Archives of Scientific Psychology 2013, 1, 7–13
DOI: http://dx.doi.org/10.1037/arc0000002

2169-3269

American Psychological Association

Archives of Scientific Psychology

http://www.apa.org/pubs/journals/arc

Phrasing Questions in Terms of Current (Not Future) Knowledge Increases Preferences for Cue-Only Judgments of Learning

Ivo Todorov
Stockholm University

Nate Kornell
Williams College

Max Larsson Sundqvist
Stockholm University, Sweden

Fredrik U. Jönsson
Stockholm University and Stockholm Brain Institute, Stockholm, Sweden

ABSTRACT

Effective learning demands knowledge about what learning strategies are most effective. Much research has addressed what students' know about how to improve memory. However, to effectively study it is also important to accurately feel (i.e., monitor) how well or poorly you have learned; for example, a glossary list, because such monitoring is closely related to the decisions students make about what to restudy. Such monitoring, termed judgments of learning (JOLs), predict later recall of glossaries (i.e., word pairs) more accurately when they are made after a delay, while viewing the first word only (cue) compared with both words in a word pair (cue and target). We investigated whether people recognize the benefit of cue-only responses when making JOLs and whether their preferences depend on how JOL prompts are phrased. Forty participants studied glossaries and then made delayed cue-only and cue-target JOLs. When the JOL prompts were phrased as predictions of future memory performance, only 15% of the participants preferred the better cue-only strategy. When JOLs were instead phrased as assessments of the current state of learning, 55% preferred the cue-only strategy. To conclude, students do not seem to recognize the value of cue-only JOLs, but they picked the superior JOL strategy more often when the JOL phrasing focused their attention on their knowledge state at the time of the JOL, rather than on a future state. This indicates that study-advice to students should not only include information about how to improve memory, but also about how to improve monitoring.

SCIENTIFIC ABSTRACT

Judgments of learning (JOLs) predict later recall more accurately when they are made, after a delay, based on a cue alone compared with a cue and target. We investigated whether people recognize the benefit of cue-only responses when making JOLs and whether their preferences depend on how JOL prompts are phrased. Forty participants studied glossaries and then made delayed cue-only and cue-target JOLs. In one condition, where the JOL prompts were phrased as predictions of future memory performance, only 15% of the participants preferred the cue-only strategy, replicating Jönsson and Kerimi (2011). In another condition, where JOLs were phrased as assessments of the current state of learning, 55% preferred the cue-only strategy. To conclude, students do not seem to recognize the value of cue-only JOLs, but they picked the superior JOL strategy more often when the JOL phrasing focused their attention on their knowledge state at the time of the JOL, rather than on a future state.

Keywords: metacognitive knowledge, learning strategies, judgments of learning, the delayed JOL effect, metacognition
Supplemental Materials: http://dx.doi.org/10.1037/arc0000002.supp
Data Repository: http://dx.doi.org/10.3886/ICPSR34645.v1

This article was published June 17, 2013.

Ivo Todorov, Department of Psychology, Stockholm University, Stockholm, Sweden; Nate Kornell, Department of Psychology, Williams College; Max Larsson Sundqvist, Department of Psychology, Stockholm University; Fredrik U. Jönsson, Department of Psychology, Stockholm University and Stockholm Brain Institute, Stockholm, Sweden.

This research was supported by grants from the Swedish Research Council (2006-1851 & 2009-2334) to Fredrik U. Jönsson. We are indebted to Andrea Gradin for help with the data collection.

For further discussion on this topic, please visit the *Archives of Scientific Psychology* online public forum at http://arcblog.apa.org.

Correspondence concerning this article should be addressed to Fredrik U. Jönsson, Department of Psychology, Stockholm University, S-106 91 Stockholm, Sweden. E-mail: fredrik.jonsson@psychology.su.se

One type of metacognitive knowledge is strategic knowledge, which according to Pintrich (2002) encompasses knowledge of the relative efficacy of general learning strategies; for example, that spaced repetition is better than massed repetition (Kornell et al., 2010), or the advantage of using imagery (Dunlosky & Hertzog, 2000; Hertzog et al., 2009) or associative elaboration (Presley, Levin, & Ghatala, 1984) over simple repetition. Strategic knowledge also includes thinking, problem solving and, of importance to the present study, awareness of *metacognitive strategies.* The latter refers to the idea that "students can have knowledge of various metacognitive strategies that will be useful to them in planning, *monitoring*, and regulating their learning, and thinking" (Pintrich, 2002, p. 220; emphasis added). In a recent publication issued by the U.S. Department of Education, addressing the issue of improving students' learning and teacher instructions, Pashler and colleagues (2007) included one very well-established metacognitive strategy in their list of recommendations, namely *the delayed cue-only Judgments of Learning (JOL) strategy*. It is a strategy students can use to enhance the accuracy in their monitoring of their own learning progress. The present research investigated students' knowledge of the value of this monitoring strategy. More precisely, the study aimed to answer two questions: First, do students know how to best monitor their own learning in that they recognize the value of delayed cue-only JOLs? Second, will they respond differently depending on how the JOL prompt is formulated? To our knowledge, only one previous study has investigated student's preference for the delayed cue-only JOL strategy (Jönsson & Kerimi, 2011), and no previous research has investigated whether this knowledge can be activated by framing the JOLs. Before turning to these two questions in detail, we briefly review prior research on the accuracy of delayed JOLs.

