that individuals' JOLs were predictive of actual memory performance, $t(51) = 18.17$, $p < .001$, Cohen's $d = 2.52$.

List composition

- **JOL ratings.** A paired-samples *t*-test found that list composition only had a marginally significant effect on the

Fig. 3 Word-pair difficulty results from Experiment 2a. Panel 1 depicts mean JOL ratings given for easy, medium, and difficult word pairs. Panel 2 shows the mean proportion of easy, medium, and difficult word pairs

selected for re-study. Panel 3 depicts the mean proportion of easy, medium, and difficult targets recalled. Error-bars represent 95% CIs corrected for within-subject comparisons (Morey, 2008)

JOL ratings given for medium word pairs. More specifically, JOL ratings given for medium word pairs were greater when presented in a difficult ($M = 52.18$, $SD = 11.09$) as compared to an easy ($M = 48.01$, $SD = 11.09$) context; however this difference did not reach statistical significance, $t(51) = 1.92$, $p = .06$, Cohen's $d = 0.22$, 95%

CI $[-0.01, 0.45]$, $BF_{10} = 0.82$; means are presented in Fig. 4 (panel 1). Given that this result was trending toward significance, and based on the results of Experiment 1, we decided to combine data for Experiments 1 and 2a to examine whether there was an overall effect of Context and if this effect differed between experiments. A 2

Fig. 4 List composition results from Experiment 2a. Panel 1 depicts the mean JOL ratings given for medium word pairs when presented in an easy and difficult context. Panel 2 shows the mean proportion of medium word pairs selected for re-study when presented in an easy and difficult

context. Panel 3 depicts the mean proportion of medium targets recalled when presented in an easy and difficult context. Points represent individual subjects, and error-bars represent 95% CIs corrected for within-subject comparisons (Morey, 2008)

(Context: Easy vs. Difficult) $\times$ 2 (Experiment: 1 vs. 2a) mixed analysis of variance (ANOVA) found significant main effects of Context, $F(1,150) = 15.18$, $p < .001$, $\eta^2_{\text{partial}} = 0.09$, such that medium word pairs were given higher JOL ratings in the difficult ($M = 56.15$, $SD = 9.88$) as compared to the easy ($M = 51.32$, $SD = 9.88$) context, and Experiment, $F(1, 150) = 4.27$, $p = .041$, $\eta^2_{\text{partial}} = 0.03$, such that medium word pairs were given higher JOL ratings in Experiment 1 ($M = 55.63$, $SD = 14.85$) as compared to Experiment 2a ($M = 50.09$, $SD = 17.15$). Critically however, there was no evidence that list composition differentially impacted JOL ratings across Experiments 1 and 2a, as no significant interaction was found, $F(1, 150) = 0.17$, $p = .68$, $\eta^2_{\text{partial}} = 0.001$, $BF_{10} = 0.20$.

- *Re-study selection.* In line with our hypothesis, a paired-samples *t*-test revealed that list composition had a significant effect on participants' decisions to re-study medium word pairs. More specifically, the proportion of medium word pairs selected for re-study was greater when presented in an easy ($M = 0.56$, $SD = 0.17$) than when presented in a difficult ($M = 0.41$, $SD = 0.17$) context, $t(51) = 4.47$, $p < .001$, Cohen's $d = 0.45$, 95% CI [0.23, 0.66]; means are presented in Fig. 4 (panel 2).
- *Cued-recall.* Again, to analyse the cued-recall data, we calculated the mean proportion of medium word pair targets correctly recalled when presented in an easy and difficult context (means are presented in Fig. 4, panel 3). Using a paired-samples *t*-test it was found that list composition did not have a significant effect on the proportion of targets recalled. More specifically, there was no difference in the proportion of targets recalled when medium word pairs were presented in an easy ($M = 0.50$, $SD = 0.10$) as compared to a difficult ($M = 0.50$, $SD = 0.10$) context, $t(51) = 0.09$, $p = 0.93$, Cohen's $d = 0.007$, 95% CI [-0.15, 0.16], $BF_{10} = 0.15$.

Discussion

Experiment 2a further demonstrated that list composition can impact how individuals evaluate their learning as measured by JOLs. More specifically, although the effect of list composition for JOL ratings was only marginally significant, we saw that this effect was not statistically different from the effect observed in Experiment 1. In contrast to Experiment 1 however, the results of Experiment 2a demonstrate that list composition can impact individuals' control behaviours as measured by the decision to re-study a given item. More specifically, the proportion of medium word pairs selected for re-study was greater when presented in an easy, as compared to a difficult context. However, given evidence that individuals'

JOLs are causally related to study choice (Metcalf & Finn, 2008), it is important to consider whether individuals' re-study decisions would still be impacted by list composition without the explicit requirement to monitor learning. This issue merits consideration because it is conceivable that re-study decisions made following a JOL could be directly influenced by the JOL itself. In other words, if an individual provides a high JOL rating, it may signal to him/her that re-study is unnecessary, and conversely, a low rating may motivate the individual to select the item for re-study. By this view, re-study decisions are entirely redundant with JOL ratings, and any effect of list composition on such decisions could simply reflect an individual's tendency to make re-study selections that are consistent with the JOL provided to that item. Moreover, it has been demonstrated that requiring individuals to explicitly predict their future memory performance can impact the study strategies they implement (Mitchum, Kelley, & Fox, 2016). Therefore, it is important to assess whether list composition is in fact impacting individuals' re-study decisions directly, or if the effect seen in Experiment 2a is an indirect consequence of the effect of list composition on JOLs.

