THE CONTRIBUTION OF PAST TEST PERFORMANCE, NEW LEARNING, AND
FORGETING TO JUDGMENT-OF-LEARNING RESOLUTION

A thesis submitted
to Kent State University in partial
fulfillment of the requirements for the
degree of Master of Arts

By

Robert Ariel

August, 2010
Thesis written by
Robert Ariel
B.A., University of Illinois at Chicago, 2007
M.A., Kent State University, 2010

Approved by

John Dunlosky, PhD  Advisor
Maria Zaragoza , PhD  Chair, Department of Psychology
John R. D. Stalvey, PhD  Dean, College of Arts and Sciences
# TABLE OF CONTENTS

- LIST OF FIGURES ...................................................................................................................... iv
- LIST OF TABLES ........................................................................................................................ v
- ACKNOWLEDGMENTS ............................................................................................................. vi
- INTRODUCTION ....................................................................................................................... 1
  - METHODS - EXPERIMENT 1 ................................................................................................... 6
    - Participants .......................................................................................................................... 6
    - Materials and Procedure ................................................................................................... 6
    - Results ............................................................................................................................... 8
    - Discussion ......................................................................................................................... 16
- INTRODUCTION - EXPERIMENT 2 ......................................................................................... 19
  - METHODS ............................................................................................................................ 20
    - Participants ........................................................................................................................ 20
    - Materials and Procedure .................................................................................................. 20
    - Results ............................................................................................................................... 21
    - Discussion ......................................................................................................................... 27
- GENERAL DISCUSSION .......................................................................................................... 28
- REFERENCES ............................................................................................................................ 34
LIST OF FIGURES

FIGURES

1  Trial 2 resolution as a function of JOL group in Experiment 1.  iJOL = immediate JOLs; pJOL = prestudy JOLs. Prompt/no prompt = whether participants were prompted with prior recall outcome. Error bars represent standard error of the mean.

2  Trial 2 mean judgments of learning (JOL) in Experiment 1 conditionalized on trial 1 and trial 2 recall.  iJOL = immediate JOL; pJOL = prestudy JOL. Prompt/no prompt = whether participants were prompted with prior recall outcome. Error bars represent standard error of the mean.

3  Trial 2 mean resolution as a function of JOL group in Experiment 2.  iJOL = immediate JOLs; pJOL = prestudy JOLs. Prompt/no prompt = whether participants were prompted with prior recall outcome. Error bars represent standard error of the mean.

4  Trial 2 mean judgments of learning (JOL) in Experiment 2 conditionalized on trial 1 and trial 2 recall.  iJOL = immediate JOL; pJOL = prestudy JOL. Prompt/no prompt = whether participants were prompted with prior recall outcome. Error bars represent standard error of the mean.
LIST OF TABLES

TABLES

1 Decomposition of resolution (gamma correlations) presented in Figures 1 and 3.
ACKNOWLEDGMENTS

This project would not have been possible without the assistance, guidance, and support of several individuals. Specifically, I thank John Dunlosky for advising me on this project. I would also like to thank the members of my committee: Maria Zaragoza, William Merriman, and John Gunstad.
INTRODUCTION

Since Arbuckle and Cuddy’s (1969) seminal article, the accuracy of people’s judgments of learning (JOL) has been intensely scrutinized (Dunlosky & Metcalfe, 2009). JOLs are predictions of the likelihood of recalling recently studied items, and their accuracy is typically estimated using the following method. Learners study a list of paired associates (e.g., dog – spoon); immediately after studying a given pair, they judge the likelihood of recalling it on the upcoming test trial (i.e., dog - ?). Accuracy is then measured by comparing JOLs with recall performance: resolution refers to the degree to which JOLs discriminate between the recall of one item relative to another, whereas calibration refers to the degree to which the magnitude of judgments relates to the absolute level of performance. The effects of repeated study-test trials on JOL accuracy have been of major interest, partly because restudy trials can have both beneficial effects (increases in resolution, Koriat, 1997) and deleterious effects (decreases in calibration, Koriat, Sheffer, & Ma’ayan, 2002) on the accuracy of JOLs.

Though some progress has been made in understanding the decrease in calibration across trials (Finn & Metcalfe, 2007; Finn & Metcalfe, 2008; Scheck & Nelson, 2005; Serra & Dunlosky, 2005), explanations for the equally important increase in resolution across repeated study-test trials have not been systematically explored. Toward filling this gap, we empirically evaluate the degree to which three factors—memory for past test performance, new learning, and forgetting—contribute to the effects of repeated study-
test trials on JOL resolution. We describe these factors next and then discuss our approach to estimating their joint effects.

People’s JOLs are influenced by various factors (or cues), and the degree to which these factors are diagnostic of future test performance will influence JOL resolution (Koriat, 1997). During repeated study-test trials, one factor that is particularly diagnostic of future recall performance is prior test performance. People are extremely accurate at identifying which items they correctly recalled on a previous retrieval attempt (Gardiner & Klee, 1976), and incorporating memory of past test performance could potentially lead to higher accuracy in discriminating between items that will versus will not be recalled in the future. Several researchers have suggested that increases in resolution observed following test trials are a result of relying on past test performance (Finn & Metcalfe, 2008; King et al, 1980). For instance, Finn and Metcalfe (2007) examined the influence of memory for past test (MPT) on JOLs across two study-test trials. They found that JOLs made on trial 2 were more highly associated with trial 1 recall than with trial 2 recall. This finding suggests that after an initial test trial, people rely on their memory for performance on the previous test trial to make subsequent JOLs.

No published experiments have estimated the joint contribution of MPT and other potentially influential factors to trial 2 judgment resolution. In fact, researchers have almost exclusively investigated the effects of single factors on JOLs in isolation (for recent exceptions, see Benjamin, 2005; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Metcalfe & Finn, 2008). Two other factors that may contribute to improvements in resolution across trials are new learning and forgetting. That is, after an initial study-test
trial, JOL resolution may be influenced by (a) new learning for pairs that had not been previously recalled or (b) forgetting of previously recalled pairs before the next test.

