Monitoring in event-based prospective memory tasks

Dale Watkins Cockman
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses

Part of the Psychology Commons

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_theses/50

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master’s Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
MONITORING IN EVENT-BASED PROSPECTIVE MEMORY TASKS

A Thesis
Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

in

The Department of Psychology

by

Dale Watkins Cockman
B.A., University of North Carolina at Greensboro, 1998
December 2003
Table of Contents

List of Tables......................................................iii
Abstract......................................................................iv
Introduction..........................................................1
Overview of Pilot Experiment....................................12
Method for Pilot Study..............................................15
  Participants .....................................................15
  Materials and Equipment .....................................15
  Procedures ....................................................15
Results and Discussion of Pilot Study .........................18
  Results ..................................................................18
  Discussion .......................................................20
Overview of Experiment 1.........................................23
Method for Experiment 1 ..........................................26
  Participants ......................................................26
  Materials, Equipment, and Procedures ....................26
Results and Discussion of Experiment 1 .....................27
  Correlational Analyses ....................................32
Overview of Experiment 2.........................................37
Method for Experiment 2 ..........................................39
  Participants ......................................................39
  Materials and Equipment .....................................39
  Procedures ....................................................40
Results and Discussion of Experiment 2 .....................42
  Correlational Analyses ....................................45
General Discussion ..................................................51
References ..........................................................58
End Notes..................................................................60
Appendix A: Post Experimental Questionnaire ..............61
Appendix B: General Knowledge Questions for Experiment 2..63
Vita..........................................................................66
List of Tables

1. Response Latencies to LDT by Learn Type and Condition in the Pilot Study..........................19

2. Correlations Between PM Accuracy and Response Latencies on LDT in the Pilot Study..............21

3. Response Latencies to LDT by Learn Type and Condition in Experiment 1.............................29

4. ANCOVA Adjusted Response Latencies to LDT by Learn Type and Condition in Experiment 1.............30

5. Prospective Memory Performance and Confidence Ratings in Experiment 1................................32

6. Correlations in the Distinctive Experimental Condition in Experiment 1.................................34

7. Correlations in the Organizational Experimental Condition in Experiment 1..............................35

8. Prospective Memory Performance and Confidence Ratings in Experiment 2...............................45

9. Response Latencies to LDT by Learn Type and Condition in Experiment 2.............................46

10. ANCOVA Adjusted Response Latencies to LDT by Learn Type and Condition in Experiment 2.........47

11. Correlations in the Easy Experimental Condition in Experiment 2........................................48

12. Correlations in the Moderately Difficult Experimental Condition in Experiment 2.......................49
Abstract

Event based prospective memory (PM) refers to remembering to perform a particular action upon the presentation of a particular cue in the environment. Until recently, most models of event-based PM performance have suggested that the realization of the target event occurs automatically. The DARC model (Smith, 2000) is among the first to suggest that monitoring is required to notice the target event, in the form of a consistent, non-strategic dedication of resources. The predictions of the DARC model are contrasted with those of Einstein & McDaniel (Noticing + Search, 1996), Goschke & Kuhl (1996), Ellis (1996). The pilot study and experiment one test the idea that items distinctively encoded will be more memorable and more fluently processed, leading to better PM performance during the target task. Pilot data suggest that less monitoring is engaged when target items are more memorable. Experiment 1 attempted to replicate that finding and included a direct measure of retrieval fluency. Faster retrieval was associated with better PM performance in the distinctive condition. However, the same did not hold in the organizational condition. Experiment two manipulates the retrieval fluency of the target events when produced as answers to general knowledge questions (Benjamin, Bjork, & Schwartz,
Target events more quickly generated at the time of target event encoding were expected to be associated with higher confidence that the target event will be recognizable, which should lead to less monitoring. In this case retrieval fluency would be misleading as an index of the need to monitor for the target items. Retrieval fluency did not reliably predict LDT performance in Experiment 2. Results of both experiments are discussed in light of the above mentioned models and McDaniel & Einstein’s multiprocess framework (2001). Results are consistent with the notion that automatic and controlled processes are involved in the realization of an intention in an event based PM task.
Introduction