Experiment 2b

To test whether list composition has a direct effect on re-study decisions that is independent of its effect on JOLs, Experiment 2b followed the same procedure as Experiment 2a but eliminated the requirement for individuals to provide item-by-item JOL ratings. Insofar as list composition has a direct effect on re-study decisions, we expected to replicate the findings of Experiment 2a such that medium word pairs would be selected for re-study more often when presented in an easy as compared to a difficult context.

Method

Participants Twenty undergraduate students from the University of Guelph psychology participant pool participated in this study in exchange for course credit (mean age = 18.25 years, $SD = 0.44$, 13 female). The eligibility criteria used was the same as Experiments 1 and 2a. The sample size was chosen based on a power analysis using the *BUCSS* package for *R* (Anderson & Kelley, 2018), using the *t*-value reported for the effect of list composition on re-study decisions found in Experiment 2a, an assurance of .5, and the power to detect an effect of .80. All experimental procedures were approved by the Research Ethics Board (REB) at the University of Guelph.

Materials and design All materials and counterbalancing procedures were the same as those used in Experiments 1 and 2a.

Procedure The procedure was the exact same as in Experiment 2a except that participants were not prompted to make a JOL rating prior to their re-study decision.

Results

Pair difficulty Similar to Experiments 1 and 2a, the mean re-study decisions, and cued-recall performance were calculated for easy, medium, and difficult word pairs (see Fig. 5, panels 1 and 2, respectively), and analyzed using linear trend analyses. As in Experiment 2a, re-study decisions were found to follow a significant positive linear trend, such that as word-pair difficulty increased, the proportion of word pairs selected for re-study also increased, $t(19) = 4.85$, $p < .001$, $R^2_{\text{alerting}} = 0.80$. As well, similar to Experiments 1 and 2a, cued-recall performance was found to follow a significant negative linear trend, such that as word-pair difficulty increased the proportion of targets correctly recalled decreased, $t(19) = -17.47$, $p < .001$, $R^2_{\text{alerting}} = 1.0$. Again, these results validate the difficulty of the experimenter-generated word pairs.

List composition

Re-study selection. Of primary interest given our hypothesis was the comparison of re-study decisions for medium word pairs when presented in an easy and difficult context. As in Experiment 2a, a paired-samples t -test

revealed that list composition had a significant effect on participants' decisions to re-study medium word pairs such that the proportion of medium word pairs selected for re-study was greater when presented in an easy ($M = 0.39$, $SD = 0.17$) than when presented in a difficult ($M = 0.18$, $SD = 0.17$) context, $t(19) = 3.90$, $p < .001$, Cohen's $d = 0.86$, 95% CI [0.34, 1.36]; means are presented in Fig. 6 (panel 1).

Cued-recall. Again, to analyse the cued-recall data, we calculated the mean proportion of medium word pair targets correctly recalled when presented in an easy and difficult context (means are presented in Fig. 6, panel 2). As in Experiments 1 and 2a, a paired-samples t -test found that list composition did not have a significant effect on the proportion of targets recalled. More specifically, there was no difference in the proportion of targets recalled when medium word pairs were presented in an easy ($M = 0.51$, $SD = 0.18$) as compared to a difficult ($M = 0.48$, $SD = 0.18$) context, $t(19) = 0.45$, $p = .66$, Cohen's $d = 0.13$, 95% CI [-0.43, 0.68], $BF_{10} = 0.25$.

Cross-experiment analysis To examine whether list composition had a differential effect on re-study decisions between Experiments 2a and 2b a 2 (Context: Easy vs. Difficult) $\times$ 2 (Experiment: 2a vs. 2b) mixed ANOVA was conducted (see

Fig. 5 Word-pair difficulty results from Experiment 2b. Panel 1 shows the mean proportion of easy, medium, and difficult word pairs selected for re-study. Panel 2 depicts the mean proportion of easy, medium, and

difficult targets recalled. Error-bars represent 95% CIs corrected for within-subject comparisons (Morey, 2008)

Fig. 6 List composition results from Experiment 2b. Panel 1 shows the mean proportion of medium word pairs selected for re-study when presented in an easy and difficult context. Panel 2 depicts the mean

proportion of medium targets recalled when presented in an easy and difficult context. Points represent individual subjects, and error-bars represent 95% CIs corrected for within-subject comparisons (Morey, 2008)

Fig. 7, which depicts the means for all four conditions). Critically, there was no significant interaction between Context and Experiment, $F(1,70) = 0.98$, $p = .32$, $\eta^2_{\text{partial}} = 0.01$, $BF_{10} = 0.40$. There was, however, a significant main effect of Context, $F(1, 70) = 32.43$, $p < .001$, $\eta^2_{\text{partial}} = 0.32$, such that the proportion of medium word pairs selected for re-study was greater when presented in an easy ($M = 0.51$, $SD = 0.33$), as compared to a difficult ($M = 0.35$, $SD = 0.30$) context, regardless of experiment. There was also a significant main effect of Experiment, $F(1, 70) = 7.75$, $p = .006$, $\eta^2_{\text{partial}} = 0.10$, such that the proportion of medium word pairs selected for re-study was greater in Experiment 2a ($M = 0.49$, $SD = 0.33$) than in Experiment 2b ($M = 0.28$, $SD = 0.26$), regardless of context.

