ABSTRACT

Title of Document: HOW DO JUDGMENTS OF INTENTION

                            (JOIS) DIFFER FROM JUDGMENTS OF

                            LEARNING (JOLS)?

                            Petra Scheck, Doctor of Philosophy, 2006

Directed By: Dr. Thomas S. Wallsten
                            Professor of Psychology

The purpose of this research was to investigate whether metacognitive judgments for

prospective memory (PM; also called Judgments of Intention - JOIs) differ from

judgments of learning (JOLs). Specifically, this researcher tested whether JOIs

exhibit the "delayed-JOL effect", in which JOLs made at a delay following study are

more accurate predictors of performance than JOLs made immediately following

study. Results from the first two experiments showed no delayed-JOI effect. In

Experiments 3 – 5 a time-based paradigm was used to investigate whether the type of

judgment prompt might explain this lack of delayed JOI effect. Results suggest that

participants are able to make moderately accurate predictions of prospective memory

(PM) performance and that delayed JOIs may be more accurate than immediate JOIs

when prompted with a cue that does not include either the recognition nor recall

target.

HOW DO JUDGMENTS OF INTENTION (JOIS) DIFFER FROM JUDGMENTS

OF LEARNING (JOLS)?

By

Petra Anne Scheck

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2006

Advisory Committee:
Professor Thomas S. Wallsten, Chair
Richard D. Brecht
Hendrik J. Haarmann
Michael R.P. Dougherty
David E. Huber

Dedication

This dissertation is dedicated to my loving husband Brian, without whose unwavering support I could never have completed it.

To Dr. Thomas S. Wallsten, for his kind mentorship and keen eye for passive voice.

To the late Dr. Thomas O. Nelson, who I believe owes me a beer.

Acknowledgements

I would like to thank David Schulzinger, who spent endless hours coding and

checking data and never grumbled.

Table of Contents

List of Tables

List of Figures

How do Judgments of Intention (JOIs) differ from

Judgments of Learning (JOLs)?

Imagine that you have the task of returning a video to the rental store on your way to work tomorrow morning. You judge whether or not you'll remember, and based on your judgment either create an external reminder (e.g. write yourself a note), create an internal reminder (e.g., "Seeing the VCR before I walk out the door in the morning will prompt me to remember the video task") or choose not to create an explicit reminder (e.g. "I don't need a reminder; I'll remember to take the video back.").

The video task is a prospective memory (PM) task, or a memory for an intention to do something in the future. Predicting whether or not you will remember it is an example of metacognitive monitoring, while choosing a cue is an example of employing a memory strategy which may or may not improve your chance of performing the task. This paper focuses on the monitoring aspect of the memory task (i.e. people's accuracy at predicting whether or not they will remember to do a PM task), although the factors one considers when deciding whether to construct a reminder and the type of reminder one chooses are also important aspects of prospective memory.

Though begun in the 1970's by Meacham and Singer (1977), investigation of PM phenomena lacked a laboratory paradigm until Einstein and McDaniel introduced one in 1990. In Einstein and McDaniel's experiment, college-age and elderly participants studied short lists of target nouns. Then the targets were interspersed among other nouns as part of a short term memory (STM) task in which participants were asked to recall the lists of nouns. For the PM task, participants were instructed to press a key if

a target (which appeared relatively infrequently) occurred on the STM list. Results showed that younger participants performed better than older participants on the STM task, but there were no age differences on the PM task. Use of an external memory aid improved performance on the PM task for both groups.

Since the adoption of this paradigm, research has shown that carrying out an event-based PM task typically produces declines in performance on the on-going task, even on trials in which no PM target appears (e.g., Smith, 2003; however, note that in some PM experiments, the decline in performance on the on-going task is necessitated by the paradigm because of a pause or stop in the on-going task to allow for the PM task to be carried out), and that performance on PM tasks tends to decline as the complexity of the PM task increases (Marsh, Hancock & Hicks, 2002). Additionally, there are age-related performance decrements for some PM tasks but not others (Maylor, Smith, Della Sala & Logie, 2002; Vogels, Dekker, Brouwer & DeJong, 2002), and among those with suspected memory impairment, more valuable objects (e.g., a wallet) required an equivalent number of cues to prompt retrieval as less valuable objects (e.g., a pen; Reese & Cherry, 2002).

Only a few studies of PM have investigated any metacognitive aspect, but Cohen, Dixon, Lindsay, Dixon & Masson (2003) have speculated that participants may use metacognitive monitoring to pace their progress through a PM task and allocate attention to the different aspects of the task. Clinical studies have shown that self-rated memory self-efficacy is a significant predictor of performance on a PM task (McDonald-Miszczak, Gould & Tychynski, 1999). Respondents with traumatic brain injuries rated their prospective memory failures as occurring with equal frequency as

those of normal controls, but close friends and relatives rated the patients' failures as comparatively more frequent (Roche, Fleming & Shum, 2002).

To date, no published literature has examined whether learners are able to use item-by-item metacognitive judgments, such as Judgments of Learning (JOLs) to discriminate between PM tasks that they will vs. will not remember to do. This may be due in part to some important difference in the traditional JOL paradigm and the paradigm for PM tasks. Investigation of JOLs typically includes a phase in which learners study pairs of items (e.g., OCEAN – TREE) and are then shown the first item in the pair and asked to make a judgment (either immediately or at some delay) about how likely they are to remember the second item when cued with the first on a later test (e.g., "How confident are you that in about ten minutes from now you will be able to recall the second word when prompted with the first (0 = definitely won't recall, 20 = 20% sure, 40..., 60..., 80..., 100 = definitely will recall?)". After judgments have been made for all items, there is usually a criterion test against which the metacognitive judgments can be compared.

The paradigm for PM is similar to that for JOLs in that it typically has a learning phase and a criterion test. However, the PM test typically occurs in the context of some other on-going activity (often a STM task) during which the PM cues occur relatively infrequently. Predictive judgments could easily be included in a PM paradigm by inserting them between the study phase and the criterion test, but with an important difference: participants would not be asked to monitor their performance on a single task (as in the classic JOL paradigm), but rather on a task with two components: recognition and recall [Note: some researchers have further dissected the

PM to include recognition, verification, retrieval of the response and coordination of the PM response with any on-going activity (Cohen, Dixon, Lindsay, Dixon & Masson, 2003)]. Successful PM performance involves both recognition that a cue has occurred, and recall of the action the cue was meant to prompt (though of course one might have successful recall even if the cue does not occur or one fails to recognize the cue, though presumably some other cue has served in it's place to indicate the need to recall the task).

To include metacognitive judgments in a PM experiment, one might use paired-associate items in which the first item in the pair is meant to act as the recognition cue and the second item in the pair is meant to be recalled. Participants would be asked to study these target pairs so that they could recall the second item in the pair when prompted with the first, but also so that they would recognize the first item in the pair as a recognition target. The prompt for the judgment would be different from those used for JOLs, as the criterion test would be different. Instead of asking a participant how likely he or she would be to recall the second item in a pair when prompted with the first, one might instead ask the participate to rate his or her confidence that he or she would be able to recognize the first item as a cue and respond by recalling the second item in the pair. At test, the first items in the target pairs would appear during another on-going task (e.g., recall of short lists of words). The criterion for correct performance on the PM task would be to recognize a word presented on in a STM list as having comprised the first item from one of the studied paired-associates and correctly recalling the second item. Figure 1 illustrates some of these differences in

paradigm between metacognitive judgments made for recall of paired associates and those made for PM.

A central question of this paper is whether learners can use metamemory judgments to predict performance on a PM task. In addition to the accuracy of predictions, this research also investigates other ways in which judgments about performance on a PM task might be like traditional JOLs, and thus also asks whether they exhibit the "delayed-JOL effect", in which delayed JOLs are more accurate predictors of recall performance than are immediate JOLs (Nelson & Dunlosky, 1991). I chose this established empirical effect to determine the ways these judgments might differ from those made for traditional paired-associate learning predictions. This effect seems like an ideal phenomenon to investigate the similarities between the two types of judgments as it is a robust and easily-produced phenomenon in the context of a paired-associate learning task.