The Value of Delayed Cue-Only JOLs: How to Improve Monitoring Accuracy

JOLs are ratings of the probability of correctly recalling recently studied items on a future test. JOLs are thought to be a metacognitive tool for monitoring ongoing learning and are used for determining the extent to which the material that needs to be learned has indeed been learned (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991; Rhodes & Tauber, 2011b). The accuracy of metacognitive monitoring, including JOLs, is assumed to be closely related to decisions students make about studying, such as study time allocation, item-selection for restudy, and decisions to terminate the study process (Dunlosky & Hertzog, 2000; Metcalfe, 2002; Metcalfe & Dunlosky, 2008; Metcalfe & Finn, 2008; Nelson & Dunlosky, 1991; Thiede, Anderson, & Therriault, 2003; but see Kimball, Smith, & Muntean, 2012). These study decisions can impact subsequent learning. Thus, it is important to be able to accurately monitor ones' learning. Faulty JOLs may lead to faulty study decisions and suboptimal learning, while accurate JOLs can have the opposite effect.

In a paired associate learning paradigm, Nelson and Dunlosky (1991) demonstrated that JOLs made with a slight delay after study are highly accurate in terms of discriminating between well- and not so well-learned items. However, a prerequisite for this high JOL accuracy to occur is that only the cue word is shown while making JOLs (i.e., the first word in a word-pair). When JOLs are delayed but both the cue and target words are shown to the participant, the JOL accuracy is considerably lower (Dunlosky & Nelson, 1992). According to the monitoring-dual-memories hypothesis, delayed cue-only JOLs tend to be accurate because they allow for an attempt to retrieve a memory from long-term memory before the JOL is made (Dunlosky & Nelson, 1992). The high JOL accuracy of delayed cue-only JOLs, as opposed to delayed cue-target JOLs, or JOLs made directly after study, is well established and has shown to be robust across studies (e.g., Connor, Dunlosky, & Hertzog, 1997; Dunlosky & Nelson, 1992, 1994; Kelemen & Weaver, 1997; Kimball & Metcalfe, 2003; Koriat & Ma'ayan, 2005; Meeter & Nelson, 2003; Nelson & Dunlosky, 1991; Thiede & Dunlosky, 1994; Weaver & Kelemen, 1997; although the effect diminishes when feedback is provided, Kornell & Rhodes, in press; see Metcalfe & Dunlosky, 2008; Rhodes & Tauber, 2011b; Sikström & Jönsson, 2005, for reviews). But do students know about the superiority of this metacognitive strategy for monitoring their learning?

Do Students Have Knowledge About the Value of Delayed Cue-Only JOLs?

One recent study by Jönsson and Kerimi (2011) suggests that most students do not know the best way of assessing their ongoing learning. Immediately after studying an item, the participants were asked to select one of the four monitoring strategies used by Dunlosky and Nelson (1992). The available strategy choices were between delaying the JOL or making it right away (immediate vs. delayed JOLs), as well as deciding whether both cue and target, or only the cue, should be displayed in the JOL prompt. The participants were asked to select the monitoring strategy that would allow them make the most accurate prediction of later memory performance. The preference for the delayed cue-only JOL strategy ranged from 25% to 34% across two experiments, and the participants did not reliably update their strategy choices on the basis of task experience. However, the strategy choice interacted with item difficulty. When the participants were asked about the best strategy *after* the end of the second experiment, a question that should not be affected by item characteristics, 55% of them preferred the delayed cue-only JOL strategy. To conclude, even after experience with the task, about half of the students did not seem aware that to improve monitoring accuracy, they should use the delayed cue-only JOL strategy. The results also hint on something else, which the present study tests; namely, that the participants more often prefer the better monitoring strategy when their attention is not directed toward learning the items.

Because Jönsson and Kerimi let the students self-regulate their choice of JOL strategy, not all students experienced all four strategies, which may have altered their JOL strategy preference. Therefore, in the present study the students first experienced both one inferior (delayed cue-target JOL; dJOLct) and one superior JOL strategy (delayed cue-only JOL; dJOLc), before being asked about which one leads to more accurate judgments, thus guaranteeing equal exposure to both strategies. Furthermore, to make the choice easier, we only included two delayed JOL strategies (i.e., no immediate JOLs). In sum, the present study aimed to further investigate the Jönsson and Kerimi (2011) findings, but with another methodology.

Will Students Respond Differently Depending on How the JOL Prompt Is Formulated?