Retrospective memory refers to the retrieval of one’s prior experiences. For example, remembering what one had for dinner two nights ago or answering a trivia question both require the retrieval of past experience. Although the vast majority of experimental research on human memory concerns this form of memory (hereafter denoted RM), only recently has research been devoted to the more practical aspect of how memory is used to fulfill goals. In other words, memory is not used only for the passive storage and retrieval of past experiences, but also for the prediction and regulation of future experiences. Prospective memory (hereafter PM; Brandimonte, Einstein, & McDaniel, 1996) is one example of just such a practical use of memory. PM refers to situations in which people have to carry out a delayed intention (Ellis, 1996). That is, one must retrieve a previously established intention (e.g., give a colleague a message) at some later point in time. Many intentions are delayed because they cannot be carried out immediately. For example, the colleague to whom a message must be delivered is out of the office, necessitating one to establish the intention to deliver the message later. Of course, when the colleague returns, one must realize that an intention had been established, followed by retrieval and delivery of the message. This realization must often occur while one is engaged in some other primary activity that may be unrelated to the intention.
Clearly, PM involves a retrospective component. The intention itself is a retrospective memory in the form of a thought to oneself, or a request from another (e.g., Ellis, 1996; Goschke & Kuhl, 1993). But the realization and retrieval of that memory is the “prospective” component. As such, PM is often viewed as having two general components: retrospective (content) and prospective (intent). One important dichotomy that has been introduced in the PM literature is that of time-based and event-based activities (e.g., Einstein, McDaniel, Richardson, Guynn, & Cunfer, 1995). Time-based PM tasks are those that involve remembering to perform an intended action at a certain time, or after a particular amount of time has passed (e.g. remembering to take medicine or to make a phone call at 4pm). Event-based PM tasks, the type of primary interest here, involve remembering to perform an intended action when an appropriate environmental cue is encountered. Examples of event-based PM tasks include remembering to deliver a message to a friend, or remembering to buy bread on the way home from work.

Recent theoretical work has seen the development of preliminary models of event-based PM. One early example of such a model is Einstein and McDaniel’s (1996) noticing + search model. This model characterizes the retrieval of an event-based intention as following two separate stages: noticing the familiarity that arises from perceiving the cue, and then searching for the significance, or relevance, of that cue (i.e., that it is associated with an intention). The noticing stage is assumed to be automatic—that is,
one does not need to actively search for the cue, nor is remembering
the specific intention associated with that cue relevant to noticing
it in the first place. Once the cue has been noticed, a controlled
search for the significance of the cue ensues. One implication of
this model is that the cue must be powerful enough to elicit
noticing. If the event goes unnoticed, then no search will take
place, and the intended action will not be carried out. PM failure
may also occur when the directed search fails. That is, the search
for the significance of the noticed event may fail.

Based on this model, manipulating the distinctiveness or the
familiarity of the target event relative to its local context should
influence PM performance. McDaniel and Einstein (1993) demonstrated
that characteristics of the target event can be manipulated to make
the target more or less noticeable. McDaniel and Einstein (1993,
Exp 2) demonstrated this by asking participants to study and
immediately recall lists of 6 words, and to press a key whenever a
target word occurred. In this example, the short-term recall task
is used to represent the “ongoing” activity in which one is usually
engaged. Pressing the key on a keyboard when a target word appears
represents the prospective activity, and it occurs only a few times
throughout many trials of the ongoing activity. Variations on this
method constitute the basic event-based PM paradigm. Familiar and
unfamiliar words were drawn from published norms (Toglia and Battig,
1978). Examples of familiar words included targets fuse and movie,
unfamiliar words included targets sone and yolif. Word
distinctiveness was defined by the target word’s familiarity relative to the majority of the items in each list. That is, participants in the distinctive condition were given a target word dissimilar in familiarity relative to the other list items. PM performance benefited from lower cue familiarity (mean PM accuracy for unfamiliar targets = .95 vs. .55 for familiar targets) and from distinctiveness relative to local context (mean PM accuracy for distinctive condition = .89 vs. .60 for the non-distinctive condition).

A role for conceptual processing of the target event was implicated in a study by McDaniel, Robinson-Riegler, and Einstein (1998, Experiment 1). Prospective memory performance was higher in a condition where the targeted meaning of homographic words was held constant between formation of the intention and later perception of the cue as compared to a condition in which the targeted meaning was different at test than when the intention was established. A levels of processing effect was demonstrated in experiment 3, in which PM performance was higher when target items were studied semantically (i.e. by generating an adjective to the studied item) rather than non-semantically (i.e. by generating a rhyme for the studied item). Thus, it appears that PM performance is influenced by qualities of the cue itself (e.g. familiarity, distinctiveness relative to local context) as well as by the processing (at encoding and retrieval) performed on the cue.
One implication of the Noticing + Search model (Einstein & McDaniel, 1996) is that monitoring for the target event is not necessary. That is, when the task is to deliver a message to a colleague, it is not necessary to monitor for that colleague between the encoding and retrieval of the intention. Rather, when that colleague is present, a directed search should automatically be initiated to determine the significance of “noticing” that colleague. This is consistent with the intuition that event-based PM tasks are used by people specifically so that the intention can be kept “out of mind” while performing other important activities. A cue might even be selected to conform to one’s opinion that it will be easy to notice when it appears at some later point in time.