Some theories in the literature offer explanations for the delayed-JOL effect. Spellman & Bjork (1992) speculated that the delayed-JOL effect might be caused by the delayed judgment acting as a spaced study trial, which would be expected to boost recall for items that are retrievable at the time of judgment. Nelson and Dunlosky (1992) countered that if this was the case, we should expect to see better recall performance for items given delayed JOLs than those given immediate JOLs, which not only does not always obtain, but occasionally reverses (i.e. recall performance is sometimes better for items given immediate JOLs). They proposed instead the Monitoring Dual Memories (MDM) hypothesis, which states that both short-term and long-term memories are accessible when making immediate JOLs, which may cause

difficulty in making accurate judgments, as one cannot distinguish whether an item is retrievable because it has been stored in long-term memory or because it is still active in short-term memory. However, when making a delayed JOL, sufficient time has passed that only long-term memory is accessible, making it easier for one to determine whether an item resides there in memory.

A series of experiments was designed to determine whether people can make accurate PM predictions. In the first experiment, participants made Judgments of Intention (JOIs, i.e., JOLs for PM) as if the PM task was a unitary task (i.e., not composed of a recognition and recall aspect). In the second experiment, participants were asked to make both JOLs and JOIs and were asked to make their judgments about either the recall aspect of the task, the recognition aspect of the task, or the entire PM task. In the third through fifth experiment, I used a time-based paradigm to investigate whether the type of judgment prompt might explain the results of the first two experiments.

The first study is designed to investigate whether learners are able to accurately monitor PM when asked to monitor recognition and recall as a unitary task (i.e., both the recognition and the recall aspect are conceptualized and explained to participants as a single "step"). Because both JOLs and JOIs share a recall aspect in the criterion test, it seemed reasonable to predict a delayed JOL effect for both types of judgments, even if participants are less accurate at monitoring the recognition aspect of the task. However, because JOLs for a paired-associate learning task are less than perfectly accurate in predicting recall performance, especially when made immediately following study I predict moderate accuracy for PM predictions. It

would be very surprising if JOIs exceeded the level of accuracy usually observed for JOLs (gamma correlations of .30-.40 for immediate JOLs and .70-.90+ for delayed JOLs), as participants are being asked to judge a more complex memory task that occurs in the context of an on-going STM task.

The explanations for delayed-JOL effect also have some bearing on the hypotheses for this experiment. If Spellman and Bjork are correct in saying that delayed judgments create a recall boost for items that are retrievable at the time of judgment, we would expect to see better PM performance for items given delayed judgments. The MDM hypothesis, however, predicts no difference in PM performance based on the timing of the judgments.

Additionally, Experiment 1 manipulates the length of the on-going STM task to determine whether longer on-going tasks cause a decline in PM performance.

Experiment 1

Method

Participants. Thirty-four students in a senior level psychology course at the University of Maryland participated as part of a course requirement.

Design. This experiment used a mixed-factors design with timing of judgment (immediate or delayed) manipulated between subjects and length of the on-going STM trial (three or six words) manipulated within subjects. Length of STM trial was randomized.

Procedure. Participants were given both written and oral instructions before participating in the experiment. Ten paired associates (e.g. BREAD – J) were displayed one at a time at a six-second rate. Participants were instructed to learn the

pairs so that they would be able to respond with the letter if they saw the noun on the later test.

Either immediately after the offset of the learning trial (immediate JOI condition) or after a 30-second filled delay (delayed JOI condition), participants were prompted with the first item in each pair and asked to make their JOI with the following question, "How confident are you that in about ten minutes from now you will be able to respond with the letter when you see the word (0 = definitely won't complete the PM task, 20 = 20% sure, 40..., 60..., 80..., 100 = definitely will complete the PM task)?". It was specified in the instructions that they should judge the entire task (i.e., both recognition and recall). At the offset of the learning and judgment phase, participants engaged in a brief distracter task which took them about 10 minutes.

During the test phase, participants were shown 20 lists of nouns, each three or six words long, at a two-second rate. They were asked to study each list so that they could remember all the nouns in each list. If, during a the presentation of a list, they recognized one of the cues from the learning phase, they were instructed to press the key representing the letter associated with the word (e.g. if "BREAD" was present in the list, they were to press the letter "J"). Half of the lists contained cues from the learning phase. No more than one cue appeared during any list and the cue was never the last word in any trial. At the end of each STM trial participants were prompted to record the words in the list with the cue "????". Participants were permitted to make their PM response any time during the presentation of the STM list until they were cued to recall the list. After they were prompted for STM recall, they were no longer permitted to respond to the PM cue.

A distracter between learning and recall contained 10 multiple-choice questions about personal preferences. This task was self-paced. Questions were displayed on the computer monitor and participants input their responses on the keyboard.

Materials. The PM learning lists contained paired associates composed of a noun and a letter which was neither the first nor last letter of the noun. Each PM list contained 10 paired associates.

The STM lists each contained 3 or 6 concrete nouns. Half contained cue words from the learning list. Participants recalled their answers on a lined sheet of paper. Responses to the PM task were made on the keyboard and recorded by the computer.

Results

Proportion correct (and standard errors) for PM and STM performance are reported in Table 1.

PM performance. Performance on the PM task was scored as "correct" if the participant pressed the correct key in the trial in which the PM cue appeared. An analysis of variance (ANOVA) on number of correct PM responses showed no significant main effect of JOI timing (immediate or delayed), $F(1, 32) = 2.68$, $p > .10$, or length of the STM trial, $F(1, 32) = .28$, $p > .10$, on PM performance. The interaction was also not significant, $F(1, 32) = .08$, $p > .10$.

STM performance. Performance on the STM task was scored as "correct" if the participant wrote down the correct word in the order it appeared in the list. An ANOVA on number of correct STM responses showed no significant main effect of

JOI timing (immediate or delayed) on STM performance for trials with PM targets, $F(1, 32) = .77$, $p> .10$, or for trials without PM targets, $F(1, 32) = .60$, $p> .10$. However, there was an effect of length of the STM trial for trials with PM targets, $F(1, 32) = 63.13$, $p< .05$, and for trials without PM targets, $F(1, 32) = 120.58$, $p< .05$. Participants performed better on trials with three items than those with six. The JOL timing by list length interaction was not significant either for target trials, $F(1, 32) = .13$, $p> .10$ or trials without targets, $F(1, 32) = 1.27$, $p> .10$.

Predictive accuracy of JOIs. Gamma correlations were computed between JOIs and PM performance for each participant for both immediate and delayed JOIs. A t-test showed no significant difference in predictive accuracy between immediate and delayed JOIs, $t(23) = .95$, $p> .10$, with a mean gamma of $\gamma = .37$ for items given immediate JOIs and $\gamma = .36$ for items given delayed JOIs[1].

Discussion

The finding that delayed JOIs were not more accurate than immediate JOIs may be because participants are unable to accurately monitor recognition, recall, or because they are unable to predict performance that occurs in the context of another on-going task. The lack of a "delayed JOI effect" cannot be attributed to any difference in performance, as performance on the perspective memory task did not differ as a function JOI timing nor as a function of the length of the STM task. However, because it cannot be determined from the data from the first experiment what caused the lack of delayed JOI effect, a second experiment was designed to examine one possible basis for the finding, namely that participants were unable to make accurate monitoring judgments about the combined recognition plus recall task.

Although no delayed-JOI effect obtained, it is important to note that there was no difference in PM performance for item given immediate as compared to delayed judgments. This provides evidence for the MDM explanation of delayed-JOL effect and suggests that merely making delayed judgments is not sufficient to improve performance. The JOIs themselves may be useful in determining which items need further study or other reminders in order for the intentions to be carried out, but there is no advantage in this case of making the JOIs at a delay.