Given that memory for past test performance is such a diagnostic cue, our questions are, Do people’s’ JOLs also predict new learning and forgetting when they are made on trial 2? That is, do these factors contribute to JOL resolution?

Finn and Metcalfe (2008) provide some evidence that people’s JOLs may in part predict new learning on trial 2. For items that were incorrectly recalled on trial 1, people gave higher JOLs to those items that were subsequently recalled on trial 2 than those that were not recalled. However, the difference in JOL magnitude between these classes of items was small, which suggests that the impact of new learning on JOLs may be negligible (Finn & Metcalfe, 2008). Furthermore, people may also incorporate cues that predict forgetting into their trial 2 judgments, which also could contribute to JOL resolution. Rawson et al (2002) provide some evidence that people’s predictions incorporate assessments about retention. Participants read brief texts and then made judgments either about their performance on a later test or about current comprehension of the text they read. Judgment magnitude was lower for performance predictions than comprehension judgments, suggesting that people incorporated information about potential forgetting into their judgments (see also Koriat et al., 2004). Most important, the relative contribution of MPT, new learning, and forgetting to JOL resolution has not yet been explored.

The goal of the present experiment was to estimate the joint contribution of these factors to JOL resolution following a study-test trial. To do so, we relied on both a recent
methodological advance (prestudy JOLs; Castel, 2008) and an analytical approach to decompose resolution (the prejudgment-recall-and-monitoring analysis; Nelson, Narens, & Dunlosky, 2004). During trial 1, participants studied paired associates (e.g., dog-spoon), made a JOL immediately after studying each pair and subsequently completed a cued-recall test (dog - ?). During a second study-test trial, participants made immediate JOLs (as on trial 1) or made JOLs immediately before studying each item (prestudy JOL). The purpose of the prestudy JOL was to eliminate item-specific cues at the time of judgment that people may use when making a JOL (Castel, 2008). Participants in the prestudy JOL group cannot evaluate their memory-for past-test performance, new learning, or forgetting when making their JOLs on trial 2. Thus, when making each JOL on trial 2, some participants were told whether they had previously recalled the response to the to-be-judged item. This prompt, when combined with prestudy JOLs, allowed us to examine the sole contribution of MPT to trial 2 JOLs, because the only information participants can use in this circumstance is how they performed on the previous test. If MPT is solely responsible for JOL resolution on trial 2, than participants in the prestudy JOL group who receive the recall prompt will make just as accurate JOLs as either of the immediate JOL groups.

To estimate the joint contribution of MPT, new learning, and forgetting to trial 2 resolution, we used the decomposition introduced by Nelson et al. (2004). Resolution on trial 2 was separated into three measures reflecting the contribution of each of the above factors. The contribution of MPT can be estimated by computing trial 2 resolution comparing items that were recalled on trial 1 to only items that were not recalled on trial
1. The contribution of new learning can be estimated by computing trial 2 resolution for only items not recalled on trial 1, and the contribution of forgetting can be estimated by computing resolution for only items recalled on trial 1. Given that these three estimates completely account for trial 2 resolution, the size of the estimates indicates the relative contribution of each factor to JOL resolution.
METHOD - EXPERIMENT 1

Participants

One hundred and twenty-five students from KSU participated for either course credit in Introductory Psychology or for $10. A 2 (JOL type: immediate or prestudy) X 2 (Recall prompt: prompt about previous recall performance or no prompt, as described below) full-factorial design was used. Participants were randomly assigned to either the immediate JOL group without prompts (henceforth, \textit{iJOL no-prompt} group, total \textit{n} = 31; paid \textit{n} = 13), the immediate JOL group with prompts (\textit{iJOL prompt} group, total \textit{n} = 32; paid \textit{n} = 12), the prestudy JOL group without prompts (\textit{pJOL no-prompt} group, total \textit{n} = 31; paid \textit{n} = 12), or the prestudy JOL group with prompts (\textit{pJOL prompt} group, total \textit{n} = 31; paid \textit{n} = 14).

Materials and Procedure

Sixty unrelated noun-noun paired associates were used in this experiment (e.g., \textit{icebox-acrobat}). Participants completed the experiment individually on a computer. They were instructed that their goal was to study word pairs, make JOLs, and complete a paired-associate recall test. Participants completed this study-judge-test cycle twice for the same pairs. The order of presenting pairs was randomized on each trial.

Trial 1 was identical for all participants: They studied items individually for 6 seconds. Immediately after studying an item, it was replaced with this prompt for a JOL: “For the pair you just studied, what is the likelihood that you will be able to recall the
second word when later presented with the first word during the upcoming test?"

Participants typed any value between 0 (I definitely won’t be able to recall this item) and 100 (I definitely will recall this item). After studying all items, participants completed a paired associate recall test for all pairs.

During trial 2, participants restudied the same 60 word pairs (6 s/pair) and again made JOLs. Participants either made JOLs immediately after studying an item (as in trial 1) or prior to studying the item (prestudy JOLs). Participants in the iJOL groups were given the same JOL prompt as on trial 1. Prestudy JOLs were obtained using the following prompt (Castel, 2008): “For the pair you are about to study, what is the likelihood that you will be able to recall the second word when later presented with the first word during the upcoming test?” The same scale was used for all JOLs. Also, immediately prior to making either JOL, some participants also were told whether they correctly recalled the current item using this recall prompt: “When tested, you correctly recalled (or did not correctly recall) the response to the pair that you just studied (or that you are about to study).” After participants finished studying the word pairs, a final test was administered.
RESULTS

Recall and Judgments of learning

The proportion of items correctly recalled did not differ between the iJOL no-prompt group (Trial 1: $M = .21$, $SE = .02$; Trial 2: $M = .51$, $SE = .04$), the iJOL prompt group (Trial 1: $M = .24$, $SE = .03$; Trial 2: $M = .54$, $SE = .04$), the pJOL no-prompt group (Trial 1: $M = .24$, $SE = .02$; Trial 2: $M = .54$, $SE = .04$), and the pJOL prompt group (Trial 1: $M = .21$, $SE = .02$; Trial 2: $M = .53$, $SE = .04$) on either trial 1, $F(3,125) = .65$, $MSE = .01$, $p = .59$, or trial 2, $F(3,125) = .13$, $MSE = .01$, $p = .94$.