Ellis (1996) also suggests the role of an automatic component in the bringing to mind of the delayed intention. Her framework draws upon the distinction between brute and hierarchical retrieval (Tulving, 1983). Hierarchical retrieval (in the context of PM) depends on integrating an intention into an already existing hierarchy, such as a daily routine. Brute retrieval is the retrieval of an item “through its own merits”. Ellis suggests that brute retrieval is in operation in most event-based PM tasks—the presence of the target event itself is sufficient to cue the intention.

Ellis’s (1996) model does not, however, entirely close the door with regard to monitoring for the PM cue. Her framework divides PM tasks into five phases: a) encoding of intention and action; b) retention interval; c) performance interval; d) initiation and
execution of intended action; e) evaluation of outcome. In the case of shorter term PM tasks, the intention may be maintained in consciousness during the retention interval, until an opportunity to carry out the intention arises (i.e. the performance interval). These sorts of tasks are essentially vigilance tasks. That is, one consistently and actively searches for an opportunity to carry out the intention. In longer term PM tasks, the intention is thought to leave consciousness for a period of time, and in these situations, realization of the delayed intention is largely independent of monitoring for the PM cue—conscious capacity is not required during the retention interval.

Goschke and Kuhl (1996) suggest that cognitive resources may or may not play a role in the realization of a delayed intention, depending on the demands of the particular task. When the intended action is simple and well specified, and when the cue is well defined, realization of the intention can likely rely solely on the increased level of activation associated with the encoded intention and action schema. This is thought to be the case when the intended action is routine, or even when a declarative representation of the intended action is relied upon. One interesting aspect of these ideas is that event-based PM and time-based PM are not entirely separated from one another. Instead, a different type of persistence is assumed to be active depending on the type of task, but the same general framework applies regardless of the task type. As just mentioned, monitoring may not be necessary in certain cases—
when the PM cue and intended action are well specified and noticeable. When the PM cue is less noticeable, self-initiated retrieval of the PM cue is often necessary in order for the intended action to be carried out. For example, the PM cue in a time-based PM task (e.g. when the clock reads 4:00 pm) will not be noticed unless one looks at a clock. In this type of task, monitoring the clock, particularly as the target time gets closer, is a reliable predictor of successful PM performance (Einstein, McDaniel, Smith, & Shaw, 1998). Given the correlation between monitoring and performance in a time based task, it seems plausible that some event-based PM tasks may also rely on or at least benefit from monitoring, such as in instances where one believes the cue to be poorly specified or difficult to notice.

Results obtained by some (e.g., Marsh & Hicks, 1998; McDaniel et al., 1998) in which dividing attention during the retrieval phase of a PM task decreased performance, suggest an involvement of controlled retrieval processes in PM. That is not to say that one is necessarily consciously monitoring for the cue, only to say that some sort of attentional capacity is necessary at least when the cue is encountered. Interestingly, one recent model, the DARC model (Decision Activity, Recollection and Consciousness; Smith, 1999), suggests that some constant level of resources is required in monitoring for the cue, in the form of a general monitoring hypothesis. Smith’s model predicts that attentional capacity is required “for making decisions about how to interpret and respond to
the environment before the environmental cue event occurs” (Smith, 1999, p.8). In other words, the mere establishment of an intention requires some amount of cognitive resources in the form of monitoring during any subsequent ongoing activity. If resources are dedicated to monitoring for the target event once an intention has been encoded, then performance on ongoing tasks should suffer relative to when no intention has been encoded. The DARC model predicts that the PM task will be successfully completed to the extent that monitoring occurs. That is, the amount of monitoring for the target during an ongoing task, as measured by the cost to an ongoing task, should be positively correlated with PM performance.