Experiment 2

The goal of the second experiment was to determine which of the differences between the classic JOL paradigm and the new JOI paradigm contributed to the lack of a delayed-JOI effect. Specifically, I designed this experiment to determine whether monitoring the recognition aspect of the task or the combined PM task was sufficiently different from monitoring a recall-only task to eliminate the increased accuracy for delayed judgments. Participants were asked to give judgments either about their performance on the recognition aspect, the recall aspect, or the combined PM task.

Previous studies have assessed the separated recognition and recall components of the PM task (e.g., Cohen, Dixon, Lindsay, Dixon & Masson, 2003). This is the first study in which participants were asked to make judgments about both aspects of the task.

Method

Participants. Thirty-nine students in a senior level psychology course at the University of Maryland participated as part of a course requirement.

Design. This experiment used a mixed-factors design with the criterion test for which the judgment was made (recognition, recall or both) manipulated within subjects and timing of judgment (immediate or delayed) manipulated between subjects. The nature of the criterion test varied over blocks. Thus participants studied, made JOIs and were tested on items of one JOI-type and then repeated the sequence for the other JOIs types. Order of the blocks was counterbalanced across participants.

Procedure. Participants were instructed to learn paired associates as in the previous experiment. JOIs followed the same procedure as previously described with one exception: participants were asked to predict either their performance on the recognition aspect of the task, the recall aspect of the task, or their successful completion of the entire PM task. The cue for recognition memory JOIs was, "How likely are you to remember to push the spacebar to indicate that you recognize this word as a cue?" The cue for recall task JOIs was, "How likely are you to recall the task that was associated with the cue?" The cue for JOIs made for the combined PM JOIs was, "How likely are you to both remember to push the spacebar to indicate that you recognize this word as a cue AND recall the task that was associated with the cue?" For all judgments, participants were asked to respond on a scale from 0% to 100%, where 0% perfect confidence that they *will not* remember the item and 100% represents perfect confidence that they *will* remember the item.

The test phase was the same for all conditions regardless of the type of JOI. It was identical to that in the first experiment except that participants were instructed to press the spacebar if they recognized one of the cues from the learning phase, and

were asked at the end of the STM trial to recall the letter that was paired with the noun, even if they did not recognize the item as a PM cue.

Half of the STM lists contained PM cues, with no more than one cue appearing on any trial. Order of the presentation of the items in the STM trials was random, with the exception that PM targets did not appear in the last or penultimate position.

Materials. The learning lists contained paired associates composed of a noun and a letter which was neither the first nor last letter of the noun. Each PM list contained 10 paired associates.

A short distracter between study and test occurred for all three JOI conditions. The STM recall list contained short lists of concrete nouns. Half of these lists contained one cue word from the learning list. Participants recalled their answers on a lined sheet of paper. Responses to the PM task were made on the keyboard and recorded by the computer.

Results

As proportion correct recognition, recall, PM performance and STM performance have no direct bearing on the accuracy of PM predictions and did not, in Experiment 1, show a pattern of results relevant to the experimental question, they are not reported here.

Predictive accuracy of JOIs. I computed gamma correlations between JOIs and PM performance for each participant for both immediate and delayed JOIs for each type of JOI and aspect of PM performance (see Table 2).

T-tests revealed no significant effect of delay for the gammas between JOI and recognition in the condition in which participants were asked to monitor

19

recognition. In fact, the gamma values of .31 and .37 are very close to those observed in the first experiment (which were $\gamma = .37$ and $\gamma = .36$ for immediate and delayed JOIs, respectively).

There was a classic delayed JOL effect for the gammas between JOI and recall in the condition in which participants monitored recall. This is the condition in which the judgments most closely matched the traditional JOL paradigm.

There was also a significant difference in predictive ability for immediate versus delayed JOIs for the gammas between JOIs and PM performance in the condition in which participants were asked to monitor PM. However, this outcome is difficult to interpret because the gammas are out of the range usually observed for immediate (.30-.40) and delayed (.70-.90+) JOLs.

Discussion

It is difficult to draw conclusions based on the first two experiments, especially since a small number of data points contributed to the gamma correlation for each participant. The evidence so far does not distinguish whether monitoring the recognition aspect or monitoring the recall aspect leads to difficulty in monitoring the PM task.

Although Thiede & Dunlosky (1994) showed a delayed-JOL effect for recognition, they used paired-associates tested during 10AFC recognition in which the correct recognition target appeared as a choice on every test trial, while this JOI paradigm employs embedded single-item recognition in which the recognition cue occurred relatively infrequently. Thiede & Dunlosky's paradigm also differs from

that used in the first two experiments both because the cue for the JOL was different and because the recognition test was not embedded in a STM task.

Additionally, T. Nelson (personal communication, February 11, 2004) suggested that one explanation of why the predictive accuracy of recognition JOIs does not improve with a delay between study and judgment may be that the cue used is not sufficient to provide the insight into memory that produces the delayed JOL effect.

A study by Nelson & Dunlosky (1997) illustrates how the information present at the time of judgment might influence whether or not one is able to accurately monitor memory. When Nelson and Dunlosky asked learners to make JOLs with both the cue and target present at the time of judgment they did not observe a delayed-JOL effect. Presumably, this was because the presence of the target prevented the type of monitoring of the memory state that leads to very good accuracy at predicting performance. It is possible that this phenomenon alone explains why we did not observe an effect of delay for recognition memory in the JOI paradigm. The cue for the JOI was the cue alone (BREAD - ?) and not the cue-target pair, but while BREAD is the cue for the recall task, it is the target for the recognition task (i.e. the recognition task was to press the spacebar when a cue word was displayed in the context of the STM list).

The presence of the target in the JOI cue may be only a partial explanation of the lack of delayed JOI effect. It is also possible that the embedded nature of the PM task also contributed to difficulty in monitoring memory for intentions. Although embedded item recognition is the type of recognition Einstein and McDaniel (1990)

used in their PM paradigm, it is not the only type of recognition cue available (or used) in real-world PM tasks. Marsh, Hicks, Cook, Hansen and Pallos (2003) say that there are many types of PM tasks (e.g., time-based, activity-based, event-based, novel or habitual) and there may be just as many types of recognition memory.

For this series of experiments, I chose to continue using embedded item recognition in the JOI paradigm as it is used commonly in laboratory investigations of PM and seems to be the most common type of recognition in real-world PM tasks. Indeed, it is very difficult to imagine examples of PM tasks that might include recognition or forced-choice recognition. However, I altered the JOI cue so that the target for recognition was not included to determine if the presence of the recognition cue is what caused the lack of delayed JOI effect.

<div align="center">Experiment 3</div>

Experiment 3 again used item recognition embedded in a STM task. However, this experiment investigated whether the classic delayed JOL effect would occur for JOIs when neither the recall nor recognition target was presented in the prompt for the judgment. Another important difference between this experiment and the previous two is that in this one, the cue to perform the prospective memory task was the time displayed on the screen during a STM task rather than the occurrence of a word.

Method

Participants. Twenty-five undergraduate students participated as part of a course requirement or for extra credit in a course. Two of those participants failed to follow directions and I discarded their data, leaving a sample size of 23.

Design. This experiment used a mixed-factors design with timing of judgment (immediate or delayed) manipulated within subjects and type of judgment cue manipulated between subjects. Judgments were prompted with either a cue-only prompt for recall, which included the target for recognition (a JOL-type prompt, e.g., "How likely are you to remember to do the prescribed action at the correct time?, 7:03pm - ?"), or with a cue that included no explicit cue for recall and no target for recognition (a JOI-type prompt, e.g. "How likely are you to remember to do the prescribed action at the correct time?").

Procedure. Participants received both written and oral instructions. One paired associate (e.g. 7:03pm – COOK PASTA) was displayed for 2 seconds at the beginning of each trial. Participants made their metacognitive judgment either immediate after the offset of study of the PM task, or after a 30 second filled delay. Delay in this experiment was produced by having participants complete word scrambles. For immediate JOLs trials, participants completed the scrambles after both study of the PM item and judgment, while for delayed JOL trials they completed the scrambles after study of the PM item but before making their judgment.