Means across participants’ mean JOLs did not differ for the iJOL no-prompt group (Trial 1: $M = 35.3$, $SE = 2.94$; Trial 2: $M = 41.3$, $SE = 3.46$), the iJOL prompt group (Trial 1: $M = 41.9$, $SE = 2.69$; Trial 2: $M = 39.8$, $SE = 2.74$), the pJOL no-prompt group (Trial 1: $M = 40.6$, $SE = 2.71$; Trial 2: $M = 40.1$, $SE = 3.24$), and the pJOL prompt group (Trial 1: $M = 36.3$, $SE = 3.60$; Trial 2: $M = 38.6$, $SE = 3.48$) on either trial 1, $F(3,125) = 1.16$, $MSE = 327.40$, $p = .33$, or trial 2, $F(3,125) = .12$, $MSE = 38.60$, $p = .94$.

JOL Resolution

To examine JOL resolution, we computed gamma correlations between each participant’s JOLs and their recall performance across items.\(^1\) During trial 1, all participants made immediate JOLs, and as expected, resolution did not differ between

\(^1\) Similar analyses were conducted using signal-detection measures of discriminative accuracy and results yielded the same outcomes as gamma correlations. Thus, to remain consistent with the vast literature on JOLs, we report only gamma correlations.
the iJOL no-prompt group \((M = .44, SE = .05)\), the iJOL prompt group \((M = .54, SE = .04)\), the pJOL no-prompt group \((M = .46, SE = .04)\), and the pJOL prompt group \((M = .43, SE = .06)\), \(F(3, 124) = 1.07, MSE = .07, p = .36\).

Most important, consider resolution on trial 2. As evident from inspection of Figure 1, resolution for the iJOL groups was higher than the pJOL group that received the recall prompt, which indicates that general knowledge of past test performance is not the only factor contributing to trial 2 resolution. Consistent with this observation, a 2 (JOL type) x 2 (Recall prompt) ANOVA revealed a main effect of JOL, \(F(1, 118) = 54.89, MSE = 5.64, p < .001, \eta_p^2 = .32\). Resolution was also greater for the prompt groups compared to the non-prompted groups, \(F(1, 118) = 12.77, MSE = 1.31, p < .001, \eta_p^2 = \)
.10. The JOL type x Prompt interaction approached significance, $F(1, 118) = 3.57$, $MSE = .37, p = .06$.

**Trial 2 JOLs Conditionalized on Trial 1 and Trial 2 Recall**

To provide an initial assessment about the potential contribution of the three factors to JOL resolution, we examined JOL magnitude on trial 2 for various classes of item. In particular, we conducted an analysis similar to Finn and Metcalfe (2007; see also Finn & Metcalfe, 2008) in which JOLs were conditionalized on trial 1 and trial 2 recall performance. JOL magnitude was computed for 4 subsets of items: (1) Items not recalled on trial 1 that were also not recalled on trial 2 (NN items), (2) items not recalled on trial 1 that were recalled on trial 2 (NR items), (3) items recalled on trial 1 that were recalled on trial 2 (RR items), and (4) items recalled on trial 1 that were not recalled on trial 2 (RN items). Mean JOLs are presented in Figure 2.²

To examine the influence of MPT, we compared NR to RR items. If people are relying on MPT to make trial 2 JOLs, then JOLs will be higher for RR items than NR items. Trial 2 recall performance is held constant in these comparisons (all items were recalled on trial 2), so the influence of other factors such as new learning and forgetting should not influence judgment magnitude (Finn & Metcalfe, 2007). A 2 (Item status: RR vs. NR) x 2 (JOL type) x 2 (Recall prompt) ANOVA revealed an effect for JOL type, $F(1, 120) = 18.21$, $MSE = 12011.53, p < .001$, $\eta^2_p = .13$, which indicates that immediate JOLs were higher than prestudy JOLs. Participants also made higher JOLs for RR than

---

² Values for overall JOL magnitude on trial 2 for each group can not be obtained by averaging magnitude across NN, NR, RR, and RN items because the number of items within each of these subclasses is not equal.
NR items, $F(1, 120) = 180.48, \text{MSE} = 53044.409, p < .001, \eta^2_p = .60$. The Item Status x JOL Type, $F(1, 120) = 50.45, \text{MSE} = 14827.35, p < .001, \eta^2_p = .30$, and the Item Status x Prompt interaction, $F(1, 120) = 16.37, \text{MSE} = 4812.35, p < .001, \eta^2_p = .12$, were significant. The three-way interaction approached significance, $F(1, 120) = 3.32, \text{MSE} = 975.42, p = .07$. These interactions revealed two important effects: (a) JOL differences between NR and RR items were greater for the iJOL groups than for the pJOL prompt group, and (b) such differences arose because JOL magnitudes for previously recalled items were greater for the iJOL groups than for the pJOL prompt group. Put differently, JOLs were the highest when participants not only knew whether a response had been recalled but also knew specifically which response had been recalled.