Discussion

The results of Experiment 2b demonstrate that even in the absence of making a JOL, list composition continued to impact individuals' re-study decisions. Replicating the findings from Experiment 2a, we found that the proportion of medium word pairs selected for re-study was greater when presented in an easy, as compared to a difficult context. Therefore, list

composition does appear to have a direct influence on re-study decisions that is independent of its effect on JOLs. It is also important to point out, however, that in comparing the results of Experiments 2a and 2b, it was evident that eliminating the requirement to make a JOL *did* reduce the overall likelihood with which participants selected a given item for re-study. Therefore, the explicit act of making a JOL appears to have consequences for metacognitive control – a finding that is consistent with research demonstrating that JOLs are not inert markers of perceived learning (Mitchum et al., 2016; Soderstrom, Clark, Halamish, & Bjork, 2015). In fact, our results suggest that, in addition to altering subsequent memory performance (Mitchum et al., 2016; Soderstrom et al., 2015) and study-time allocation policies (Mitchum et al., 2016), JOLs can also induce reactive shifts in processing that affect decisions regarding what learners choose to re-study.

General discussion

The results of the current series of experiments provide new insights on the correspondence between monitoring and control, and shed light on how list composition

Fig. 7 Combined results for medium items in Experiments 2a and 2b. Panel 1 depicts the mean proportion of medium word pairs selected for re-study when presented in an easy and difficult content across Experiments 2a and 2b. Panel 2 depicts the mean proportion of medium targets recalled

when presented in an easy and difficult context across Experiments 2a and 2b. Error-bars represent 95% CIs corrected for within-subject comparisons (Morey, 2008)

affects these two processes. First, as reported by Zawadzka and Higham (2016) and Susser et al. (2013), the results of Experiments 1 and 2a support the notion that JOLs are sensitive to the relative differences between classes of items. Specifically, our manipulation of list composition demonstrated that stimuli of equal difficulty are reliably judged to be more or less memorable depending on the relative context in which they are presented, with this effect being small to medium in magnitude ($\eta^2_{\text{partial}} = 0.09$). Second, with regard to metacognitive control strategies, our results indicate that list composition may not impact all control strategies in the same fashion. Specifically, in Experiment 1 we saw no effect of list composition on study-time allocation, such that stimuli of equal difficulty were allocated the same amount of study time regardless of the context in which they were presented. Conversely, Experiments 2a and 2b demonstrate that individuals' decision to re-study was impacted by list composition, as stimuli of equal difficulty were more likely to be selected for re-study when presented in an easy as compared to a difficult context. Furthermore, Experiment 2b demonstrated that this effect was not contingent upon having to explicitly make a JOL, and that list composition can affect re-study decisions directly.

Mechanisms underlying the effects of list composition

The results from two of the three experiments reported here demonstrate that JOLs are sensitive to relative differences among items in a study list. Although our experiments were not designed to tease apart different mechanistic accounts of the list composition effects reported here and elsewhere (Susser et al., 2013; Zawadzka & Higham, 2016), our findings both fit well with previous accounts (Zawadzka & Higham, 2016) and unveil new insights into the processes that underlie the monitoring and control of learning.

According to the signal detection theory (SDT) approach advanced by Zawadzka and Higham (2016), a JOL can be conceptualized as reflecting the confidence with which a learner will later remember a given item. By this view, different items vary continuously in the degree of evidence they elicit for future recall with confidence (JOLs) being proportional to the degree of evidence. Importantly, the learner is thought to parse this evidence continuum into a set of discrete criteria – each of which has an associated JOL rating. Using an example similar to Zawadzka and Higham (2016), if an individual were to partition their evidence continuum in increments of 20, a JOL rating of 60 for a given item would mean that the level of evidence in favour of future recall of that item is equal to or greater than their

criterion of 60, but less than their criterion of 80. To understand how list composition affects JOLs, Zawadzka and Higham (2016) posited that when a given set of items (e.g., medium-difficulty) are studied alongside a set of easier items, learners adjust their confidence criteria to accommodate the presence of these easier items. Critically, the confidence criteria associated with the high end of the JOL scale are shifted upward, such that only items that elicit a high degree of evidence for future recall surpass these criteria and are consequently assigned a high JOL. Therefore, on average, a medium item would be assigned a lower JOL when studied in the presence of easier items than when studied alone, as that item would be relatively less likely to surpass the higher criteria that are shifted upward. Conversely, the opposite shift would occur when a given set of items (e.g., medium-difficulty) are studied alongside more difficult items, resulting in these items receiving a higher JOL on average than when studied alone. Zawadzka and Higham (2016) supported this interpretation using receiver operating characteristic (ROC) analyses that revealed selective shifts in JOL decision criteria as a function of study context. A key insight derived from this account is that a given item does not elicit more or less evidence for future recall depending on the context within which it is studied. By this view, medium items are not *perceived* as more or less memorable when studied alongside easier or more difficult items, respectively. Rather, the difference in mean JOLs for these items is a consequence of learners shifting their evidence criteria to accommodate the range of evidence elicited by different classes of items.