In her third experiment, Smith (1999) asked participants to study six target items that would serve as cues for a later PM task. Once the six target items were successfully memorized, participants began a lexical decision task (LDT) with either embedded or delayed PM instructions. The LDT required a word/non-word decision for letter strings presented one at a time. Those in the embedded instruction condition were told that any time they encountered one of the six target words during the LDT they should press the F1 key instead of making their LDT response. Participants in the delayed condition were instructed that they should remember to press the F1 key when one of the target words appeared, but that they did not have to do this during the LDT and would instead be expected to press F1 at some later time in the experiment.
If capacity is involved in monitoring for a PM target, then it should, on average, take longer to make lexical decisions in the embedded condition than in the delayed condition. In fact, lexical decision latencies were greater in the embedded condition than in the delayed condition, presumably as a result of monitoring required in the embedded condition that was not present in the delayed condition. Additionally, more participants in the embedded condition whose PM performance was at or above the mean for that group had slower reaction times in making their LDT responses than participants whose performance was below the mean. These results were interpreted as consistent with a general monitoring hypothesis in which dedication of more capacity to monitoring for the target events led to greater performance at the expense of performance on the LDT.

It is plausible that while some capacity-consuming monitoring does occur during the delay between encoding the intention and encountering the target event in the environment, such monitoring need not necessarily take place during the entire period of time, nor should the same amount of monitoring be consistently applied during that time. For example, deciding in the morning to buy bread and milk on the way home from work in the evening may not require any rehearsal of the intention or monitoring for the cue (the grocery store) during the work day. Monitoring for the grocery store may not take place at all until the drive home has begun. Or, perhaps, one might periodically remind himself throughout the day of
the intention to shop for groceries after work, “maintaining the activation of the cue-intention association so that it is more readily activated when the triggering event occurs” (McDaniel & Einstein, in press). For example, McDaniel and Einstein (1993, experiment 2) demonstrated that the specificity of instructions influences performance on a PM task. Participants instructed to respond to specific words (leopard, lion, and tiger) performed better than those instructed to make their PM response whenever they saw an instance of an animal. One possibility, although speculative, is that the specificity of instructions influenced the quality or quantity of monitoring during the ongoing task. In other words, performance by participants with specific instructions may have been better because they dedicated more resources to monitoring for the targets, an interpretation consistent with the DARC model.

However, another possibility is that participants receiving specific instructions actually monitored less than those receiving general instructions. It may be that participants receiving specific instructions believed the targets would be more recognizable than did participants instructed to make their response whenever they saw any instance of an animal. This oversimplification is not intended to suggest that resources dedicated to monitoring are strategically applied in a conscious manner — monitoring may be determined outside of awareness by factors such as perceptual distinctiveness, retrieval fluency, or any number of other factors. Therefore, a pilot study was conducted
to both replicate Smith’s (1999) original finding of slowed LDT latencies, and also to examine the generality of those results. If the mere existence of an intention taxes resources, then the manner in which the intention is learned should not change the level of monitoring required. However, if the monitoring is more context-specific or flexible, then the level of monitoring applied may depend on how people perceive the relative difficulty of the PM task itself.
Overview of Pilot Experiment

The DARC model predicts that performance on a PM task will improve to the extent that capacity is made available for monitoring. Given the observed relationship between monitoring and background task performance, and the predictions of the DARC model, one would expect PM performance to correlate positively with latencies on an LDT. That is, PM performance improves as a result of increased monitoring, which comes about as a result of reallocating capacity from the ongoing task (LDT) to monitoring for the PM target. If this prediction is true, then manipulating the initial encoding, and therefore the perceived memorability, of PM targets should not influence the amount of capacity dedicated to monitoring. Smith’s (1999) model suggests only a very general monitoring hypothesis. In other words, the presence of a delayed intention supports a strategic reallocation of attentional capacity in order to monitor for an event-based target. However, the nature of this monitoring strategy was not well-specified by Smith. She compared an event-based prospective condition only with a control condition that did not have any such intention.

Consideration of the numerous factors that surround intention formation suggests that monitoring may be more or less likely depending on the nature of the PM targets expected. One such factor may be the perceived likelihood of noticing the targets at some later time. For example, if one believes that an event-based target will be readily perceived or noticed, conscious allocation of
attention to monitoring for those events should be less useful. Alternatively, if one believes that an event-based cue will be difficult to notice in the face of ongoing activity, then such monitoring may be more likely. The real or perceived memorability of the event-based targets is therefore an important factor that may affect the degree to which people allocate fixed attentional resources toward monitoring. Well-known or easily retrieved targets may be accompanied by a sense of confidence that those items more likely to be noticed or retrieved when experienced at some future time (Benjamin, Bjork, & Schwartz, 1998). However, targets that are relatively more difficult to retrieve initially may produce a low sense of confidence in future memorability, therefore increasing the likelihood that a monitoring strategy will be used.