After completing the PM study and judgment phase participants completed a STM task. They saw several lists of six nouns each at a 2-second rate under instructions to study each list in order to remember all the words from each trial. After studying each list they were prompted with ???? which was their cue to recall the items from the STM list. During the STM task, a fake time was displayed in the lower right-hand corner of the computer screen. If they recognized that it was for the PM task, they were to press the spacebar to interrupt the trial. If it was in fact a time

at which they were supposed to do some task, they then were asked "What are you supposed to do at this time?" Additionally, they were asked to recall the time at which they were to complete the task. If they pressed the spacebar incorrectly, they were shown a screen which read "I am sorry, this is not the time to complete your task." If, after the entire STM task was completed, the participant had not pressed the spacebar or had done so, but had only indicated incorrect times, he or she received the prompt "What were you supposed to do during this trial?" and also asked to recall the time at which he or she was to do it. A total of 20 trials continued in this fashion, each with a new PM task for each trial.

It is important to note that the fake time displayed on the monitor did not coincide with the time of day (except perhaps by coincidence) and did not progress at the same speed as real time. Within each trial however, it only moved forward at a constant rate, though some trials "progressed through time" faster than others.

Materials. The learning lists contained paired associates composed of a time and a task that was described with a noun and a verb. The STM task contained short lists of concrete nouns.

Results

Performance on the recognition part of the PM task was scored as a hit if the participant pressed the spacebar anytime during the STM list after which the correct time had appeared so.

Predictive accuracy of judgments. I computed gamma correlations between judgments and recognition, recall and PM performance for each participant for both immediate and delayed judgments. The mean gammas are reported in Table 3.

An ANOVA for gamma correlations between judgments and recognition performance revealed no significant main effect of prompt type (JOL or JOI), a borderline-significant effect of judgment timing, $F(1, 23) = 3.47$, $p = .07$, and no significant interaction.

An ANOVA for gamma correlations between judgments and performance on the recall performance revealed neither significant main effects nor a significant interaction.

An ANOVA for gamma correlations between judgments and performance on the entire PM task revealed no significant main effect of prompt type, a borderline-significant effect of judgment timing, $F(1, 23) = 3.67$, $p = .06$, and no significant interaction.

Strangely, the gamma correlations are again out of the range usually observed, even for monitoring recall. The only correlation in the expected range is that for immediate judgments of the JOL-type. Moreover, there was no delayed-JOL effect even for the JOL-type cues. This is an extremely unusual result, as the delayed-JOL effect is very robust.

Experiment 4

Because of the peculiar gamma correlations in this experiment, I replicated it. The excessive length of the session may have caused participants to lose interest in the task, thus adversely affecting their ability to provide reasonable JOIs or JOLs. (Some participants did not finish in the two hours allotted.) Experiment 4 corrects this possible problem.

Method

The method for Experiment 4 was identical to that of Experiment 3 except that the presentation time for the PM and the STM items was reduced to one second and there were only 16 trials.

Participants. Forty undergraduate students participated as part of a course requirement or for extra credit in a course. Seven of those did not follow instructions and I discarded their data. This left 33 participants.

Results

Predictive accuracy of JOIs. I computed gamma correlations between judgments and PM performance for each participant for both immediate and delayed judgments. The mean gammas are reported in Table 4. Separate ANOVAs examined each of the three types of gamma correlation (i.e., between judgment and recognition performance, judgment and recall performance and judgment and PM performance).

The ANOVA for gamma correlations between judgment and recognition performance showed no significant effect of JOL timing or of type, nor a significant interaction.

The ANOVA for gamma correlations between judgment and recall showed no significant effect of judgment timing or type (JOL or JOI), but a borderline-significant interaction, $F(1, 18) = 3.11$, $p = .09$. Pairwise t-tests showed that for JOL-type judgments, there was a borderline-significant difference in accuracy for immediate as compared to delayed judgments $t(8) = 2.25$, $p = .05$. For JOI-type judgments, this difference was not significant, $t(10) = .06$, $p > .10$.

The ANOVA for gamma correlations between judgment and PM performance showed a significant interaction, $F(1, 23) = 5.32$, $p < .05$, a significant effect of judgment timing, $F(1, 23) = 4.85$, $p < .05$, and a borderline-significant effect of judgment type, $F(1, 23) = 3.58$, $p = .07$. Pairwise t-tests showed that for JOL-type judgments, there was a significant difference in accuracy for immediate as compared to delayed judgments $t(11) = 2.86$, $p < .05$. For JOI-type judgments, this difference was not significant, $t(12) = .09$, $p > .10$.

Although many of the gamma correlations are still outside the expected range the delayed JOL predictions were more accurate than the immediate for both recall and PM. This difference did not emerge for the JOI-recognition judgments. These results suggest that the increase in accuracy due to delay may be specific to retrospective recall tasks and not to prospective memory tasks, and therefore that there may be a psychological difference between the two types of tasks.

One explanation for the out-of-expected-range gammas is that prediction on this time-based task was difficult because the fake time progressed in a way participants were not used to, (i.e., more quickly). Although "time" advanced monotonically within each trial, it varied between trials (e.g., one trial might start out at 11:06am, but the next trial could start at 8:51am), and corresponded to the actual time only by chance. Because this is very different from what people are used to experiencing in their daily lives, it may increase the difficult of making metacognitive predictions about such a situation.

It might also be the case that the out-of-expected-range gamma values were caused by the time-based (as opposed to event-based) prospective memory task and that the fast or fake time was irrelevant.

Experiment 5

It is difficult to interpret the relative accuracy of prediction (as measured by gamma) in both Experiment 3 and 4. Thus I designed Experiment 5 as an additional attempt to obtain a delayed-JOI effect. Given the possible explanations for the prior results, it may make sense to test the accuracy of JOIs in an event-based task. However, as there is some debate over the distinction between time- vs. event-based tasks (e.g., 4:00 pm can be conceptualized as either a time or an event or even the time at which a specific event is scheduled to occur, any of which could trigger recall of the intention), this adaptation of the experiment involved an additional trial of the time-based paradigm.

Method

Participants. Thirty-four undergraduate students participated to partially fulfill a course requirement or for extra credit in a course.

Design. This design was identical to that in Experiment 4 except that the instructions explained fake time in more detail and allowed more trial runs, after which there was ample opportunity to ask questions.

Results

Predictive accuracy of judgments. I computed gamma correlations between judgments and PM performance for each participant for both immediate and delayed judgments. The mean gammas are reported in Table 5. Separate ANOVAs examined

28

each of the three types of correlations (i.e., between judgment and recognition performance, judgment and recall performance and judgment and PM performance).

The ANOVA for gamma correlations between judgment and recognition performance showed no significant effect of judgment type, judgment timing (immediate vs. delayed), but a borderline-significant interaction, $F(1, 18) = 4.19$, $p = .06$. Paired-samples t-tests showed no significant difference between immediate vs. delayed judgments for predicting recognition performance when the judgment was prompted by a JOL-type cue, $t(8) = .41$, $p > .10$, but a significant advantage of delayed judgments in predicting recognition performance when the judgment was prompted with a JOI-type cue, $t(10) = 3.34$, $p < .05$. The t-test comparing gamma correlations between JOLs and JOIs made at a delay was non-significant, $t(21) = 1.29$, $p > .10$.

The ANOVA for gamma correlations between judgment and recall showed no significant effect of judgment timing, or of type of judgment (JOL or JOI), nor a significant interaction.

Likewise, the ANOVA for gamma correlations between judgment and PM performance showed no significant effect of judgment timing, or of type of judgment, nor a significant interaction.

It is possible that the lack of significant effects for recognition and recall is due to a power problem. A small number of trials for each participant (8 each for immediate and delayed judgments) resulted in many ties between judgments, and hence, few valid gamma correlations. Pearson r correlations require assumptions about the properties of the scale on which participants make their judgment that are

not testable in the current paradigm and are also based on a small number of observations (I discuss this shortcoming of the design in further detail below), so while they may overcome gamma's problem with ties, they may not provide much additional information in this analysis. However, it should be noted that Pearson correlations between predictive judgments and performance outcomes (see Table 6) revealed the identical pattern of results as did gamma correlations.