With respect to the findings reported here, we believe this SDT approach may help explain why we found no influence of list composition on self-paced study-time allocation for medium items despite finding such an effect for re-study decisions. Although both of these control functions have been found previously to be inversely related to JOLs (Dunlosky & Thiede, 1998; Finn, 2008; Koriat, 2008; Metcalfe & Finn, 2008; Thiede & Dunlosky, 1999), there is strong evidence to suggest that self-paced study-time allocation in particular is dictated by the intrinsic difficulty of an item in a data-driven fashion (Koriat, 2008; Koriat, Ma'ayan, & Nussinson, 2006). Therefore, insofar as our manipulation of list composition did not change the perceived difficulty/memorability of medium items, the absence of a list composition effect on study-time allocation for these items is entirely consistent with the theoretical positions of both Koriat et al. (2006, 2008) as well as

Zawadzka and Higham (2016). In contrast, the fact that a list composition effect *was* observed for re-study decisions alludes to the possibility that such decisions are more mechanistically aligned with JOLs. That is, it could be the case that the decision to re-study a given item is a function of both the perceived utility of re-study (evidence signal) together with the placement of a single, binary criterion that defines the boundary between the decision to re-study and the decision to not re-study (see Fig. 8 for a graphic illustration). Accordingly, to the extent that list composition recalibrates the placement of decision criteria rather than affecting perceived memorability, one might expect to

Fig. 8 Graphic illustration of how shifts in re-study criteria could accommodate the effect of list composition on re-study decisions. Panel 1 depicts criterion placement when medium word pairs are presented alongside easy word pairs (easy context). Panel 2 depicts criterion placement when medium word pairs are presented alongside difficult word pairs (difficult context). Critically, in both panels it is shown that the perceived evidence that re-study will benefit medium items does not change between contexts; however, more medium items fall to the right of the criterion placement in panel 1 as compared to panel 2 and therefore more medium items would be selected for re-study in the easy as compared to the difficult content

observe a parallel influence of list composition on JOLs and re-study decisions, but not on JOLs and study-time allocation. Indeed, this pattern is precisely what was observed here.

Although this interpretation is broadly consistent with the SDT account forwarded by Zawadzka and Higham (2016), it is also important to note that previous research by Hanczakowski, Zawadzka, Pasek, and Higham (2013) and Zawadzka and Higham (2015) has questioned whether the context-driven criterion shifts that appear to occur for judgments with multiple scale criteria (e.g., 0–100 JOLs) can also be assumed to occur for judgments based on a single criterion (e.g., yes/no re-study decisions). For example, Hanczakowski et al. (2013) demonstrated that the well-known *underconfidence-with-practice* effect (i.e., the finding that JOLs begin to display underconfidence after multiple study-test cycles; see Koriat, Sheffer, & Ma'ayan, 2002) was observed when participants made 0–100 scale JOLs, but not when binary yes/no JOLs were elicited. This finding was later interpreted as evidence that single criterion decisions may not be as susceptible to criterion shifts as decisions with multiple criteria (Zawadzka & Higham, 2015, 2016). Despite this finding, however, other research has demonstrated clearly that context-driven criterion shifts can in fact occur for other types of binary decisions. For example, in tests of recognition memory using the remember/know procedure (Tulving, 1985; see Yonelinas, 2002, for a review), similar contextual manipulations implemented at both study (McCabe & Balota, 2007) and test (Bodner & Lindsay, 2003) were associated with a difference in the likelihood with which participants provided a “remember” response to items endorsed as previously studied (i.e., “old”). McCabe and Balota (2007) showed that this difference in the tendency to classify recognition decisions as involving “remembering” was a direct consequence of a criterion shift driven by the expected level of memorability of test items. More specifically, when relatively weak items were studied in the context of strong items (repeated 5X), participants were more likely to provide a “remember” response to weak items at test when they were told that they would be tested on the weak items only as compared to when they expected to be tested on both the weak and strong items. Similarly, Bodner and Lindsay (2003) found that the likelihood of a “remember” response at test for items initially studied in a medium levels-of-processing (LOP) task changed as a function of whether these items were tested alongside items initially studied in a

shallow or deep LOP task. Together, these results imply that list composition can indeed influence criterion placement for at least some kinds of binary decisions (i.e., “remember” responses). Therefore, it seems reasonable to suggest that the list composition manipulation employed here altered participants’ re-study criterion. Further research is needed to better understand the conditions under which such criterion shifts are likely to occur, and whether single and multiple criteria decisions are always affected by context in the same fashion.

Thus far we have interpreted our findings within the SDT framework proposed by Zawadzka and Higham (2016), which emphasizes the role of criterion shifts in accounting for the effect of list composition on JOLs. An alternative account, however, is that the observed influence of list composition on JOLs and re-study decisions reflects a genuine difference in the underlying evidence signal, such that medium items in the easy context condition are truly perceived as less memorable than medium items in the difficult context condition. By this view, the observed dissociation between re-study decisions and study-time allocation could simply reflect the fact that study-time allocation for medium items may be less sensitive to relatively small changes in perceived memorability owing to variations in list composition. However, we find this explanation unsatisfying in that it begs the question of *why* study time is less sensitive. Furthermore, study time was just as sensitive as re-study decisions to overall differences in item difficulty. We acknowledge that distinguishing between shifts in decision criteria and genuine changes in perceived memorability/confidence is challenging (see Pansky & Goldsmith, 2014; Portnoy & Pansky, 2016), and that we cannot definitively conclude that the influence of list composition on JOLs and re-study decisions observed here is truly a product of context-driven re-calibration of scale criteria. Nevertheless, we believe this criterion-based explanation most naturally accommodates the observed dissociation between study-time allocation and re-study decisions. We would also like to point out that, regardless of the specific mechanism by which list composition affects JOLs and re-study decisions, the present findings have practical implications in that they demonstrate that some, but not all measures of metacognitive control are affected by such manipulations of study context.