Previous research in the retrospective memory literature has shown that utilizing a combination of organizational and distinctive information about a given stimulus at encoding should lead to greater memorability of the item than the use of organizational encoding alone. In one study (Hunt & Smith, 1996), participants were presented with a list of items. For each item on the list, they were asked to generate either one organizational or one distinctive cue. An organizational cue is one that is based on the similarity of the target item to other items on the list. A distinctive cue is one based on some difference between the target item and the other items on the list. In their study, participants studied lists containing five categorized words by writing one word
associated with each of the categorized words that was not true of the other four. Twenty lists were learned for a total of 100 words. Participants successfully recalled the original word when cued with the distinctive cue they had generated earlier 97% of the time. When participants generated an organizational word at study, cued recall performance was much poorer (.59).

Manipulating the perceived memorability of PM targets by having an individual generate a distinctive cue in response to a presented category exemplar is beneficial because both organizational and distinctive processing are engaged at encoding. When only organizational processing is engaged, recall performance suffers (Hunt & Smith, 1998). The availability of organizational and distinctive processing at encoding should not only lead to better retrospective memory, but also to better PM performance than when the item is studied by generating an organizational cue. The following pilot experiment is a both a test of this idea and a general replication of Smith’s (1999) third experiment.
Method for Pilot Study

Participants

125 LSU students were randomly assigned to either an experimental (i.e. embedded) or control (i.e. delayed) PM condition. Participants received extra credit in undergraduate psychology courses for their participation.

Materials and Equipment

Four categorized word lists of five items each were compiled from category norms (Battig & Montague, 1969) for the study portion of the experiment. All instructions were presented on IBM-compatible PC’s. The learning portion of the experiment (stimulus presentation, cue generation and category-cued recall test) was completed using Micro Experimental Laboratory Professional software package (Schneider, 1988). The LDT was accomplished with a program written in Turbo Pascal, and consisted of 300 letter strings (150 words, 150 pronounceable non-words). Five words were selected as control items, as in Smith (1999), and five were PM targets (sports: golf, hockey, racing, boating, soccer; or weapons: bomb, club, gun, knife, rifle). PM targets and control items appeared every 50 trials, PM targets beginning with trial #70, control items beginning with trial #80.

Procedures

Participants were tested in groups of one to four, and the session lasted approximately one-half hour. Instructions were given that the participant would see five items on the computer screen at
once, all from the same category. Participants were instructed to try to remember the words for a later memory test. Those in the organizational condition were asked to type in one thing about the item at the top of the list that made it similar to the other four items on the list. All five words in each list appeared at the top of the list exactly once. Once a cue had been generated for each of the five category exemplars, a new category list was presented. This continued until a cue was generated for all five items in all four lists. Those in the distinctive condition did the same, with the exception of the type of cue they were asked to generate. Study instructions for the distinctive condition were to type in one thing about the item at the top of the list that made it different from the other four items on the list.

The target list (weapons or sports) was always studied last. Immediately following the study portion of the experiment, instructions for a category-cued recall test were given, and category labels were presented for recall in the same order as the lists had been studied. Once recall was complete, participants read instructions for the LDT, and in the embedded group, for the PM task. All were instructed that they would see strings of letters on the computer screen, and that they should press the key marked ‘Y’ if the string was a word, and the key marked ‘N’ if the string was not a word. The Y/N response was to be made as quickly and accurately as possible. After each response was made, a screen that read “waiting” was presented. This remained on screen until the
participant pressed the space bar to begin the next trial. Before beginning the LDT, those in the control group were asked to think back to the target list they just recalled, and were informed that they would be asked to remember those words later in the experiment.

Those in the experimental group were given the same instructions for the LDT, except they were instructed to press the forward slash (‘/’) key after they made their lexical decision whenever they encountered one of the target words (i.e., sports or weapons exemplars), but before they pressed the space bar to begin the next trial. Once the LDT was completed, participants were both debriefed and asked to recall the 5 target items.
Results and Discussion of Pilot Study

Results

All differences reported were significant at alpha = .05. Any participants unable to recall either four or five of the PM cues at initial recall were removed from all analyses, leaving 61 participants in the PM groups (32 distinctive, 29 organizational) and 47 participants in the control groups (24 distinctive, 23 organizational). Initial analyses were conducted to determine any differences in latencies between words and the five control words. No differences were found, and all words (five control and 140 remaining non-target words) were aggregated for subsequent analyses. Prospective memory performance did not differ significantly by t-test for the organizational and distinctive learning groups, whose mean performance was .71 and .61, respectively.