Figure 2. Trial 2 mean judgments of learning (JOL) in Experiment 1 conditionalized on trial 1 and trial 2 recall. iJOL = immediate JOL; pJOL = prestudy JOL. Prompt/no prompt = whether participants were prompted with prior recall outcome. Error bars represent standard error of the mean.
Concerning the other factors, higher JOLs for NR than NN items would suggest that participants’ JOLs are sensitive to new learning on trial 2. Consistent with this prediction, a 2 (item status: NR vs. NN) x 2 (JOL type) x 2 (Prompt) ANOVA revealed an effect for item status, $F(1,122) = 32.84$, $MSE = 1617.06$, $p < .001$, $\eta_p^2 = .21$, and an Item status x JOL type interaction, $F(1,122) = 28.63$, $MSE = 28.63$, $p < .001$, $\eta_p^2 = .19$, which arose because only the iJOL group was sensitive to factors pertaining to new learning. JOLs also appeared to reflect sensitivity to predicting forgetting for the iJOL groups, because they gave lower JOLs to RN (forgotten) than RR (remembered) items. However, differences between RR and RN items were not significant for any group.

*Estimating the Contributions of MPT, New Learning, and Forgetting*

Based on the previous analysis of JOL magnitudes (Figure 2), MPT and new learning appear to contribute most to JOL resolution on trial 2. To estimate their contributions, we computed partial correlations using the decomposition described by Nelson et al. (2004). Overall resolution (Figure 1) was decomposed into 3 correlations reflecting (1) discriminations between items that were recalled versus were not recalled on trial 1 (or $\gamma_{RN}$, which reflects the influence of MPT), (2) discriminations between dyads of items that were not recalled on trial 1 ($\gamma_{NN}$, which reflects the influence of monitoring new learning), and (3) discriminations between dyads of items that were recalled on trial 1 ($\gamma_{RR}$, which reflects the influence of monitoring forgetting).³

³ Subscripts here refer to the recall status of dyads of items on trial 1 (e.g., RR refers to two items that were both recalled on trial 1), whereas the similar nomenclature for the previous analysis (Figure 2) referred to the recall status of groups of items across trials (e.g., RR referred to all items that were recalled on both trials 1 and 2).
Overall resolution is comprised of each of the above discriminations as weighted in this equation (Nelson et al., 2004):

\[ \gamma = (P_{RN} \star \gamma_{RN}) + (P_{NN} \star \gamma_{NN}) + (P_{RR} \star \gamma_{RR}) \]  

[1]

The parameters \( P_{RN}, P_{NN}, \) and \( P_{RR} \) reflect the proportion of dyads that contributed to the computation of each respective correlation. A low value for \( P \) indicates that the corresponding factor will have a limited influence on overall resolution.

Mean parameters for Equation 1 are presented in Table 1. First consider \( \gamma_{RN} \), which is relevant to the contribution of MPT to resolution on trial 2. A 2 (JOL type) x 2 (Recall prompt) ANOVA examining \( \gamma_{RN} \) revealed an effect for JOL type, \( F(1,121) = 69.17, MSE = 9.57, p < .001, \eta^2_p = .37 \), which indicates that on trial 2, the iJOL groups were more accurate at discriminating between items previously recalled versus not recalled. Prompting participants about their past performance also boosted \( \gamma_{RN} \), \( F(1,121) = 12.77, MSE = 3.12, p < .001, \eta^2_p = .16 \). However, this prompt had a larger influence on the pJOL groups, resulting in a JOL Type x Prompt interaction, \( F(1,121) = 6.07, MSE = .84, p < .05, \eta^2_p = .05 \).

Next we compared \( \gamma_{NN} \) values to evaluate the contribution of monitoring new learning to JOL resolution on trial 2. Note that in the pJOL groups, participants should not be able to monitor new learning, because they made JOLs prior to actually studying the items. Consistent with this prediction, the values for both pJOL groups did not differ from zero, \( ts < 1.0 \). By contrast, \( \gamma_{NN} \) values for both iJOL groups were significantly
greater than zero, \( t_s > 4.5 \), which suggests that monitoring of new learning significantly contributed to JOL resolution on trial 2. A 2 (JOL type) x 2 (Recall prompt) ANOVA revealed a significant effect of JOL type, \( F(1,121) = 23.64, MSE = 3.82, p < .001, \eta_p^2 = .17 \). The main effect for recall prompt, \( F(1,121) = .11, MSE = .02, p = .74 \), and the interaction, \( F(1,121) = .38, MSE = .06, p = .54 \), were not significant.

Inferential statistics for \( \gamma_{RR} \), which reflect the contribution of forgetting to trial 2 resolution, will not be presented because these values are based on only a small subset of participants (\( N = 31 \)). The reason for this outcome is evident from the low \( P_{RR} \) values, which indicate that few of the items that were recalled on trial 1 were not recalled on trial 2. Thus, even if participants were incorporating information about forgetting into their trial 2 JOLs, doing so could not influence their resolution.
Table 1. Decomposition of resolution (gamma correlations) presented in Figures 1 and 3.

<table>
<thead>
<tr>
<th>JOL Group</th>
<th>Proportion of dyads</th>
<th>Partial Gammas</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P_{RN}$</td>
<td>$P_{NN}$</td>
</tr>
<tr>
<td><strong>Experiment 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iJOL no-prompt</td>
<td>.41 (.03)</td>
<td>.59 (.04)</td>
</tr>
<tr>
<td>iJOL prompt</td>
<td>.47 (.04)</td>
<td>.53 (.04)</td>
</tr>
<tr>
<td>pJOL no-prompt</td>
<td>.41 (.03)</td>
<td>.57 (.03)</td>
</tr>
<tr>
<td>pJOL prompt</td>
<td>.50 (.04)</td>
<td>.50 (.04)</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>iJOL no-prompt</td>
<td>.66 (.04)</td>
<td>.10 (.03)</td>
</tr>
<tr>
<td>iJOL prompt</td>
<td>.68 (.03)</td>
<td>.09 (.02)</td>
</tr>
<tr>
<td>pJOL no-prompt</td>
<td>.67 (.04)</td>
<td>.10 (.03)</td>
</tr>
<tr>
<td>pJOL prompt</td>
<td>.66 (.04)</td>
<td>.10 (.04)</td>
</tr>
</tbody>
</table>