Discussion

Though virutally none of the results from Experiment 5 are significant they never-the-less are informative about whether people are able to make accurate metacognitive judgments about PM tasks and whether the delayed-JOL effect also occurs for JOIs.

The correlations between predictive judgments and recognition performance must be interpreted in light of the fact that the participants never made predictions specifically about the recognition aspect of the test. The difference in correlations for JOLs as compared to JOIs suggests that the type of judgment prompt caused the participants to incorporate different information in their predictions of performance on the target task if they were asked to predict recall as compared to PM. Specifically, they appear to be better able to incorporate predictions of recognition performance when making JOIs at a delay as evidenced by a significant improvement in accuracy if JOI predictions made at a delay, but not of JOLs. Although this experiment does not answer the original question in it's entirety, it provides evidence of different psychological processes for JOLs as compared to JOIs.

Figure 2 plots the data of Experiment 5. It suggests an interesting relationship between judgments of recall and of PM. Though not statistically significant, the trend in this figure appears to be for increased predictive accuracy of both JOLs and JOIs for their target tasks (those depicted by filled shapes and solid lines) when made at a delay following study. The relationship between the JOLs and performance on the PM task and between JOIs and performance on the recall task (not the targets of the predictions; depicted in unfilled shaped with dashed lines) show overall slightly less accurate predictions (not unexpected, since in both cases this is not the aspect of performance they were asked to predict), and possibly a different pattern of results for JOLs and compared to JOIs. This pattern of correlations between recognition and performance suggests that participants attempt to account for predicted performance on only the recall aspect of the test when making JOLs and on both the recognition and recall aspect of the test when making JOIs.

The trends evident in the data from Experiment 5 suggest that individuals are able to predict their performance on a PM task and that, as with recall tasks, these predictions are more accurate when made at a delay following study of the PM task. The next section elaborates the implications of this interpretation in the context of results from previous experiments.

Experiments 3, 4 and 5

Because some of the difficulty in interpreting the results from experiments 3, 4 and 5 may have been due in part to a lack of power, the data from these experiments was combined and analyzed together. This is possible because of the very similar

31

methodology used across the three experiments (the specific differences are noted above). No trials were excluded from the experiment for the purposes of this analysis.

Results

Predictive accuracy of judgments. I computed gamma correlations between judgments and PM performance for each participant for both immediate and delayed judgments. The mean gammas are reported in Table 7. Separate ANOVAs examined each of the three types of correlations (i.e., between judgment and recognition performance, judgment and recall performance and judgment and PM performance).

The ANOVA for gamma correlations between judgment and recognition performance showed a significant effect of judgment timing (immediate vs. delayed), $F(1, 57) = 5.41$, $p < .05$. Neither the effect of judgment type nor the interaction was significant.

The ANOVA for gamma correlations between judgment and recall showed a borderline-significant effect of judgment timing $F(1, 64) = 3.52$, $p = .06$. Neither the effect of judgment type nor the interaction was significant.

The ANOVA for gamma correlations between judgment and PM performance showed a significant effect of judgment timing, $F(1, 87) = 7.31$, $p < .05$. Neither the effect of judgment type nor the interaction was significant.

The combined data of these last three experiments shows a consistent pattern of more accurate predictions for delayed judgments than for immediate judgments, and no consistent difference between JOLs and JOIs.

Distribution of judgments. The proportion of correct and incorrect items assigned each judgment value (0, 20, 40, 60, 80 or 100) are shown in Figures 3

through 5 for recognition, recall and PM, respectively. Note that the distribution of judgments was actually the same in all cases (i.e., participants only made one JOL or JOI per item), however, performance on the criterion test differs among the figures.

ANOVAs were computed for each distribution to compare the proportion of items assigned a judgment. Separate ANOVAs were computed for items depending judgment timing and on whether the item was correct at the criterion test. Judgment type (JOL and JOI) and judgment value were treated as between-subjects variables because not all participants used all values on the scale. Because of the number of unplanned analyses, a Bonferroni correction ($\alpha = .0125$) was used for each set of comparisons to control experiment-wise error.

For items given immediate judgments and recognized correctly, there was a significant effect of judgment value $F(5, 283) = 5.50$, $p < .01$, but no significant main effect of judgment type. The interaction was also not significant. For items given immediate judgments and recognized incorrectly, there were no significant main effects nor a significant interaction. For items given delayed judgments and recognized correctly, there was a significant effect of judgment value $F(5, 297) = 5.23$, $p < .01$, but no significant main effect of judgment type. The interaction was also not significant. Likewise, for items given delayed judgments and recognized incorrectly, there were no significant main effects nor a significant interaction.

For items given immediate judgments and recalled correctly, there was a significant effect of judgment value $F(5, 304) = 4.59$, $p < .01$, but no significant main effect of judgment type. The interaction was also not significant. For items given immediate judgments and recalled incorrectly, there were no significant main effects

33

nor a significant interaction. For items given delayed judgments and recalled correctly, there was a significant effect of judgment value $F(5, 301) = 5.50$, $p <.01$, but no significant main effect of judgment type. The interaction was also not significant. Likewise, for items given delayed judgments and recalled incorrectly, there were no significant main effects nor a significant interaction.

For items given immediate judgments and correct PM performance, there was a significant effect of judgment value $F(5, 265) = 3.65$, $p <.01$, but no significant main effect of judgment type. The interaction was also not significant. For items given immediate judgments and incorrect PM performance, there were no significant main effects nor a significant interaction. For items given delayed judgments and correct PM performance, there was a significant effect of judgment value $F(5, 258) = 3.68$, $p <.01$, but no significant main effect of judgment type. The interaction was also not significant. For items given delayed judgments and incorrect PM performance, there were no significant main effects nor a significant interaction.

Although visual inspection of the distribution curves suggests some small differences in the distribution of judgments for JOLs and JOIs, (specifically a tendency to use the extremes of the scale when making immediate JOLs for correct items), statistical analysis does not bear this out. The only consistent evidence in this analysis is that for correct items, proportion of items assigned a particular judgment increases with judgment value. This does not obtain for incorrect items. This is to be expected and is an indication that participants are able to use the judgment scale when making both JOLs and JOIs.

General Discussion

The results of these experiments, taken as a whole, suggest that participants are able to make moderately accurate predictions of PM performance and that just as do traditional JOLs, these predictions may enjoy some increase in accuracy when made at a delay rather than immediately following study. However, accurate JOIs require a prompt that does not include the recognition target. The first two experiments demonstrated no effect of delay for PM judgments, but in both cases used a judgment prompt that included the recognition target (e.g.,"BREAD – ?" if they were supposed to respond to recognition of the item "BREAD" by pressing the letter J). The latter experiments used a judgment prompt that did not include the recognition target, and showed a trend toward increased accuracy of both JOIs and JOLs when made at a delay following study.

Prompting a metacognitive judgment of performance on a PM task with a cue that includes the recognition target is analogous to asking someone "What are you supposed to do at 11am today?" rather than asking them, "Will you remember the task(s) you intend to do?" Although the former may seem to be a better memory cue as it offers another study trial of the recognition target and provides a rehearsal of recall, the later enables the person to ascertain if both the recognition and recall cues are in memory, and hence, better predict whether they will be able to remember to do both aspects of the PM task. These results suggest that the latter prompt will result in a more accurate judgment of performance on the PM task, especially if the judgment is made after a filled delay following learning the PM task.