There is one final issue that merits further scrutiny. To this point, we have suggested the null effect of list composition on study-time allocation may reflect the fact that study time is largely driven by bottom-up differences in item difficulty (Koriat, 2008; Koriat et al.,

2006), and therefore that list composition ought to have little effect on this measure insofar as it affects the placement and calibration of decision criteria rather than perceived difficulty/memorability. However, it is possible that list composition may well have affected study time for medium items, but that its influence was masked by participant-level variation in allocation policies. Past research has shown that the relation between JOLs and various metacognitive control functions including self-paced study time and re-study selection is moderated by many factors including time pressure and one's learning goals (Son & Metcalfe, 2000; Thiede & Dunlosky, 1999). For example, Thiede and Dunlosky (1999) showed that participants instructed to learn only a small number of items (low-learning goal) tended to choose easier items for re-study, yielding a positive correlation between re-study selection and JOLs. In contrast, participants instructed to learn as many items as possible (high-learning goal) tended to choose more difficult items for re-study, yielding a negative correlation between re-study selection and JOLs. Although Thiede and Dunlosky (1999) found that different learning goals did not appear to substantially change how participants allocated their study time, it remains possible that different learning goals could have played a role in how list composition affected study-time allocation in the present study insofar as different participants spontaneously adopted different learning goals. If this were the case, it would be expected that those who adopted a low-learning goal may have allocated more time to medium items in the difficult as compared to easy context condition, whereas those who adopted a high-learning goal may have allocated more time to medium items in the easy as compared to difficult context condition. If roughly half of participants adopted each of these learning goals, the net effect of context on study-time allocation for the medium items would be near zero. To investigate this possibility, we conducted a median-split analysis based on cued-recall performance and analyzed the effect of list composition on study time separately for the two groups (High/Low cued-recall performance). Participants whose cued-recall performance was greater than the median were presumed to have adopted a high-learning goal, whereas those whose performance was lower, a low-learning goal. Critically, a 2 (Context: Easy vs. Difficult) $\times$ 2 (Learning Goal: High vs. Low) mixed ANOVA found a significant interaction between Context and Learning Goal, $F(1, 98) = 5.05$, $p = .03$, $\eta^2_{\text{partial}} = 0.05$. However, subsequent investigation of

the cell means revealed that participants who presumably adopted a high-learning goal allocated more study time to medium items in the difficult ($M = 4.54$, $SD = 2.46$) as compared to the easy ($M = 3.79$, $SD = 2.46$) context, whereas those participants who presumably adopted a low-learning goal allocated more study time to medium items in the easy ($M = 3.47$, $SD = 1.52$) as compared to the difficult ($M = 2.90$, $SD = 1.52$) context, though neither of these simple contrasts reached significance, $p = .12$, $BF_{10} = 0.47$ and $p = .07$, $BF_{10} = 0.73$, respectively. Therefore, if anything, our results are opposite in direction to what would be predicted by the learning goals hypothesis. It is therefore unlikely that this hypothesis, at least as typically conceived, can account for the null effect of list composition on study-time allocation.

Metacognitive control strategies and reactivity to JOLs

Another interesting finding from the present series of experiments is the results from the cross-experimental analysis between Experiments 2a and 2b. While we did not see a difference in the effect of list composition on re-study selection between the two experiments, we did see an effect of Experiment such that participants were in general more likely to select items for re-study when they were required to make an explicit JOL rating. This finding expands upon the results from Mitchum et al. (2016) in that the explicit requirement to make a JOL not only affects the amount of study time allocated to a given item (Mitchum et al., 2016), but also whether an item is selected for re-study. Specifically, our results suggest that the requirement to provide item-by-item JOLs increases the overall likelihood with which a given item is selected for re-study. This finding is also broadly consistent with Mitchum et al.'s (2016) insight that requiring participants to provide an explicit JOL likely invites them to consider the fact that some items may be more memorable than others, which may in turn increase the caution with which they approach learning. Although we did not instruct participants to explicitly monitor their learning by making item-by-item JOLs in Experiment 2b, it is certainly possible that prompting individuals to make a re-study decision in the absence of a JOL may have nevertheless induced them to monitor their learning in a similar fashion. However, it is important to note that this spontaneous monitoring does not appear to be

happening as frequently or as judiciously, as the overall proportion of words selected for re-study was smaller than when the requirement to monitor was explicitly implemented by having participants make JOLs.

It is also informative to consider the results of Experiments 2a and 2b in the context of those reported by Metcalfe and Finn (2008). In their study, the authors examined whether manipulations that produced illusory differences in JOLs also produced a corresponding difference in the likelihood of selection for re-study. Across multiple different experiments and experimental paradigms, they showed that re-study decisions aligned with JOLs rather than objective memory performance, leading them to conclude that JOLs have a causal influence on re-study choice. While this pattern of findings is certainly informative and weighs against the idea that metacognitive control is driven by direct-access to stored memory traces, it does leave open the question of whether re-study decisions are sensitive to such manipulations in the absence of JOLs, as the authors did not include a condition in which participants were required to make a re-study decision in the absence of a prior JOL. Thus, based on their results alone, it is unclear whether re-study decisions are a rather uninteresting by-product of JOLs themselves, or whether they are based on similar information as JOLs but are mechanistically independent. The results of Experiment 2b support the latter alternative by demonstrating that the parallel effect of list composition on JOLs and re-study decisions observed in Experiment 2a was not contingent on the requirement to make a JOL.