Note: Values above can be used to estimate overall $\gamma$, using this equation: $\gamma = (P_{RN} * \gamma_{RN}) + (P_{NN} * \gamma_{NN}) + (P_{RR} * \gamma_{RR})$. Because values above are group means and not individual participant’s values, estimates of the overall $\gamma$ using these group values will not be identical to the actual values (Figure 1) due to rounding errors. However, when individual participant’s values are imputed into the equation, the exact values reported in Figure 1 are obtained.
Although a great deal is now known about the heuristic nature of judgments of learning (Dunlosky & Mefcalfe, 2009), the extent to which multiple factors jointly influence JOL resolution has received little empirical attention. The present results suggest that a multi-factor approach will be required to provide a complete explanation for JOL resolution, especially when JOLs can be influenced by task experience. Two factors in particular—MPT and new learning—appeared to contribute to improvements in JOL resolution following a study-test trial.

The contribution of MPT and new learning to overall resolution on trial 2 (Figure 1) can be estimated by examining the P and γ parameters in Table 1. Consider three key observations this table. First, the P_{RN}, P_{NN}, and P_{RR} parameters do not differ between groups. Thus, differences in resolution on trial 2 are not a function of differences between groups in the number of items previously recalled, newly learned, or forgotten on trial 2. In contrast, they arise because of differences between groups in discriminative accuracy for each subset of items, which is represented by γ_{RN}, γ_{NN}, and γ_{RR}, respectively.

Second, concerning the P parameters, only P_{RN} and P_{NN} differ significantly from zero. Thus, in the present experiment, only MPT and new learning could potentially contribute to overall resolution on trial 2. Finally, the proportion of items that contribute computationally to estimates of the influence of new learning to JOL resolution (54.8%,
see $P_{NN}$ values in Table 1) was slightly larger than the number of items contributing to estimates of MPT (44.8%, $P_{RN}$). However, because $\gamma_{RN}$ is larger than $\gamma_{NN}$ in the iJOL groups and the pJOL prompt group, MPT resulted in higher boosts to JOL resolution on trial 2 than did new learning.

Why does MPT boost resolution? One answer to this question arises from closer inspection of outcomes in Figure 2: For items that had been correctly recalled on both trials (RR), JOL magnitude is lower for prompted pJOLs than for the iJOLs. Moreover, the influence of MPT on JOL resolution (Table 1) was diminished when people only had knowledge for past test performance (as in the pJOL-prompt group) compared to when they could engage in actual retrieval of past test performance (as in the iJOL groups). Note, in the later case, people not only had knowledge of past test performance, but they also knew which item that they had correctly retrieved. Thus, one source of participants’ increased JOLs for items recalled on prior test trials is related to knowing that a specific item had been recalled, and this is probably due to “remembering” of the specific recall episode, not just the knowledge that recall was successful. We return to this intriguing issue in the General Discussion.

In Experiment 1, we used a standard method to demonstrate and explore JOL resolution across multiple trials – study times were experimenter-paced, the retention interval was short, and study on trial 2 occurred immediately after recall on trial 1. We suspect that manipulating these parameters will moderate the joint contribution of various factors to JOL resolution. For instance, if a longer retention interval was used on trial 2 (so that some items recalled on trial 1 would not be recalled on trial 2), then people’s
JOLs may also pick up forgetting that occurs on trial 2. In Experiment 2, we systematically explored the influence of a longer retention interval on judgment resolution by combining pre-study JOLs and analytic methods used here.
In Experiment 1, forgetting for items did not contribute to JOL resolution on trial 2. However, it is unclear whether this finding arose because people’s JOLs do not predict potential forgetting or because of the lack of forgetting that occurred with the relatively short retention interval between study and the test on trial 2. To evaluate these alternatives, the retention interval between study and test on trial 2 was increased to 1 week. Pilot data indicated that a 1 week retention interval was sufficient to increase forgetting between test 1 and test 2, and hence would produce a $P_{RR}$ value greater than zero. Thus, if forgetting does contribute to resolution for JOLs made on trial 2, we expected $\gamma_{RR}$ values to be significantly greater than zero, and hence contribute to overall JOL resolution.
METHOD

Participants

One hundred and thirty-three students from KSU participated for course credit in Introductory Psychology. A 2 (JOL type: immediate or prestudy) X 2 (Recall prompt: prompt about previous recall performance or no prompt, as described below) full-factorial design was used. Participants were randomly assigned to either iJOL no-prompt group \((n = 33)\), the iJOL prompt group \((n = 32)\), pJOL no-prompt group \((n = 32)\), or the pJOL prompt group \((n = 36)\).

Materials and Procedure

Eighty unrelated noun-noun paired associates were used in this experiment. The procedure was the same as Experiment 1 with the following exceptions. First, prior to the initial study-judge-test cycle, each item was presented for 4 seconds for study. The purpose of this study trial was to increase learning that would occur during trial 1. After this familiarity trial, participants engaged in the first study-judge-test trial. Second, during the study-judge-test cycles for trial 1 and trial 2, items were presented for 4 seconds for study instead of 6 seconds. Third, the retention interval between study and test on trial 2 was increased to 1 week from the time of the second study trial. The increased retention interval between study and test on trial 2 was expected to increase the likelihood of forgetting. All participants were informed on the second trial that items would be tested one week later.
RESULTS

Recall and Judgments of learning

The proportion of items correctly recalled did not differ between the iJOL no-prompt group (Trial 1: $M = .36, SE = .04$; Trial 2: $M = .12, SE = .04$), the iJOL prompt group (Trial 1: $M = .41, SE = .03$; Trial 2: $M = .16, SE = .03$), the pJOL no-prompt group (Trial 1: $M = .31, SE = .03$; Trial 2: $M = .10, SE = .03$), and the pJOL prompt group (Trial 1: $M = .36, SE = .03$; Trial 2: $M = .11, SE = .03$) on either trial 1, $F(3,135) = 1.29, MSE = .05, p = .28$, or trial 2, $F(3,135) = 1.69, MSE = .03, p = .17$.