This series of experiments certainly does not answer all the questions related to metacognition and PM. It was difficult to design an experiment that would allow participants to make JOIs for cues in which neither the recognition nor recall target are present. Such a task requires a new trial for each item, each with sufficient time to allow for possible forgetting of the PM task. Thus, each participant could run only a few trials, providing virtually no power to detect any effects. Making judgment timing a between-subjects variable is a possible solution, but that eliminates the power inherent to a within-subjects manipulation. This problem might be better solved by finding a way to have participants study and make judgments for multiple cues in a trial. For example, one might have an experiment that includes study, followed either by immediate JOI or JOI after a filled delayed, then a judgment prompted by a cue that does not include the recognition or recall target, with this series repeated for all items in the list before the memory trial begins. The memory trial itself would then contain multiple recognition targets to which participants should respond be recalling the appropriate task or response. This solution introduces another type of variance into the design however, as participants must then make judgments all at once for events or cues that may occur many minutes apart. One might argue that this is often how we make PM judgments in daily life, but it is a significant departure from the usual laboratory paradigm for JOLs in which judgments and test are made in blocks to minimize difference in time for each item between judgment and test.

Another puzzling outcome in this series of experiments is failure to find consistent evidence of the usually robust delayed-JOL effect. One possible reason for

this failure is that Experiments 3 – 5 utilized a time-based PM paradigm. As noted

previously, there is some discussion in the literature regarding whether time- vs.

event-based tasks exist on a continuum or constitute a sharp distinction. There are

persuasive arguments for the idea of a continuum, but also some empirical evidence

that performance differs on the two types of PM tasks, (Kliegel, Martin, McDaniel, &

Giles, 2001; Park, Hertzog, Kidder, Morrell and Mayhorn, 1997). The delayed-JOL

effect may be more difficult to obtain in the time-based paradigm, which may in turn

make a time-based task a less than ideal context in which to test the hypothesis of a

delayed-JOI effect. It is possible to adapt the current paradigm to use the occurrence

of a particular number or letter in the display to replace the advancing fake time as the

recognition target. However, the similarity of a changing number or word to the

advancing time highlights the fuzzy distinction between time- and event-based PM

tasks.

The final three experiments differed from the first two in another possibly

important way. In the first two experiments the PM was embedded in the STM task in

a manner such that some of the STM items were the recognition cues for the PM task.

In the latter three experiments, the cues for the PM task were extraneous to the STM

task and occurred in the form of a changing time display on the screen. It is possible

that this difference also contributed to the lack of evidence for a delayed-JOI effect in

the first two experiments. This hypothesis could be tested by comparing PM

performance when the recognition cue is part of vs. extraneous to the task in which it

is embedded. This design has a real-world parallel in that we sometimes engage in

tasks that themselves contain potential or idiosyncratic recognition targets of our PM

task (e.g., grading a student's paper may remind one of an appointment with that individual), while in other tasks the recognition target may be extraneous to the task (e.g., a different student's paper may not remind us of the appointment, or the time of day may be extraneous to the task even though we intended it to act as a recognition target for the appointment).

It is possible that the performance differences on event- vs. time-based PM tasks is actually due to a difference in embedded vs. extraneous recognition cues. Although it is difficult to imagine a time-based cue that is embedded in the task (i.e., most people must look at the clock to know the exact time) it is easy to think of examples in which recognition cues are embedded in time-based tasks, either by chance (the paper by the student happens to remind you of the appointment) or by design (you intend to use the supermarket billboard on your way home as a reminder to buy bread). It may be the case that time-based cues are more likely to be extraneous to the task, while event-based cues may be embedded in the task on which attention is already focused.

A final possibility for the difficulty in obtaining the delayed-JOL effect may be that because PM tasks occur during some other on-going activity (the STM task in this case), the on-going activity distracts from the recall aspect, or represents a more complex task overall, which may make the delayed-JOL effect more difficult to obtain. This may be because participants are not able to mentally instantiate the context of the on-going task when making their metacognitive judgments. Although there may be costs to forgetting the hypothetical appointment with a student (e.g.,

embarrassment, having to re-schedule), in this paradigm there is no real cost to failing to recognize the recognition cue or failing to recall the correct response.

One issue that future work on metacognition and PM may focus on is the role of costs in forming of PM cues. It is neither possible nor practical to create external cues for all intentions that we create throughout the day, and nearly as impossible to create and retain internal cues for all intentions. To determine how to best allocate memory and attention resources, it is important to be able to judge which intentions require cues and which are sufficiently likely to be recalled without specific cues (possibly because of reliance on habit or belief that idiosyncratic cues will provide sufficient reminders). It is important to determine both how well people judge which items will require an external cue and how people factor in costs (either in terms of the energy required to create the cue or the consequences if the PM task is forgotten) when deciding whether or not to form an external PM cue for a particular item. Although such costs are not intrinsic in the laboratory paradigm, they could be included by tying costs and pay-offs to task performance.

Other questions might address whether predictive accuracy is improved when learners formulate their own PM cues (as compared to those produced by the experimenter) and whether other features of the judgment prompt affect the accuracy of PM judgments. This question also has important implications for everyday PM performance as we sometimes are given external PM cues (e.g., someone adds something to our calendar, or asks us to pick up bread at the grocery store), and other times we form our own cues. It would be interesting to know if people are able to tell

whether an assigned cue is a good onr for them (and therefore rely on it) or a poor

cue, in which case they might reinforce it by forming their own.

It is clear that we rely on memory for intentions in order to get through our

days in both work and personal life. Although little work to date has focused on the

accuracy of metacognitions about intention memory, one's functioning in the world

often relies on whether one will remember an intended task. There is considerable

work to be done in determining the circumstances under which people can best make

predictions about intentions. Addressing these issues will further our understanding of

memory and metamemory.

Footnote

1. Note that the degrees of freedom reported in this analysis are lower than might be

expected given the number of participants. This is because not all participants

contributed gamma correlations to the analysis because of tied judgments or memory

performance that was at floor or ceiling.

Table 1

Experiment 1- Proportion correct (and standard errors) for the PM and STM tasks

	JOL timing			
	Immediate (n = 18)		Delayed (n = 16)	
	3 items	6 items	3 items	6 items
PM memory (hits)	.32 (.07)	.33 (.08)	.46 (.07)	.50 (.08)
Proportion correct on STM target trials	.93 (.03)	.67 (.03)	.94 (.02)	.71 (.03)
Proportion correct on STM lure trials	.95 (.02)	.75 (.04)	.99 (.01)	.74 (.03)

Table 2

Experiment 2 – Mean gamma correlations (and standard errors) between judgments and performance on each aspect of the PM task and the combined (recognition + recall) PM task.

| | Monitoring Condition | | | | | |
| | Recognition | | Recall | | PM (recog+recall) | |
	Immediate Judgments	Delayed Judgments	Immediate Judgments	Delayed Judgments	Immediate Judgments	Delayed Judgments
g(JOI, recog)	.31 (.05)	.37 (.05)	.02 (.04)	.17 (.05)	-.28 (.06)	.25 (.07)
g(JOI, recall)	.28 (.05)	.40 (.04)	.33 (.04)	.86 (.04)	.41 (.06)	.75 (.05)
g(JOI, PM)	.37 (.04)	.53 (.04)	-.02 (.04)	.52 (.04)	-.16 (.06)	.49 (.07)

Note. Shaded areas represent the task for which participants made JOIs in that condition.

Table 3

Experiment 3 - Mean gamma correlations (and standard errors) between judgments and performance on each aspect of the PM task and the PM task.

	Monitoring Condition			
	JOL-type (recog target present)		JOI-type (recog target absent)	
	Immediate Judgments	Delayed Judgments	Immediate Judgments	Delayed Judgments
g(JOL, recog)	.03 (.19)	.12 (.22)	-.11 (.19)	.28 (.18)
g(JOL, recall)	.39 (.17)	.02 (.19)	-.27 (.20)	-.22 (.18)
g(JOL, PM)	.21 (.22)	.29 (.18)	-.14 (.21)	.37 (.19)

Note. Shaded areas highlight the type of metacognitive judgment in that condition.

Table 4

Experiment 4 - Mean gamma correlations (and standard errors) between judgments

and performance on each aspect of the PM task and the PM task.