The present study also investigated whether JOL strategy preference changes as a function of how the JOL prompt and associated task instructions are formulated. It is well established that how a question or text is formulated (i.e., phrased or framed) can affect how we respond to the question, make choices, or behave in a task (e.g., Finn, 2008; Halkjelsvik, Jørgensen, & Teigen, 2011; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2006; Koriat, Bjork, Sheffer, & Bar, 2004; Kühberger, 1998; Schkade & Kahneman, 1998; Tversky & Kahneman, 1981). An excellent example of this is the set of experiments by Koriat, Bjork, Sheffer, and Bar (2004), where participants rated their certainty that they would be able to recall paired associates after various retention intervals. Although people almost universally know that we forget over time, Koriat and colleagues demonstrated that we do not apply this belief about forgetting when making JOLs unless it is explicitly brought to our attention. When they rephrased the JOL prompt in terms of forgetting rather than remembering, or if the retention interval was manipulated within participants (forcing a comparison between the different retention intervals), rather than between participants, the JOLs were lower for longer retention intervals. When making JOLs, participants tend to ignore information like the retention interval in favor of information more immediately experienced when making the JOLs (e.g., perceived item difficulty or retrieval fluency; Koriat, 1997). To conclude, although most of us, if not all, do know that we forget over time, Koriat and colleagues demonstrated that we do not apply our beliefs about forgetting when making JOLs unless it is explicitly brought to our attention.

Judgments and decision making researchers have long attended the phenomena that if logically equivalent choice situations are differently described, it may result in a shift in the choice preferences (Kühberger, 1998; Tversky & Kahneman, 1981). A classic example is the seminal study by Tversky and Kahneman (1981), where the participants were to choose one of two different ways to contain a disease outbreak that would claim 600 lives unless nothing was done. In one condition, the majority of the participants preferred to surely save 200 lives over a one-third probability of saving all 600 and two-thirds probability of saving no one. In another condition, the majority of the participants were instead more prone to choose the option that with one-third probability nobody would die, but with two-thirds probability all 600 would die, over the alternative that 400 would certainly die. Although the objective probabilities are identical in terms of lives saved or deaths between the two versions of the problems, peoples' preferences changed depending on how the problem was formulated (i.e., framed); that is, whether the participants attention was on deaths or on saving lives. This framing effect on people's choices between alternatives has been extensively replicated with a small to average effect size across studies (Kühberger, 1997, 1998).

In the present experiment, we investigated whether the framing of the JOL prompt (and additional instructions) had an effect on which monitoring strategy participants judged to be optimal. More precisely, the phrasing manipulation in the present study aimed to elucidate whether it was possible to activate students' metacognitive knowledge of the relative efficacy of different metacognitive strategies. We know of no such previous study. Jönsson and Kerimi (2011) found no evidence that participants understood how efficient delayed cue-only JOLs are in improving monitoring accuracy, but, as reviewed earlier, such knowledge does not always affect behavior unless it is made salient, so it is possible that Jönsson and Kerimi (2011) underestimated their participants' actual knowledge. If people do understand that JOLs are more accurate when made after a delay, we predict that they will bring this knowledge to bear on their judgments more often when the task is framed more clearly toward the monitoring situation, compared to when the task focuses on future memory performance.

The standard way of phrasing the JOL prompt in the metacognitive literature is in terms of a prediction about later memory performance (i.e., what is the probability that you will remember this item on a later memory test? e.g., Dunlosky & Nelson, 1992; Jönsson & Kerimi, 2011; Nelson & Dunlosky, 1991). The fact that the JOL prompt (and other instructions) highlight later memory performance may focus the participant's attention on the memorization of the items in preparation for the delayed test, rather than on how to best monitor learning. We hypothesized that if the JOL prompt instead were about the learning state at the time of the judgment (i.e., how well have you learned this item now?), rather than about a future memory state, it would shift attention away from memorization and toward monitoring accuracy. If they have accurate beliefs, this shift should lead them to discriminate between the better and the worse monitoring strategy.

To test this hypothesis, all participants first experienced both an inferior strategy (delayed cued-target JOLs) and a superior strategy (delayed cue-only JOLs), and then had to choose which of the two had let them make the most accurate JOLs or if they were equally good. One group received the standard JOL prompt (the prediction group), whereas the other group instead judged how well they had learned each item without having to predict future memory performance (current learning group). The instructions and the final questions about which strategy was better were also adapted to fit the phrasing of the JOL prompt. We expected the prediction group to pay relatively more attention to what improves memory, and therefore to prefer the delayed cue-target JOL strategy. By contrast, we expected the current learning group to focus more on metacognitive accuracy, and thus to select the strategy that produces the most accurate judgments. In sum, to the degree that participants understand the benefits of delayed testing for JOL accuracy, we expected them to prefer delayed cue-only JOLs, but more so when their attention is directed toward the monitoring situation instead of a future memory test.

Method

Participants

Forty students (31 women; mean age = 25.85 years; *SD* = 9.42) from Stockholm University participated in the experiment for course credits or a movie voucher. An informed consent form was signed before the experiment started.

Materials

Sixty Swedish word-pairs from the association norms of Shaps, Johansson, and Nilsson (1976) were divided into two separate lists with 30 words in each (e.g., mother - home). The mean association value and *SD* were identical for the two lists. In addition, all word-pairs had a similar association value (varying from one to three). The experiment was run on PC computers with E-prime 2.0 software (Psychological Software Tools, Pittsburgh, PA).