Means across participants’ mean JOLs did not differ for the iJOL no-prompt group (Trial 1: $M = 39.8, SE = 2.86$; Trial 2: $M = 33.8, SE = 3.31$), the iJOL prompt group (Trial 1: $M = 39.7, SE = 2.92$; Trial 2: $M = 36.3, SE = 3.53$), the pJOL no-prompt group (Trial 1: $M = 35.3, SE = 2.79$; Trial 2: $M = 29.7, SE = 3.98$), and the pJOL prompt group (Trial 1: $M = 40.4, SE = 3.27$; Trial 2: $M = 31.0, SE = 3.54$) on either trial 1, $F(3,129) = .66, MSE = 196.34, p = .58$, or trial 2, $F(3,135) = .67, MSE = 292.02, p = .58$.

JOL Resolution

As in Experiment 1, we first examined JOL resolution for trial 1. Since, all participants made immediate JOLs on this trial, resolution did not differ between the iJOL no-prompt group ($M = .54, SE = .03$), the iJOL prompt group ($M = .59, SE = .03$), the pJOL no-prompt group ($M = .55, SE = .05$), and the pJOL prompt group ($M = .54, SE = .05$), $F(3,129) = .25, MSE = .01, p = .86$. 
Most important, we examined resolution on trial 2, which is presented in Figure 3. Resolution was higher for the iJOL groups than the pJOL prompt group, which replicates findings from Experiment 1 that knowledge of past test performance is not the only factor contributing to trial 2 resolution. Consistent with this observation, a 2 (JOL type) x 2 (Recall prompt) ANOVA revealed a main effect of JOL type, $F(1, 126) = 84.40, MSE = 10.81, p < .001, \eta_p^2 = .41$. An effect for recall prompt approached significance, $F(1, 126) = 3.73, MSE = .48, p < .06, \eta_p^2 = .03$. The JOL x Prompt interaction was significant, $F(1, 126) = 4.88, MSE = .63, p < .05, \eta_p^2 = .04$, which indicates that the recall prompt did influence resolution in the pJOL group.

Figure 3. Trial 2 mean resolution as a function of JOL group in Experiment 2. iJOL = immediate JOLs; pJOL = pre-study JOLs. Prompt/no prompt = whether participants were prompted with prior recall outcome. Error bars represent standard error of the mean.
**Trial 2 JOLs Conditionalized on Trial 1 and Trial 2 Recall**

Judgments were conditionalized based on trial 1 and trial 2 recall status to obtain an initial assessment of the influence of MPT, new learning, and forgetting on JOL resolution. Mean JOL magnitude for NN, NR, RR, and NR items is presented in Figure 4. First, we compared NN to RN pairs to assess the influence of MPT on JOL magnitude. Recall that in Experiment 1, this comparison was made using NR and RR items. Given the longer retention interval in Experiment 2, new learning between study and test on trial 2 was diminished, resulting in fewer data points for NR items compared to the other three classes of items. To increase statistical power, we decided to compare NN to RN items to assess the influence of MPT. The logic behind comparing these two classes of items is similar to comparing NR to RR items. Differences in item status on trial 1 contribute to the impact of MPT on JOL magnitude while item status remains constant on trial 2 to control for the influence of other factors (in this case, potential forgetting).

A 2 (item status: NN vs. RN) x 2 (JOL type) x 2 (prompt) ANOVA revealed effects for JOL type, $F(1,131) = 5.28$, $MSE = 4417.72$, $p < .001$, $\eta_p^2 = .04$, indicating that JOL magnitude was higher when participants made JOLs after studying items compared to prior to studying them. JOL magnitude was also higher for RN items than for NN items, $F(1,131) = 216.08$, $MSE = 32075.52$, $p < .001$, $\eta_p^2 = .62$. These main effects were qualified by a JOL type x Item Status interaction, $F(1,131) = 126.50$, $MSE = 18777.67$, $p < .001$, $\eta_p^2 = .49$. An Item Status x Prompt interaction was also significant, $F(1,131) = 9.14$, $MSE = 1356.30$, $p < .01$, $\eta_p^2 = .07$. Thus, knowledge of past test performance
resulted in a larger boost in judgment magnitude, when participants made immediate JOLs compared to when they made prestudy JOLs.

![Figure 4. Trial 2 mean judgments of learning (JOL) in Experiment 2 conditionalized on trial 1 and trial 2 recall. iJOL = immediate JOL; pJOL = prestudy JOL. Prompt/no prompt = whether participants were prompted with prior recall outcome. Error bars represent standard error of the mean.](image)

Concerning the influence of new learning on judgment magnitude we compared NN to NR items. A 2 (item status: NN vs. NR) x 2 (JOL type) x 2 (prompt) ANOVA yielded an effect for item status, $F(1,74) = 14.20$, $MSE = 3739.61$, $p < .001$, $\eta_p^2 = .16$, which indicates that judgment magnitude was higher for NR items. An Item status x JOL type Interaction was also significant, $F(1,74) = 4.10$, $MSE = 1079.32$, $p < .05$, $\eta_p^2 = .05$. This interaction arose because only the iJOL groups’ judgments were sensitive to new learning.
Finally, we examined if people’s JOLs predicted forgetting on trial 2. Lower JOLs for RN items compared to RR items here would suggest JOLs are sensitive to forgetting. JOLs were significantly lower for RN items than for RR items, $F(1,123) = 27.04$, $MSE = 1311.21$, $p < .001$, $\eta^2_p = .18$. Moreover, JOLs were higher in magnitude when made immediately after study opposed to before study, $F(1,123) = 33.03$, $MSE = 42523.51$, $p < .001$, $\eta^2_p = .21$. These main effects were qualified by a Item status x JOL type interaction, $F(1,123) = 11.10$, $MSE = 538.41$, $p < .001$, $\eta^2_p = .08$. This interaction arose because only the iJOLs groups’ JOLs were sensitive to forgetting.