From the standpoint of measurement reactivity, the finding that list composition affected re-study decisions even in the absence of JOLs is also interesting when juxtaposed with the null effect of list composition on study-time allocation seen in Experiment 1. A potentially important, yet often overlooked methodological difference between measures of study-time allocation and re-study decisions is that the former is often measured *prior* to giving a JOL, whereas the latter is often measured *following* a JOL (see Metcalfe & Finn, 2008; Rhodes & Castel, 2009). This discrepancy means that measures of study-time allocation are less likely to be contaminated by explicit JOLs than are re-study decisions. Accordingly, attempts to examine whether a given variable has corresponding effects on each of these measures is often confounded by differences in the sequence in which monitoring and control are probed. However, because the requirement to make a JOL was removed in Experiment 2b, a

direct comparison can be made between the influence of list composition on re-study decisions in Experiment 2b and on study-time allocation in Experiment 1. The fact that the effect of list composition on re-study selection persisted in Experiment 2b is strong evidence against the notion that the observed dissociation between re-study selection and study-time allocation is a product of the differential order in which monitoring and control were assessed in Experiments 1 and 2a. Similarly, it is also conceivable that these sequencing differences between Experiments 1 and 2a could impact JOLs. Given that the relation between monitoring and control is often considered bi-directional (see Koriati, 2008; Koriati et al., 2006), invested study time could well have influenced JOLs in Experiment 1 in which self-paced study preceded JOLs. Although it was found that the overall magnitude of JOLs differed between Experiments 1 ($M = 58.07$, $SD = 13.30$) and 2a ($M = 52.03$, $SD = 16.31$), $t(150) = 2.45$, $p = .02$, Cohen's $d = 0.41$, there was no evidence that the effect of list composition differed across experiments. This latter finding is perhaps not surprising given that we did not find any differences in invested study time for medium items as a function of list composition.

Conclusion

Together, the findings reported here provide important new insights on the relationship between metacognitive monitoring and control. First, our results confirm previous findings that JOLs are indeed sensitive to relative differences among items (Dunlosky & Matvey, 2001; Susser et al., 2013; Zawadzka & Higham, 2016). Second, we also found evidence that although such sensitivity to relative differences does not appear to influence the length of time individuals spend studying a given item, it does affect what they choose to re-study. Thus, our results imply that the construct of metacognitive control is not entirely homogenous, and that different forms of control may be driven by different mechanisms. We believe these findings have important applied implications, and may assist in the development of more effective guidance on how to best engage in self-regulated learning.

Acknowledgements This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through a Discovery Grant to C.M.F. We would also like to thank Brianna Mitton and Silas Manning for their assistance with data collection.

Appendix 1

Stimulus pool

Table 1 Stimuli used in Experiments 1–2b

Easy	Medium (sub-list A)	Medium (sub-list B)	Difficult
paper – pen	educator – pencil	widow – estate	phrase – fashion
husband – wife	tent – forest	fur – nature	tongue – art
aluminum – foil	anchor – navy	grass – blade	tight – distance
magazine – article	earth – sphere	kingdom – tyranny	emperor – heavy
aunt – uncle	key – bone	produce – orange	pill – cellar
autumn – fall	pine – oak	quiet – ear	village – rare
salt – pepper	brain – gray	milk – jersey	fear – vision
brother – sister	highway – urban	pressure – vein	jail – bite
rubber – band	iron – engineer	cow – butter	younger – strain
honey – bee	lemon – sun	ghost – sheet	lane – desk
belly – button	victim – cell	author – series	steam – weep
tree – branch	cocktail – tip	swimming – tide	justice – growth
lung – cancer	oxygen – plant	moon – phase	arrow – nerve
candy – cane	hospital – cure	dozen – bread	murder – message
wedding – ring	wound – aid	valley – pit	blanket – tube
prince – charming	vice – deputy	youth – calf	leader – page
copper – wire	corn – whiskey	wool – shear	stem – egg
dentist – teeth	customer – window	prairie – rural	gold – kid
fishing – net	palace – servant	clover – rose	increase – ice
exchange – student	evidence – tape	division – sum	danger – quarter
queen – king	beg – knee	sky – ocean	wrote – occasion
visitor – guest	freight – railroad	industry – fuel	neck – unique
yard – sale	deer – shot	steak – ham	zero – rise
east – west	debate – ballot	physics – mass	hail – object

Table 2 Results of comparisons made for latent semantic analysis (LSA) values between word pair lists, descriptive statistics included

	<i>M</i>	<i>SD</i>	1			2			3		
			<i>t</i> (46)	<i>p</i>	<i>BF</i> ₁₀	<i>t</i> (46)	<i>p</i>	<i>BF</i> ₁₀	<i>t</i> (46)	<i>p</i>	<i>BF</i> ₁₀
1. Easy	0.4094	0.3006									
2. Medium (sub-list A)	0.1441	0.2115	3.54	<.001	>10						
3. Medium (sub-list B)	0.0864	0.0897	5.04	<.001	>10	1.23	.225	0.531			
4. Difficult	0.0068	0.0105	6.56	<.001	>10	3.18	.003	>10	4.32	<.001	>10

Note. *M* and *SD* represent mean and standard deviation, respectively

Appendix 2

Pilot study

The purpose of the pilot experiment was to validate the experimenter generated word pairs. Since the word pairs were generated by the experimenter, we wanted to ensure that the presumed difficulty of the word pairs was accurate. More specifically, we wanted to ensure the “easy” and “medium” word pairs were more related and therefore easier to learn than the “difficult” word pairs, and that the “easy” word pairs were more related and easier to learn than the “medium” word pairs. As well, because the medium word pairs were divided into two sub-lists, we wanted to ensure that these two lists were equated for relatedness and how easy they are to learn.