Design

The delayed cue-only JOL strategy is henceforth denoted as dJOLc and the delayed cue-target JOL strategy as dJOLct. A 2 × 2 mixed factorial design was used, with JOL strategy (dJOLc/dJOLct) as a within-subjects variable and phrasing (prediction/current learning) as a between-subjects variable. The dependent variables are stated separately in each analysis below. For one group (n = 20), henceforth referred to as the *current learning* group, both the instructions, the JOL prompt, and the questions about JOL strategies were formulated in terms of an assessment of the current state of learning. For the other group (n = 20), henceforth referred to as the *prediction* group, the instructions, the questions about JOL strategies and the JOL prompt were instead formulated in terms of a prediction of later memory performance, because it is typically formulated in metamemory experiments.

Procedure

Except for the phrasing manipulation, the procedure was the same for both groups. After they read the instructions, participants studied one of the two stimulus lists, made JOLs for each word-pair, answered a question about the accuracy of their JOLs (Question 1), made a global JOL (Question 2), and then took a test. Then the procedure was repeated with the second stimulus list.

The participants started with either the dJOLc or the dJOLct condition, and the order was counterbalanced. Finally, after having finished both the dJOLc and dJOLct conditions, the participants answered two questions; one about which monitoring strategy they thought was better (Question 3), and why (Question 4). The procedure will now be outlined in more detail.

The dJOLc and dJOLct conditions. Participants first studied 30 word-pairs (i.e., one of the two item lists; see the materials section), one at a time, with an interstimulus presentation interval of 500 ms (a blank screen). Following an unrelated filler task for 45 s, during which they solved mathematical equations, they were again presented with all items, one at a time, and made their JOLs. In the dJOLct condition the participants made their JOLs while viewing the whole word-pair, whereas in the dJOLc condition they were presented with the cue word only. The prediction group was asked "How sure are you, that you, in few minutes, will remember the target word if shown the cue word?", whereas the current learning group responded to "How well have you learned the association between the cue and target words?" In both cases, the rating was given on a percentage scale from 0% to 100% with increments of 20%, where a higher percentage reflected higher confidence.

Immediately after the last JOL trial the participants answered two questions (Question 1 and 2). In the first question they responded to "How well do you think your judgments will correspond with your actual performance on an upcoming memory test?", by giving an answer on a 7-graded ordinal scale (1 = *not at all*; 7 = *very well*). The scale was explained by giving examples for high, average and low correspondence. The second question was a global JOL, where the participants judged how many of the 30 target words would be remembered on the upcoming memory test by entering a number between 0 and 30 on the keyboard.

Question 1 and 2 were then followed by a 90-s filler activity, where the participants again solved different math equations, followed by a cued-recall test for the 30 recently studied items. For each item, while presented with the cue, the participants had to type in the target-word on the keyboard within 15 s. To advance to the next trial, they either pressed ENTER, or waited the full 15 s.

Following a 45-s filler activity, where they again solved math equations, this same procedure was repeated once more, but with 30 new word pairs (i.e., the other stimulus list). Hence, Question 1 and 2 were given twice; one time in the dJOLc condition and one time in the dJOLct condition. The item presentation order was uniquely randomized for every participant and every list (i.e., during study, then again when they made their JOLs, and at the cued-recall test).

Final questions 3 and 4. After having finished both the dJOLc and dJOLct conditions the participants were informed that: "You have now studied, judged and tested your memory on two different glossaries, and in two different ways. In both cases the judgment was made after a short delay. One of the judgments was made while you were presented with only the cue (e.g., tired-), whereas the second was made while you were presented with the cue and the target (e.g., tired—worn)". They were then asked to judge which of the two strategies is the better one (Question 3). The prediction group responded to "Which of these two ways of judging let you make the most accurate estimation of how well you actually performed on the later memory test? Or are they equally good*?*", whereas the current learning group responded to "Which of these two ways of judging let you make the most accurate judgment of how well you have learned the association between the cue and the target? Or are they equally good?" Both groups responded by entering 1, 2, or 3 on the keyboard (1 = *When you were presented with both the cue and the target*; 2 = *When you were only presented with the cue*; 3 = *Both strategies are equally good*). Finally, the participants were asked to briefly motivate their response to Question 3 on a piece of paper. This we refer to as Question 4. On average, the experiment lasted about 36 min (*SD* = 9 min), including instructions.

Results

In all analyses, we used an alpha level of .05, and missing values, if any, were handled by using case-wise deletion. Effect sizes are denoted by partial eta squared (η_p^2) for analyses of variance (ANOVAs) and Cohen's *d* for *t*-tests. We first present the analyses of most importance to the aims of the present study (Question 1, 3 and 4), followed by secondary analyses that may still be of interest to the reader.

Do Students Have Knowledge About the Value of Delayed Cue-Only JOLs?

Knowledge of the value of cue-only JOLs was measured in three ways. Immediately after the JOL phase, and before the cued-recall test, the participants rated on a seven-point scale how well their judgments would correspond with actual cued recall performance (Question 1). After finishing both the dJOLc and dJOLct tasks, they were asked about which of the judgment strategies let them make optimal judgments (Question 3). Finally, after having responded to Question 3, they were asked to write down why they thought the strategy they had chosen was the better one or why the two strategies were equally good (depending on their choice).