*Estimating the Contributions of MPT, New Learning and Forgetting*

To estimate the contribution of MPT, new learning, and forgetting to JOL resolution on trial 2, we decomposed overall resolution (Figure 3) into 3 partial correlations. As described previously, these correlations reflect estimates of the relative contribution of each factor to overall JOL resolution.

First, consider $\gamma_{RN}$, which reflects the contribution of MPT (Table 1). A 2 (JOL type) x 2 (Recall Prompt) ANOVA examining $\gamma_{RN}$ revealed an effect for JOL type, $F(1,125) = 102.83$, $MSE = 14.86$, $p < .001$, $\eta^2_p = .61$. This effect arose because the iJOL groups were better than were the pJOL groups at discriminating between items that were previously recalled versus items that were not on trial 2. An effect for prompt was also significant, $F(1,125) = 4.33$, $MSE = .63$, $p < .05$, $\eta^2_p = .04$. The JOL group x Prompt interaction was not significant, $F(1,125) = 3.20$, $MSE = .46$, $p = .08$, $\eta^2_p = .03$. 
Next, consider $\gamma_{NN}$, which reflects the contribution of new learning to JOL resolution. If participants’ JOLs accurately predicted new learning on trial 2, then $\gamma_{NN}$ should be significantly greater than zero. Consistent with findings from Experiment 1, $\gamma_{NN}$ did not differ from zero in the pJOL groups, $ts < 1.49$, but was significantly different from zero in the iJOL groups, $ts > 2.01$. A 2 (JOL type) x 2 (Recall Prompt) ANOVA revealed an effect for JOL group, $F(1,117) = 13.53$, $MSE = 2.44$, $p < .001$, $\eta^2_p = .11$. Effects for recall prompt, $F(1,117) = .31$, $MSE = .58$, $p = .58$, $\eta^2_p = .003$, and the interaction, $F(1,117) = 1.32$, $MSE = .25$, $p = .25$, $\eta^2_p = .01$, were not significant.

Finally, in contrast to Experiment 1, forgetting occurred after trial 2 study ($P_{RR}$ values were significantly greater than 0), so that we could estimate $\gamma_{RR}$ values to evaluate the contribution of predicting forgetting to JOL resolution. Participants in the pJOL groups were not expected to be able to predict forgetting for items, because they made JOLs prior to actually studying the items on trial 2. Consistent with this prediction, values did not differ from zero for both of the pJOL groups, $ts < 1.48$. By contrast, values for the iJOL groups were significantly greater than zero, $ts > 2.00$. A 2 (JOL type) x 2 (Recall Prompt) ANOVA revealed an effect for JOL type, $F(1,73) = 10.96$, $MSE = 4.13$, $p < .001$, $\eta^2_p = .14$. As expected, effects for recall prompt, $F(1,73) = .15$, $MSE = .06$, $p = .70$, $\eta^2_p = .002$, and the interaction, $F(1,73) = 2.79$, $MSE = .38$, $p = .10$, $\eta^2_p = .04$, were not significant.
DISCUSSION

When forgetting occurred after study on trial 2, people’s JOLs were sensitive to inter-item forgetting in a manner that significantly contributed to JOL resolution. Thus, people’s JOL resolution on trial 2 (Figure 3) was influenced jointly by at least 3 factors—MTP, new learning, and forgetting. The relative contribution of these factors is presented in Table 1. In the present experiment, 66.8% of the overall JOL resolution on trial 2 for each group was composed of discriminations between items that were recalled versus not recalled on trial 1 ($P_{RN}$ values in Table 1). These high values in combination with the high $\gamma_{RN}$ values for participants who could evaluate prior retrieval success (iJOL groups and pJOL prompt group), resulted in MPT contributing most to overall resolution. Forgetting contributed 23.5% ($P_{RR}$) to overall resolution and new learning contributed to 9.8% ($P_{NN}$). Thus, in contrast to Experiment 1 in which only MPT and new learning contributed to resolution on trial 2, forgetting also contributed. In fact, with the longer retention interval to promote forgetting, the weighted contribution of forgetting was larger than the contribution of new learning ($P_{RR} > P_{NN}$).
GENERAL DISCUSSION

Across two experiments, evidence from multiple analyses converged on the same conclusion that MPT, new learning, and forgetting contributed to JOL resolution when participants engaged in multiple study-test trials. In both experiments, MPT provided the largest contribution to differences in JOL magnitude (Figures 1 and 3) and JOL resolution on trial 2 (see the high $P_{RN}$ and $\gamma_{RN}$ values in Table 1). These results are not entirely surprising given previous research examining the influence of test experience on metamemory judgments (King et al., 1980; Finn & Metcalfe, 2007; 2008). Perhaps most important, the current approach using the prestudy JOL methodology (Castel, 2008) provides some novel insights about why MPT boosts judgment resolution.

Consider the findings from the pJOL prompt group in Experiments 1 and 2. This group made JOLs on trial 2 prior to actually studying items. However, they were provided with the outcome of their previous retrieval attempt for each item—that is they knew their past test performance for the item being judged. This retrieval outcome is an extremely diagnostic predictor of future memory. In fact, the mean gamma correlation between recall on trial 1 and trial 2 was .94 and .84 in Experiment 1 and 2, respectively. Thus, participants in the pJOL prompt groups could have demonstrated high levels of JOL resolution on trial 2 by using this prompt alone, such as by responding with high JOLs when prompted that they recalled an item and lower JOLs when prompted that they did not. However, participants in the pJOL groups did not use this strategy, and their
subsequent JOL resolution on trial 2 was relatively low. These observations raise the question, Why do people use knowledge of prior test performance differently in the iJOL groups and the pJOL prompt group? One answer lies in the differences in the subjective experiences involved in remembering a prior retrieval outcome versus knowing a prior retrieval outcome.