Method

Participants Ten undergraduate students from the University of Guelph psychology participant pool participated in this study in exchange for course credit (mean age = 18.8 years, $SD = 2.25$, eight female). The eligibility criteria used was the same as that used in Experiments 1–2b. All experimental procedures were approved by the Research Ethics Board (REB) at the University of Guelph.

Materials All materials were the same as those used in Experiments 1–2b.

Procedure

Study phase. Participants were brought into the lab and seated in front of the testing computer. After obtaining informed consent, and a general overview of the experiment was given, participants were presented with all 96 word pairs in one continuous block. Upon each presentation the participant was required to make a relatedness judgment from 1 to 10 indicating how related they thought the words were (i.e., 1 indicates not at all related, 10 indicates

highly related), and an ease of learning (EOL) judgment from 1 to 10 indicating how easy they thought it would be to learn each word pair (i.e., 1 indicates not at all easy, 10 indicates extremely easy). The order of the word pairs was randomized, and the order of the two ratings was counterbalanced. Each rating was self-paced with no experimenter-imposed time limit. Participants were not told that there would be a cued-recall test later on.

Test phase. Immediately following the study phase, participants were presented with all cues from the word pairs one at a time, and were required to recall the corresponding target word. Again, the order of the cues was randomized and all presentations were self-paced with no experimenter-imposed time limit.

Results

Relatedness judgments Four paired-samples t -tests were conducted to assess the difference in relatedness judgments given for easy, medium, and difficult word pairs. We found a significant difference in relatedness judgments for easy ($M = 9.25$, $SD = 0.64$) and medium ($M = 6.72$, $SD = 0.95$) word pairs, $t(9) = 13.13$, $p < .001$, Cohen's $d = 2.94$, CI [1.53, 4.33], medium and difficult ($M = 3.62$, $SD = 0.90$) word pairs, $t(9) = 9.73$, $p < .001$, Cohen's $d = 2.65$, CI [1.32, 3.96], and easy and difficult word pairs, $t(9) = 13.72$, $p < .001$, Cohen's $d = 5.13$, CI [2.68, 7.56]. There was no significant difference in the relatedness judgments given for medium word pairs from sub-list A ($M = 6.72$, $SD = 0.25$) and sub-list B ($M = 6.73$, $SD = 0.53$), $t(9) = 0.05$, $p = .96$, Cohen's $d = 0.008$, CI [-0.32, 0.33], $BF_{10} = 0.31$.

Mean relatedness ratings were calculated for each item. As seen in Fig. 9 most easy items were given high relatedness ratings and most difficult items were given low relatedness ratings. Furthermore, there is a high amount of overlap for the relatedness ratings given for medium items from sub-lists A and B.

Fig. 9 Number of items receiving an average relatedness rating separated by word-pair difficulty and sub-list

EOL judgments Four paired-samples *t*-tests were conducted to assess the difference in EOL judgments given for easy, medium, and difficult word pairs. There was a significant difference in EOL judgments for easy ($M = 9.31$, $SD = 1.24$) and medium ($M = 7.10$, $SD = 1.05$) word pairs, $t(9) = 5.17$, $p < .001$, Cohen's $d = 1.92$, CI [0.76, 3.04], medium and difficult ($M = 4.70$, $SD = 1.05$) word pairs, $t(9) = 7.19$, $p < .001$, Cohen's $d = 1.72$, CI [0.80, 2.2], and easy and difficult word pairs, $t(9) = 7.58$, $p < .001$, Cohen's $d = 3.13$, CI [1.47, 4.74]. There was no signifi-

cant difference in EOL judgments given for medium word pairs from sub-list A ($M = 7.22$, $SD = 0.48$) and sub-list B ($M = 6.97$, $SD = 0.56$), $t(9) = 1.39$, $p = .20$, Cohen's $d = 0.23$, CI [-0.12, 0.57], $BF_{10} = 0.67$.

Mean EOL judgments were calculated for each item. As seen in Fig. 10 most easy items were given high EOL ratings and most difficult items were given low EOL ratings. Furthermore, there is a high amount of overlap for the EOL ratings given for medium items from sub-lists A and B.

Fig. 10 Number of items receiving an average EOL rating separated by word-pair difficulty and sub-list

Cued–Recall Four paired-samples *t*-tests were conducted to assess the difference in the proportion of correct targets recalled for easy, medium, and difficult word pairs. There was a significant difference in the proportion of targets recalled for easy ($M = 0.85$, $SD = 0.19$) and medium ($M = 0.61$, $SD = 0.26$) word pairs, $t(9) = 3.32$, $p = .009$, Cohen's $d = 1.23$, CI [0.29, 2.14], medium and difficult ($M = 0.30$, $SD = 0.10$) word pairs, $t(9) = 8.15$, $p < .001$, Cohen's $d = 1.30$, CI [0.62, 1.96], and easy and difficult word pairs, $t(9) = 8.28$, $p < .001$, Cohen's $d = 3.31$, CI [1.60, 4.98]. There was no signif-

icant difference between the proportion of targets recalled for medium word pairs from sub-list A ($M = 0.62$, $SD = 0.11$) and sub-list B ($M = 0.60$, $SD = 0.07$), $t(9) = 0.37$, $p = .72$, Cohen's $d = 0.05$, CI [-0.21, 0.30], $BF_{10} = 0.33$.