	Monitoring Condition			
	JOL-type (recog target present)		JOI-type (recog target absent)	
	Immediate Judgments	Delayed Judgments	Immediate Judgments	Delayed Judgments
g(JOL, recog)	-.37 (.21)	.17 (.24)	.47 (.25)	.54 (.25)
g(JOL, recall)	-.02 (.29)	.62 (.16)	.39 (.22)	.39 (.17)
g(JOL, PM)	-.39 (.21)	.43 (.22)	.37 (.17)	.50 (.18)

Note. Shaded areas highlight the type of metacognitive judgment in that condition.

Table 5

Experiment 5 - Mean gamma correlations (and standard errors) between judgments

and performance on each aspect of the PM task and the PM task.

	Monitoring Condition			
	JOL-type (recog target present)		JOI-type (recog target absent)	
	Immediate Judgments	Delayed Judgments	Immediate Judgments	Delayed Judgments
g(JOL, recog)	.35 (.15)	.19 (.25)	.07 (.19)	.55 (.15)
g(JOL, recall)	.30 (.30)	.47 (.16)	.00 (.21)	.19 (.14)
g(JOL, PM)	.27 (.17)	.29 (.15)	.17 (.16)	.37 (.14)

Note. Shaded areas highlight the type of metacognitive judgment in that condition.

Table 6

Experiment 5 – Mean Pearson r correlations (and standard errors) between

judgments and performance on each aspect of the PM task and the PM task.

	Monitoring Condition			
	JOL-type (recog target present)		JOI-type (recog target absent)	
	Immediate Judgments	Delayed Judgments	Immediate Judgments	Delayed Judgments
g(JOL, recog)	.39 (.10)	.25 (.15)	.00 (.15)	.40 (.10)
g(JOL, recall)	.14 (.17)	.24 (.10)	.03 (.10)	.14 (.08)
g(JOL, PM)	.25 (.09)	.25 (.09)	.10 (.09)	.23 (.08)

Note. Shaded areas highlight the type of metacognitive judgment in that condition.

Table 7

Experiments 3, 4 and 5 - Mean gamma correlations (and standard errors) between judgments and performance on each aspect of the PM task and the PM task.

	Monitoring Condition			
	JOL-type		JOI-type	
	Immediate Judgments	Delayed Judgments	Immediate Judgments	Delayed Judgments
g(JOL, recog)	.03 (.11)	.27 (.12)	.12 (.12)	.43 (.11)
g(JOL, recall)	.16 (.15)	.41 (.11)	.06 (.12)	.28 (.09)
g(JOL, PM)	-.01 (.11)	.33 (.10)	.11 (.10)	.41 (.09)

Note. Shaded areas highlight the type of metacognitive judgment in that condition.

Figure 1. Illustration of metacognitive judgments in a JOL paradigm as compared to a JOI paradigm.

Figure 2. Gamma correlations between judgment and recall and judgment and between PM for both JOL- and JOI-type judgment prompts for Experiment 5. Correlations between judgments and the target tasks are depicted by filled shapes and solid lines. The correlations between the JOLs and performance on the PM task and between JOIs and performance on the recall task (not the targets of the predictions) are depicted in unfilled shaped with dashed lines.

Figure 3. Distribution of judgments for immediate and delayed judgments for each judgment value based on whether or not the item was recognized at test.

Figure 4. Distribution of judgments for immediate and delayed judgments for each judgment value based on whether or not the item was recalled at test.

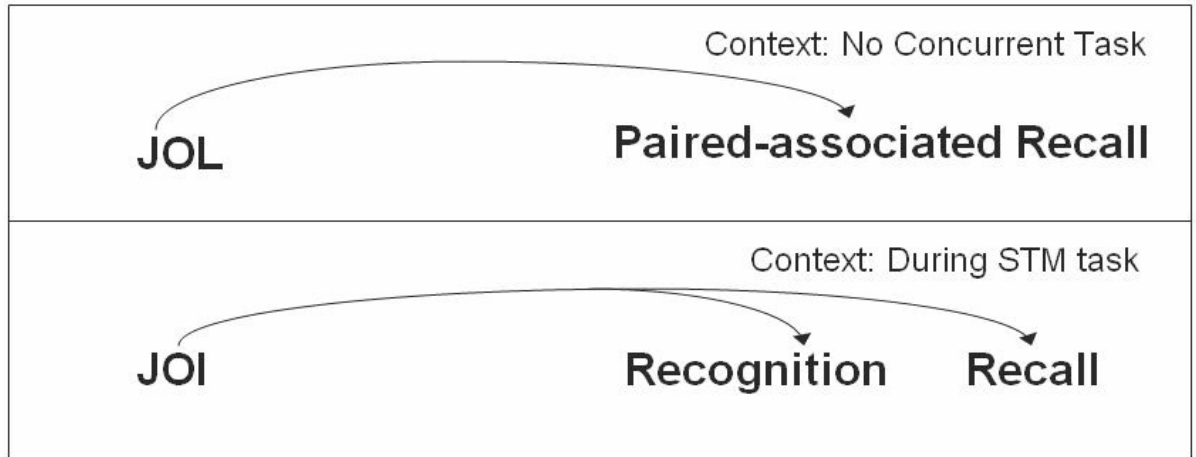*Figure 5.* Distribution of judgments for immediate and delayed judgments for each judgment value based on PM performance at test.
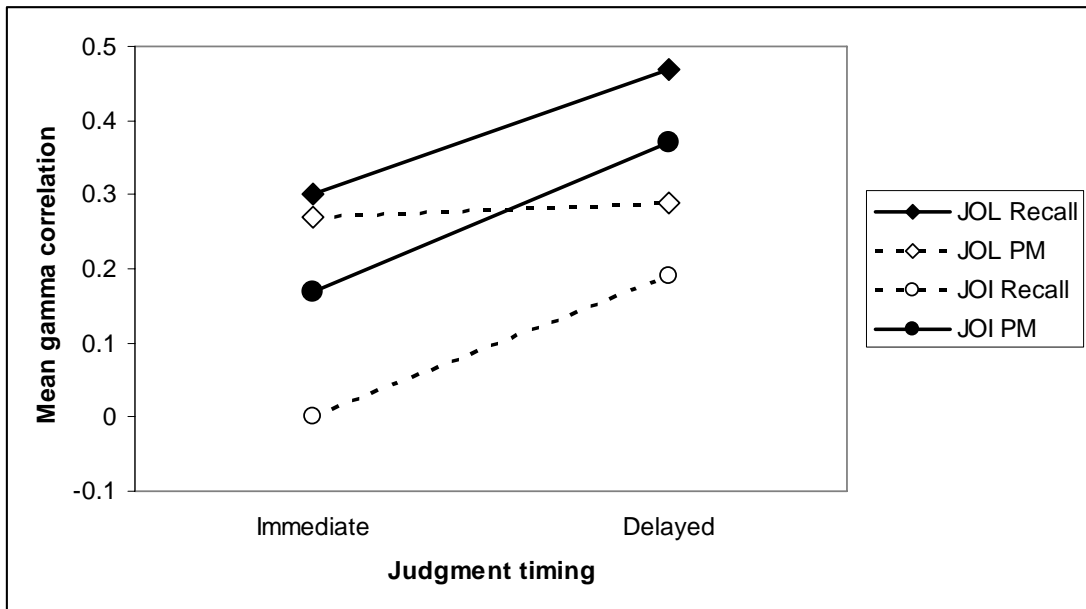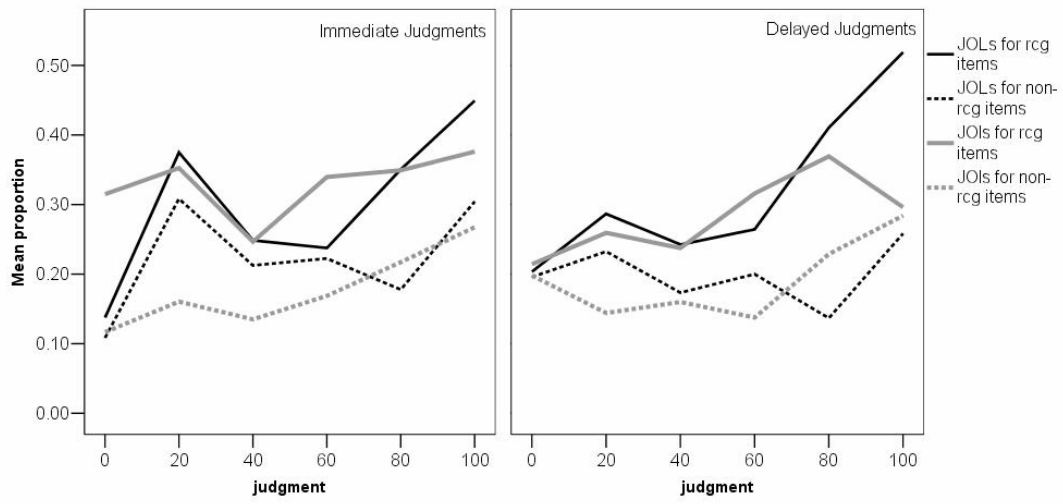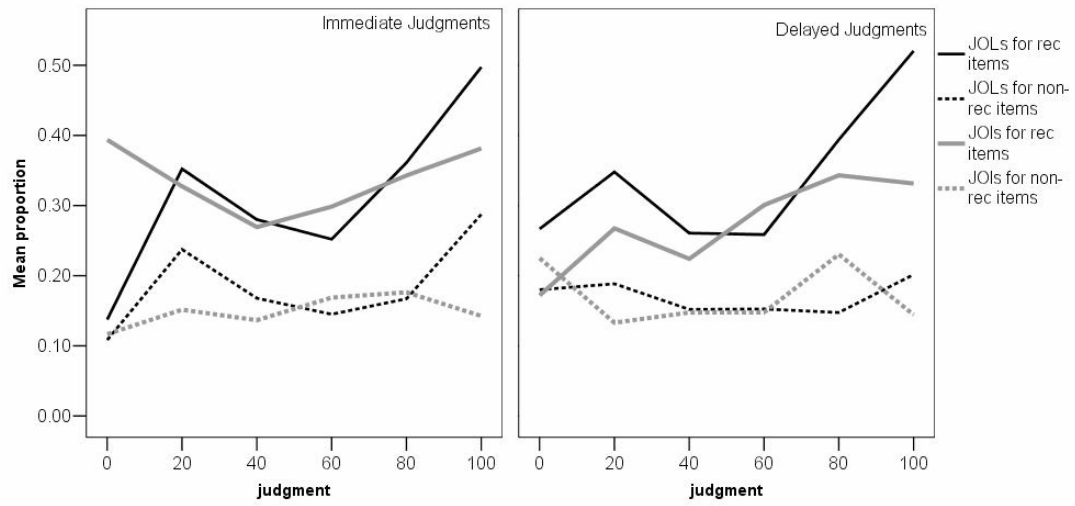
Figure 1