The subjective experiences involved in remembering versus knowing are distinct (Gardiner & Richardson-Klavehn, 2000). When we remember something—such as whether a study item was correctly recalled on a previous test—we engage in mental-time travel that involves retrieving a specific episode of the prior testing event (Gardiner & Richardson-Klavehn, 2000). Knowing requires only semantic information opposed to episodic information (Tulving, 1985). As a consequence, the subjective experience of knowing is often less personal, and as Gardiner and Richardson-Klavehn (2000) note, “There is no awareness [with knowing] of reliving any particular events or experiences” (p. 229). The subjective experience involved with remembering something versus just knowing something has been linked to higher levels of confidence in memory for that information (Tulving, 1985; Dunn, 2004; but see Gardiner & Java, 1990). Thus, the aspect of MPT that may largely boost confidence in future memory for previously recalled items may be the subjective experiences associated with remembering a specific prior retrieval outcome and not just knowing whether something was correctly recalled or not recalled. In the present experiments, the pJOL prompt groups only knew whether they correctly recalled an item (or not) because they made JOLs prior to actually studying items. In contrast, the iJOL groups could remember that they actually recalled or did not
recall an item when making their JOLs. This remembering of previous test episodes may boost people’s confidence in future memory for recalled items, and should be more fully explored in future research.

In addition to past test performance, JOLs from the iJOL groups were also sensitive to new learning (Experiments 1 and 2) and forgetting (Experiment 2). Thus, when making JOLs, people appear to integrate multiple cues that pertain to both past and future memory status for items. This finding appears to be inconsistent with recent conclusions that indicate that people do not consider new learning (Kornell & Bjork, 2009) or forgetting (Koriat et al., 2004) when making predictions about their future memory. For example, Kornell and Bjork (2009, Experiment 6) asked participants to predict how many items they would recall following multiple study-test opportunities. Participants engaged in either 1 study-test trial (ST group), 2 study-test trials (STST group), 3 study-test trials (STSTST), or 4 study-test trials (STSTSTST). Participants in each group were asked during the initial study phase (that is, on trial 1) to predict their test performance on the final test. Thus, the ST, STST, STSTST, and STSTSTST groups were asked to predict performance on a test that would occur after additional study and testing opportunities that had yet to occur. In order to accurately predict future performance, people in these groups would have to realize that additional study opportunities would result in additional learning. Across multiple experiments, participants’ average predictions did not differ between groups, despite the fact that recall improved with additional study. These findings suggest that people do not consider new learning when making predictions about their memory.
Koriat, Bjork, Sheffer, & Bar (2004) examined whether people consider forgetting when making predictions about future memory. In one experiment (Koriat et al., 2004, Experiment 1), participants were asked to make aggregate and item-by-item predictions about the number of items they would recall either immediately after study, 1-week after study, or 1-month after study. Both aggregate and average item-by-item predictions did not differ between groups, despite actual differences in final recall across each retention interval. Perhaps more surprising, in a follow-up experiment (Experiment 4c), participants were asked to make predictions about how many items they would recall after a 1-year retention interval, and even under this extreme scenario, participants still predicted they would recall the same number of items as a group who would be tested immediately after study.

The results from both Kornell and Bjork (2009) and Koriat et al. (2004) provide convincing evidence that people’s memory predictions discount the influence of new learning and forgetting following study. Furthermore, as noted above, they seem inconsistent with the current findings, which indicate that people’s JOLs appear to be sensitive to both new learning and forgetting. However, there are several important differences between those experiments and the current ones. One major difference is that findings concerning a lack of effect for new learning and forgetting have focused on the magnitude of the judgments across all items, where as the current experiments focused on discriminative accuracy of the judgments.

New learning and forgetting are item specific. That is, specific items are learned with additional practice and specific items are forgotten over time. Aggregate judgments
and average item-by-item judgments ignore the effects occurring at the item level, and JOLs are sensitive to item-level influences that are diagnostic of future memory such as item relatedness (Carroll, Nelson, & Kirwan, 1997; Koriat, 1997; Rabinowitz, Ackerman, Craik, & Hinchley; 1982) and fluency (Koriat & Ma’ayan, 2005; Matvey, Dunlosky, & Guttentag, 2001). Thus, people do need to explicitly apply beliefs about memory when making JOLs and their JOLs can still be influenced by new learning and forgetting. Consistent with this conclusion, in the present experiments, JOL resolution was influenced by new learning and forgetting only in the iJOL groups and not the pJOL groups. The reason for differences in resolution between these groups can be linked to the differences in access to idiosyncratic cues at the time of judgments. In the iJOL groups, JOLs were made in the presence of item-specific cues because they were made after studying an item. In contrast, in the pJOL groups, JOLs were made in the absence of any idiosyncratic cues.

In summary, the present experiments used a recent methodological advance (prestudy JOLs, Castel, 2008) to investigate the joint influence of multiple factors on JOLs. Comparisons of JOLs made immediate after study versus JOLs made prior to study (prestudy JOLs) on trial 2 allowed us to isolate the influence of item-specific factors on JOLs. Several important findings were revealed, and conclusions were based on converging evidence from multiple analyses. First, MPT contributed most to boosts in JOL magnitude and improvements in resolution across trials. Second, comparisons of the iJOL groups and the pJOL prompt groups revealed that the aspect of MPT that boosts people’s confidence in items is related to remembering specific past-test episodes and not
just knowing the outcome of prior retrieval attempts. Third, participants JOLs and subsequent resolution were sensitive to new learning and forgetting, but only when participants’ judgments were made after and not before study. Thus, participants’ JOLs appear to integrate information from multiple factors, and these factors jointly contribute to JOL resolution.
REFERENCES


Serra, M. J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the