The mean proportion of times the target was correctly recalled was calculated for each item. As seen in Fig. 11 most easy items had very high recall proportions and most difficult items had low recall proportions. Furthermore, there is a high amount of overlap for the mean recall proportions for medium items from sub-lists A and B.

Fig. 11 Number of items receiving an average proportion of targets recalled separated by word-pair difficulty and sub-list

References

- Anderson, S. F., & Kelley, K. (2018). BUCSS: Bias and uncertainty corrected sample size. R package version 1.0.0. <https://CRAN.R-project.org/package=BUCSS>
- Bodner, G. E., & Lindsay, D. S. (2003). Remembering and knowing in context. *Journal of Memory and Language*, 48(3), 563 – 580.
- Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*, 25(5), 691 – 700.
- Dunlosky, J., & Matvey, G. (2001). Empirical analysis of the intrinsic-extrinsic distinction of judgments of learning (JOLs): Effects of relatedness and serial position on JOLs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(5), 1180 – 1191.
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica*, 98(1), 37 – 56.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behaviour Research Methods*, 39, 175–191. <https://doi.org/10.3758/BF03193146>
- Fiechter, J. L., Benjamin, A. S., & Unsworth, N. (2016). The metacognitive foundations of effective remembering. In Dunlosky, J., & Tauber, S. K. (Eds.), *The Oxford Handbook of Metamemory* (pp. 307 – 324). Oxford University Press.
- Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, 36(4), 813 – 821.
- Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun – An R package for computations based on latent semantic analysis. *Behaviour Research Methods*, 47(4), 930 – 944.
- 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(3), 429 – 444.
- Hirshman, E., & Mulligan, N. (1991). Perceptual interference improves explicit memory but does not enhance data-driven processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(3), 507 – 513.
- Jia, X., Li, P., Li, X., Zhang, Y., Cao, W., Cao, L., & Li, W. (2016). The effect of word frequency on judgments of learning: Contributions of beliefs and processing fluency. *Frontiers in Psychology*, 6, 1995.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126(4), 349 – 370.
- Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, 36(2), 416 – 428.
- Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behaviour. *Journal of Experimental Psychology: General*, 135(1), 36 – 69.
- Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, 131(2), 147 – 162.
- Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(3), 609 – 622.
- Kruschke, J. K. (2013). Bayesian Estimation Supersedes the *t* test. *Journal of Experimental Psychology: General*, 142(2), 573 – 603.

- McCabe, D. P., & Balota, D. A. (2007). Context effects on remembering and knowing: The expectancy heuristic. *Journal of Experimental Psychology: Learning, memory, and cognition*, 33(3), 536 – 549.
- McDaniel, M. A., & Bugg, J. M. (2008). Instability in memory phenomena: A common puzzle and a unifying explanation. *Psychonomic Bulletin & Review*, 15(2), 237 – 255.
- Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, 131(3), 349 – 363.
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, 18(3), 159 – 163.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, 15(1), 174 – 179.
- Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General*, 145(2), 200 – 219.
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau. *Tutorial in Quantitative Methods for Psychology*, 4(2), 61 – 64.
- Morey, R. D., & Rouder, J. N. (2018). BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12-4.2. <https://CRAN.R-project.org/package=BayesFactor>
- Mulligan, N. W. (1999). The effects of perceptual interference at encoding on organization and order: Investigating the roles of item-specific and relational information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 54 – 69.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. <http://www.usf.edu/FreeAssociation/>.
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science*, 5(4), 207 – 213.
- Pansky, A., & Goldsmith, M. (2014). Metacognitive effects of initial question difficulty on subsequent memory performance. *Psychonomic Bulletin & Review*, 21(5), 1255 – 1262.
- Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, 51(1):195 – 203
- Portnoy, S., Pansky, A. (2016). Metacognitive effects of initial question difficulty on subsequent eyewitness memory performance. *Journal of Applied Research in Memory and Cognition*, 5(2), 159 – 167.
- R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rhodes, M., & Castel, A. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, 137(4), 615 – 625.
- Rhodes, M. G. (2016). Judgments of learning: Methods, Data, and Theory. In Dunlosky, J., & Tauber, S. K. (Eds.), *The Oxford Handbook of Metamemory* (pp. 65 – 80). Oxford University Press.
- Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review*, 16(3), 550 – 554.
- Richards, R. M., & Nelson, T. O. (2004). Effect of the difficulty of prior items on the magnitude of judgments of learning for subsequent items. *American Journal of Psychology*, 117(1), 81 – 91.
- Schwartz, B. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, 1(3), 357 – 375.
- Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(2), 553 – 558.
- Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(1), 204 – 221.
- 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(7), 1000 – 1011.
- Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 1024 – 1037.
- Tulving, E. (1985). Memory Consciousness. *Canadian Psychology/Psychologie Canadienne*, 26(1), 1 – 12.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, 46(3), 441 – 517.
- Zawadzka, K., & Higham, P. A. (2015). Judgments of learning index relative confidence, not subjective probability. *Memory & Cognition*, 43(8), 1168 – 1179.
- Zawadzka, K., & Higham, P. A. (2016). Recalibration effects in judgments of learning: A signal detection analysis. *Journal of Memory and Language*, 90(C), 161 – 176.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.