Question 1. There was no main effect of Phrasing, no main effect of JOL strategy, nor any Phrasing x JOL strategy interaction. However, as evident from Table 1, in the current learning condition the dJOLc strategy was rated as nominally better (i.e., higher correspondence between the JOLs and later memory performance) than the dJOLct strategy, with an average effect size (d = .55).

Question 3. Figure 1 shows the proportion of participants who chose the dJOLc, dJOLct, or Equal alternative as a function of Phrasing. Because only one response was correct (i.e., the dJOLc was the superior monitoring strategy for 36 participants and it was inferior for none, as indicated by the relative JOL accuracy analysis below), the answers were scored dichotomously in terms of accuracy (0 = *wrong*; 1 = *correct*). A significantly larger proportion of participants responded that the dJOLc strategy was better (55%) in the current

Table 1
Means (SD) for the Dependent Variables of the Experiment as a Function of Phrasing and Judgments of Learning (JOL) Strategy

	JOL strategy			
	Current learning		Prediction	
Dependent measure	dJOLc	dJOLct	dJOLc	dJOLct
Global JOL	.47 (.21)	.50 (.13)	.41 (.11)	.39 (.10)
Mean item-by-item JOL	.42 (.11)	.67 (.17)	.38 (.11)	.64 (.18)
Cued Recall	.33 (.09)	.84 (.16)	.30 (.09)	.83 (.16)
Absolute JOL accuracy	.10 (.07)	−.16 (.24)	.08 (.09)	−.19 (.20)
Relative JOL accuracy	.96 (.06)	.01 (.59)	.97 (.04)	.28 (.61)
Question 1 rating	5.25 (1.12)	4.65 (1.04)	4.50 (1.61)	4.60 (1.19)

Note. dJOLc = delayed cue-only JOL; dJOLct = delayed cue-target JOL. Absolute JOL accuracy was calculated by subtracting cued recall from mean JOL ratings, and relative accuracy constitutes the Goodman-Kruskal gamma correlation between the JOLs and cued recall performance.

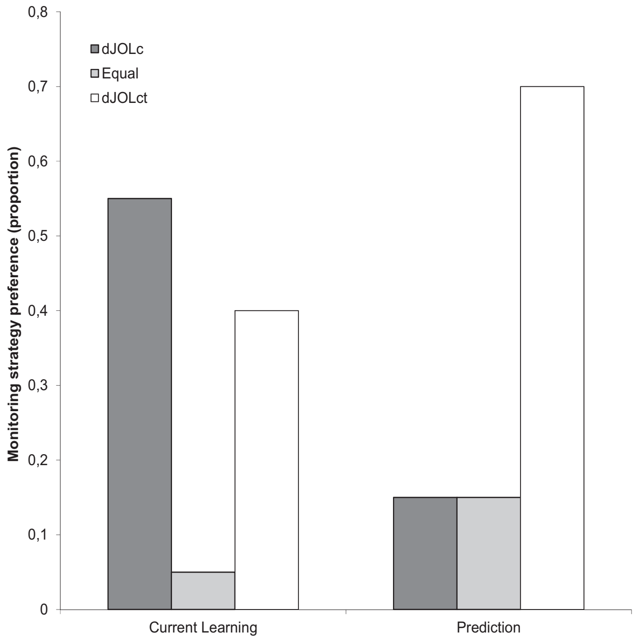

Figure 1. The proportion of participants that preferred the delayed cue-only judgment of learning (dJOLc), delayed cue-target judgment of learning (dJOLct), or equal alternative as a function of Phrasing, in response to Question 3.

learning group than in the prediction group (15%), as shown by a Mann–Whitney U test (U = 120.00, p = .01).

Question 4. All participants were asked to elaborate on their response to Question 3 (Figure 1) in a free-response format. Three independent judges scored their responses as either (1) reasoning clearly in terms of the dJOLc strategy, or showing explicit knowledge of the superiority of the dJOLc strategy, (2) reasoning about their choice of strategy, but without explicitly stating that dJOLc is better (e.g., that dJOLc and dJOLct are equally good), (3) motivating their choice solely in terms of what improves memory, or (4) unclear or ambivalent elaboration that could not be rated. We used the Kappa statistic (Landis & Koch, 1977) to measure interrater consistency and it was moderate (*Kappa* = .60, p < .001). The responses that were scored differently by any of the three raters were discussed until consensus was reached. The responses were then rescored as either accurate (1; response category 1) or inaccurate (0; response category 2–4 were collapsed). A Mann–Whitney U test showed that significantly more participants argued that the dJOLc strategy was superior in the current learning group (47.5%), compared to those in the prediction group (15%), U = 140.00, p = .04. These results closely mirror, and hence corroborate, those of Question 3. Furthermore, when the JOL question was formulated as a prediction, 55% of the participants answered Question 4 only in terms of what improves memory, compared to 30% in the current learning group, U = 130.00, p = .03.