Figure 2

Figure 3

Figure 4
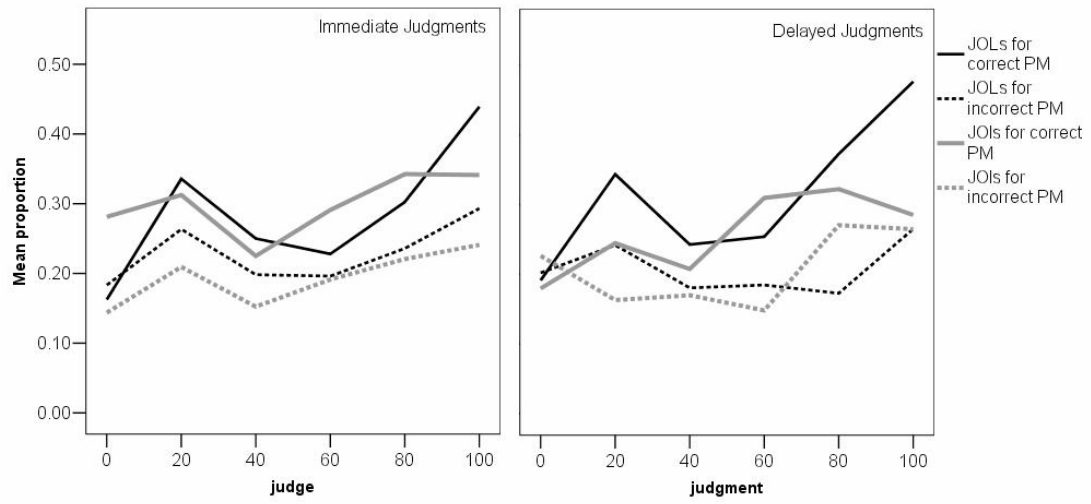
Figure 5

References

Ceci, S. J. & Bronfenbrenner, U. (1985). Don't forget to take the cupcakes out of the

oven. *Child Development, 56,*152-164.

Cohen, A., Dixon, R. A., Lindsay, D. & Masson, M. E. J. (2003). The effect of

perceptual distinctiveness on the prospective and retrospective components of

prospective memory in young and old adults. *Canadian Journal of*

Experimental Psychology 57, 274-289.

Dunlosky, J. & Nelson, T. O. (1992). Importance of the kind of cue for judgments of

learning (JOL) and the delayed-JOL effect. *Memory and Cognition, 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.

Einstein, G. O. & McDaniel, M. A. (1990). Normal aging and prospective memory.

Journal of Experimental Psychology: Learning, Memory, & Cognition,16,

717-726.

Kliegel, M., Martin, M., McDaniel, M. A.& Giles, G. O. (2001). Varying the

importance of a prospective memory task: Differential effects across time- and

event-based prospective memory. *Memory, 9,* 1-11.

Marsh, R. L., Hancock, T. W., & Hicks, J. L. (2002). The demands of an ongoing

activity influence the success of event-based prospective memory.

Psychonomic Bulletin & Review,9, 604-611.

Marsh, R. L., Hicks, J. L., Cook, G. I., Hansen J. S. & Pallos, A. L. (2003).

> Interference to ongoing activities covaries with the characteristics of an event-based intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 861–870.

Maylor, E. A., Smith, G., Della Sala, S, & Logie, R. H. (2002). Prospective and

> retrospective memory in normal aging and dementia: An experimental study. M*emory & Cognition, 30,* 871-884.

McDonald-Miszczak, L., Gould, O. N. & Tychynski, D. (1999). Metamemory

> predictors of prospective and retrospective memory performance. *Journal of General Psychology, 126,* 37-52.

Meacham, J. A. & Singer, J. (1977). Incentive effects in prospective remembering.

> *Journal of Psychology, 97,* 191-197.

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.

Nelson, T. O. & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-

> learning-effect? *Psychological Science, 317-318.*

Park, D.C., Hertzog, C., Kidder, D. P., Morrell, R. W. & Mayhorn, C. B. (1997).

> Effect of age on even-based and time-based prospective memory. *Psychological Aging, 12,* 314-327.

Reese, M. & Cherry, K. E. (2002). The effects of age, ability and memory monitoring

> on prospective memory task performance. *Aging, Neuropsychology, & Cognition, 9,* 98 – 113.

Roche, N. L., Fleming, J. M. & Shum, D. H. K. (2002). Self-awareness of prospective

    memory failure in adults with traumatic brain injury. *Brain Injury, 16,* 931-

    945.

Spellman, B. A. & Bjork, R. A. (1992). When predictions create reality: Judgments of

    learning may alter what they are intended to assess. *Psychological Science, 3,*

    513-316.

Smith, R. E. (2003). The cost of remembering to remember in event-based

    prospective memory: Investigating the capacity demands of delayed intention

    performance. *Journal of Experimental Psychology: Learning, Memory, &*

    Cognition, 29, 347-361.

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.

Vogels, W. W. A., Dekker, M. R., Brouwer, W. H. and de Jong, R. (2002). Age-

    Related changes in event-related prospective memory performance: A

    comparison of four prospective memory tasks. *Brain and Cognition 49,* 341 –

    362.