The Superior Monitoring Accuracy of Delayed Cue-Only JOLs

To test whether the dJOLc condition was associated with better monitoring compared to the dJOLct condition, also in the present study, we first calculated a Goodman Kruskal gamma correlation for each participant and condition. For two participants the dJOLct gamma correlation could not be calculated due to recall being constant (but both their dJOLc gammas were > .90). For 36 of the remaining 38 participants the dJOLc gammas were nominally higher than in the dJOLct condition, and the two gammas were identical for the other two participants.[1] The ANOVA showed a main effect of JOL strategy, $F(1, 36) = 70.59$, $p < .001$, $\eta_p^2 = .66$. The dJOLc strategy was associated with a significantly higher correlation (M = .97, SD = .05), than the dJOLct strategy (M = .14, SD = .61). This underscores the robustness of the effect (e.g., Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991), and clearly demonstrates the superiority of the dJOLc strategy in this experiment as well. There was no main effect of Phrasing and no significant Phrasing × JOL strategy interaction.

Secondary Analyses

Mean item-by-item JOL. The mean item-by-item JOL rating was calculated for every participant and entered into the ANOVA as the dependent variable. A significant main effect of JOL strategy showed that the participants made reliably higher JOLs in the dJOLct (M = .66, SD = .17) than in the dJOLc condition (M = .40, SD = .11), $F(1, 38) = 88.79$, $p < .001$, $\eta_p^2 = .70$. There was no main effect of Phrasing, nor any Phrasing × JOL strategy interaction (Table 1).

Cued recall. There was a significant main effect of JOL strategy, $F(1, 38) = 570{,}70$, $p < .001$, $\eta_p^2 = .94$. The dJOLct strategy was associated with significantly higher recall levels (M = .83, SD = .16), than the dJOLc strategy (M = .31, SD = .09). There was no main effect of Phrasing, nor any Phrasing × JOL strategy interaction (Table 1).

Absolute JOL accuracy. Absolute accuracy was calculated by subtracting the mean cued recall performance from the mean item-by-item JOL ratings, individually for every participant. The dJOLc strategy was associated with overconfidence (M = .09, SD = .08), whereas the dJOLct strategy was associated with underconfidence (M = −.17, SD = .22), as shown by a significant main effect of JOL strategy $F(1, 38) = 52.15$, $p < .001$, $\eta_p^2 = .58$. Please note that a value of zero would mean that the JOLs perfectly matched the memory performance. There was no main effect of Phrasing, nor was there any Phrasing × JOL strategy interaction (Table 1).

Global JOLs. The answers to Question 2, a value from zero to 30, were transformed to proportions, before entering them into the ANOVA. There was a main effect of Phrasing, $F(1, 38) = 4.40$, p = .04, $\eta_p^2 = .10$. The global JOLs were higher in the current learning condition (M = .48, SD = .13), than in the prediction condition (M = .40, SD = .13). There was no main effect of JOL strategy, nor any Phrasing × JOL strategy interaction.

We also analyzed the difference between the global JOLs and subsequent cued recall. This ANOVA showed a main effect of JOL strategy, $F(1, 38) = 282.76$, $p < .001$, $\eta_p^2 = .88$, with overconfidence in the dJOLc condition (M = .12, SD = .20) and underconfidence in the dJOLct condition (M = −.39, SD = .22). The pattern of results follows those of the item-by-item JOLs, but with a more severe underconfidence in the dJOLct condition. There was no significant main effect of Phrasing, nor any Phrasing × JOL strategy interaction.

Discussion

The present experiment investigated students' knowledge about the value of delayed cue-only JOLs. A minority of participants accurately judged the delayed cue-only JOL strategy to produce the greatest monitoring accuracy. However, the phrasing of the JOL prompt (and associated instructions) affected students' JOL strategy preference.

[1] It is noteworthy that in the dJOLc Condition 37/40 correlations were >.90, and the remaining three correlations were >.74.

We first discuss the prediction condition, which corresponds to previous JOL research. The superiority of the dJOLc strategy over the dJOLct strategy in terms of monitoring accuracy is well documented (Dunlosky & Nelson, 1992; Nelson & Dunlosky, 1991; see Rhodes & Tauber, 2011b for a review). Indeed, in the present study the effect replicated for nearly all participants. Despite this robust effect, most participants preferred the inferior dJOLct strategy, with only 15% stating that the dJOLc strategy leads to better monitoring (Question 3), and about half reported having made their response solely in terms of what improves memory (Question 4; cf., Jönsson & Kerimi, 2011). In line with the participants' beliefs, in the present study the dJOLct strategy led to much better memory performance in the end, despite monitoring being much worse. The instructions and Question 3 were clearly about what leads to better memory monitoring, however, not better memory. Hence, the present study replicates the previous findings by Jönsson and Kerimi (2011) that most students do not seem to realize the value of cue-only JOLs in improving monitoring accuracy, even when they only had to select between two JOL strategies rather than four.

This outcome could be symptomatic of noncompliance with the instructions (i.e., selecting the most effective strategy for learning, despite being asked to select the most effective strategy for metacognitive accuracy). However, Jönsson and Kerimi (2011) found that self-reported noncompliance did not alter their results. Rather, we suspect that the participants failed to distinguish between the concepts of monitoring and memory. That is, participants were asked to make judgments about which strategy produced the best metacognitive accuracy, but may not have seen this question as different from asking which strategy produced the most learning. The latter conclusion is corroborated by the Jönsson and Kerimi (2011) finding that the participants more often chose to reexpose both the cue and target, compared to cue-only, when the items were difficult, and vice versa for the easier items, indicating the employment of a learning strategy rather than a monitoring strategy. An alternate explanation, that participants actually thought the dJOLct strategy led to better monitoring, is contradicted by the Question 4 responses.

However, as reasoned in the introduction, phrasing JOLs in terms of the upcoming memory test may have led the participants to focus their attention on optimizing recall performance. Previous research has demonstrated that the framing of a question or task may affect how people respond to it (e.g., Halkjelsvik et al., 2011; Kahneman et al., 2006; Schkade & Kahneman, 1998; Tversky & Kahneman, 1981). The current learning manipulation was supposed to shift focus away from recall accuracy and toward monitoring accuracy, and this was exactly what we found. Students in the current learning group reliably more often selected (Question 3) and motivated their choice (Question 4) in terms of delayed cue-only JOLs, over the dJOLct alternative, as the optimal monitoring strategy. However, even with this manipulation, only about half of the participants seemed to be aware of the best monitoring strategy.

It should be noted that the result of the framing (phrasing) manipulation was not as clear for Question 1 as for the other questions (i.e., the question answered immediately after each JOL phase about how well their judgments would correspond with their later cued recall performance). Nevertheless, although not statistically reliable, the data pattern was in line with the Question 3 and 4 findings. As evident in Table 1, the participants in the current learning group nominally rated the dJOLc strategy as a better monitoring strategy than the dJOLct strategy, which was clearly not the case in the prediction group. This is on par with the Question 3 responses, where the participants more often preferred the dJOLc strategy in the current learning, compared to the prediction group.

Although the present experiment suggests that people are not aware of the metacognitive value of delayed cue-only JOLs, other findings suggest the opposite. Kornell and Bjork (2007) asked 452 UCLA undergraduates if they test themselves while studying, and if so, why. Although self-testing is a highly effective way to learn, the majority of participants (68%) reported testing themselves for a different reason: to figure out what they did and did not know (i.e., to make diagnostic "current learning" JOLs; also see Hartwig & Dunlosky, 2012; Kornell & Son, 2009). This result, which suggests a strong motivation to use tests as a metacognitive aid, seems at odds with the present findings, and those of Jönsson and Kerimi (2011), which showed that people do not seem to recognize the benefit of tests as a metacognitive aid. Yet these results may be reconcilable. It is possible that people like testing themselves if they get feedback afterward, telling them whether they were right or wrong. When they are tested without feedback (as in this and most other JOL studies), however, they do not know for sure whether their answers were right or wrong. In other words, their JOLs tend to be very accurate but this fact is not necessarily *apparent* to the studier without feedback. Thus, it is possible that people would be more likely to prefer cue-only JOLs (i.e., JOLs allowing them to make a diagnostic attempt to retrieve information from long-term memory) if they were given feedback after the retrieval attempt (i.e., if they made a delayed cue-only JOL with feedback; see Kornell & Rhodes, in press; Rhodes & Tauber, 2011a). Furthermore, whereas the present experiment and the Jönsson and Kerimi (2011) experiments answer the question of whether participants *know* that to improve their metamemory monitoring they should spontaneously pick the alternative that allows for diagnostic memory testing, the Kornell and Bjork (2007) study specifically inquired about why students self-test, *when* they do so. Though similar, the two tasks are not fully comparable.

To conclude, we replicated the previous finding (Jönsson & Kerimi, 2011) that most students are unaware of the superiority of the dJOLc monitoring strategy over the dJOLct strategy. However, the clear effect of framing on the degree of metacognitive knowledge shown by the students indicates that their responses are more accurate if their focus is shifted toward monitoring. We suspect that students do not generally think in terms of memory monitoring when they study, but rather focus on learning. This does not mean that they do not monitor their learning, nor does it imply that they do not regulate their study efforts on the basis of such monitoring; it only suggests that they may not have a naïve and explicitly available idea about the importance of monitoring for being a successful learner (in other words, they may engage in metacognitive monitoring and control without possessing sophisticated metacognitive knowledge). Hence, when monitoring their learning, students may not always use the most efficient monitoring strategy, or they may sacrifice accurate monitoring to achieve more learning. Future research should focus on how and to what extent students implement their metacognitive knowledge and whether possessing accurate metacognitive knowledge translates into increased learning efficiency.

As Schwartz and Efklides (2012) and others have noted, it may be valuable to teach students about efficient learning strategies. The present study clearly demonstrates that this does not only include memory strategies (see Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013, for a review) but also metacognitive strategies like assessing one's learning by taking a delayed test.

References

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12,* 50–71. doi:10.1037/0882-7974.12.1.50

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15,* 462–474. doi:10.1037/0882-7974.15.3.462

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

Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language, 33,* 545–565. doi:10.1006/jmla.1994.1026

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14,* 4–58. doi:10.1177/1529100612453266

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

Halkjelsvik, T., Jørgensen, M., & Teigen, K. H. (2011). To read two pages, I need 5 minutes, but give me 5 minutes and I will read four: How to change productivity estimates by inverting the question. *Applied Cognitive Psychology, 25,* 314–323. doi:10.1002/acp.1693

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19,* 126–134 doi:10.3758/s13423-011-0181-y

Hertzog, C., Price, J., Burpee, A., Frentzel, W. J., Feldstein, S., & Dunlosky, J. (2009). Why do people show minimal knowledge updating with task experience: Inferential deficit or experimental artifact? *Quarterly Journal of Experimental Psychology, 62,* 155–173. doi:10.1080/17470210701855520

Jönsson, F., & Kerimi, N. (2011). An investigation of students knowledge of the delayed judgments of learning effect. *Journal of Cognitive Psychology, 23,* 358–373. doi:10.1080/20445911.2011.518371

Kahneman, D., Krueger, A. B., Schkade, D., Schwarz, N., & Stone, A. A. (2006). Would you be happier if you were richer? A focusing illusion. *Science, 312,* 1908–1910. doi:10.1126/science.1129688

Kelemen, W. L., & Weaver, C. A. (1997). Enhanced memory at delays: Why do judgments of learning improve over time? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1394–1409. doi:10.1037/0278-7393.23.6.1394

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition, 31,* 918–929. doi:10.3758/BF03196445

Kimball, D. R., Smith, T. A., & Muntean, W. J. (2012). Does delaying judgments of learning really improve the efficacy of study decisions? Not so much. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 923–954. doi:10.1037/a0026936

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., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133,* 643–656. doi:10.1037/0096-3445.133.4.643

Koriat, A., & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language Special Issue: Metamemory, 52,* 478–492.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14,* 219–224. doi:10.3758/BF03194055

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging, 25,* 498–503. doi:10.1037/a0017807

Kornell, N., & Rhodes, M. G. (in press). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied.*

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17,* 493–501. doi:10.1080/09658210902832915

Kühberger, A. (1997). Theoretical conceptions of framing effects in risky decisions. In R. Ranyard, W. R. Crozier, & O. Svenson (Eds.), *Decision making: Cognitive models and explanations* (pp. 128–144). New York, NY: Routledge.

Kühberger, A. (1998). The influence of framing on risky decisions: A meta-analysis. *Organizational Behavior and Human Decision Processes, 75,* 23–55. doi:10.1006/obhd.1998.2781

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,* 159–174. doi:10.2307/2529310

Meeter, M., & Nelson, T. O. (2003). Multiple study trials and judgments of learning. *Acta Psychologica, 113,* 123–132. doi:10.1016/S0001-6918(03)00023-4

Metcalfe, J. (2002). Is study time allocated selectively to a Region of Proximal Learning? *Journal of Experimental Psychology: General, 131,* 349–363. doi:10.1037/0096-3445.131.3.349

Metcalfe, J., & Dunlosky, J. (2008). Metamemory. In H. L. Roediger, III (Ed.), *Learning and memory: A comprehensive reference* (pp. 349–362). Oxford, United Kingdom: Elsevier. doi:10.1016/B978-012370509-9.00159-5

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

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

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (NCER 2007–2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ncer.Ed.gov

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice, 41,* 219–225. doi:10.1207/s15430421tip4104_3

Pressley, M., Levin, J. R., & Ghatala, E. S. (1984). Memory strategy monitoring in adults and children. *Journal of Verbal Learning & Verbal Behavior, 23,* 270–288. doi:10.1016/S0022-5371(84)90181-6

Rhodes, M. G., & Tauber, S. K. (2011a). Eliminating the delayed JOL effect: The influence of the veracity of retrieved information on metacognitive accuracy. *Memory, 19,* 853–870. doi:10.1080/09658211.2011.613841

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

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9,* 340–346. doi:10.1111/1467-9280.00066

Schwartz, B. L., & Efklides, A. (2012). Metamemory and memory efficiency: Implications for student learning. *Journal of Applied Research in Memory and Cognition, 1,* 145–151. doi:10.1016/j.jarmac.2012.06.002

Shaps, L. P., Johansson, B., & Nilsson, L-G. (1976). *Swedish association norms.* (Report No. 196). Uppsala, Sweden: Uppsala University, Department of Psychology.

Sikström, S., & Jönsson, F. (2005). A model for stochastic drift in memory strength to account for judgments of learning. *Psychological Review, 112,* 932–950. doi:10.1037/0033-295X.112.4.932

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95,* 66–73. doi:10.1037/0022-0663.95.1.66

Thiede, K. W., & Dunlosky, J. (1994). Delaying students' metacognitive monitoring improves their accuracy in predicting their recognition performance. *Journal of Educational Psychology, 86,* 290–302. doi:10.1037/0022-0663.86.2.290

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211,* 453–458. doi:10.1126/science.7455683

Weaver, C. A., & Kelemen, W. L. (1997). Judgments of learning at delays: Shifts in response patterns or increased metamemory accuracy? *Psychological Science, 8,* 318–321. doi:10.1111/j.1467-9280.1997.tb00445.x

Received August 8, 2012
Revision received March 21, 2013
Accepted April 1, 2013 ■