LDT data were entered into a 2 x 2 x 2 (item type = word/non-word, condition = PM/control, learn = organizational/distinctive) ANOVA and are summarized in Table 1. Words were identified correctly more quickly than non-words, F (1, 104) = 111.6, MSE = 1822740. Those in the experimental condition made lexical responses more slowly than participants in the control condition F (1, 104) = 7.471, MSE = 393144. However, the learn by condition interaction was also significant, F (1, 104) = 5.029, MSE = 264626. The pattern of latencies in Table 1 suggests that the interaction was driven by greater latencies when PM targets were studied by generating
Table 1.

Response Latencies to LDT by Learn Type and Condition in Pilot Study.

<table>
<thead>
<tr>
<th>Learn Type</th>
<th>Words</th>
<th>Nonwords</th>
<th>Words</th>
<th>Nonwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinctive</td>
<td>756.65</td>
<td>-18.69</td>
<td>914.88</td>
<td>-36.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>706.63</td>
<td>-29.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>933.18</td>
<td>-72.69</td>
</tr>
<tr>
<td>Organizational</td>
<td>838.62</td>
<td>-31.32</td>
<td>1056.02</td>
<td>-48.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>707.28</td>
<td>-19.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>865.8</td>
<td>-33.77</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses represent standard errors.
organizational cues compared to latencies in the control condition.
When targets were studied by generating distinctive cues, mean
latency to LDT responses in the experimental condition was only
slightly greater than in the control condition. Separate 2 x 2
(learn x item type) ANOVAs were run for each condition to confirm
the nature of the interaction. The ANOVA including PM subjects
revealed a main effect of learning, $F(1, 59) = 6.737, \text{MSE} = 53779$, whereas an ANOVA on the control participants yielded no
effect of learning, $F(1, 45) = 0.484, \text{MSE} = 51105.1$. The
interaction was further clarified by a 2 x 2 (item type x condition)
ANOVA for each learning group. The main effect of condition found
for participants in the organizational group $F(1,51) = 12.279, \text{MSE} = 48175$ was absent in the distinctive group $F(1, 53) = .525, \text{MSE} = 46024$.

Correlations of prospective memory performance and response
time on the LDT for each learning group were calculated, and are
summarized in Table 2. The only significant correlation between PM
performance and response time was in the organizational learning
condition ($0.425$). The correlations were not consistent with a
general monitoring hypothesis, as they were not significant for non-
words.

Discussion

The pattern of results obtained for the LDT in the
experimental condition is clearly inconsistent with the Noticing +
Search model (Einstein & McDaniels, 1996), which would not predict
Table 2.

Correlations Between PM Accuracy and Response Latencies on the LDT in the Pilot Study

<table>
<thead>
<tr>
<th>Learn type</th>
<th>Words</th>
<th>Non-Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinct</td>
<td>0.279</td>
<td>0.079</td>
</tr>
<tr>
<td>Organiz.</td>
<td>0.425*</td>
<td>0.077</td>
</tr>
</tbody>
</table>

* Indicates significance at alpha = .05

...slowing. The short delay between establishing the intention and entering the performance interval would lead Ellis’s (1996) framework to predict an overall slowing in the experimental condition due to the intention remaining in consciousness. However, the tendency for only experimental participants in the organizational learning group to show significant slowing would not be predicted. Because both the target and the intended action were equally well specified in both the organizational and distinctive learning conditions, Goschke and Kuhl (1996) would predict no slowing in either learning condition relative to control participants. The presence of a significant slowing in one of the experimental conditions is consistent with Smith’s (1999) monitoring hypothesis. However, several aspects of the data seem to qualify this general prediction. First, the slowing in the experimental condition as compared with the control condition was significant only for the organizational learning groups. This suggests that the
quality of initial learning of PM targets affected the likelihood of a monitoring strategy. Second, the group with the greatest amount of monitoring, as measured in cost to LDT latencies, had numerically lower PM accuracy. The presence, or degree, of monitoring therefore does not guarantee better noticing of relevant PM targets in the ongoing task. Third, PM performance correlated significantly with LDT latency only for the organizational learning group, providing further evidence for a context-specific form of monitoring.

One inexplicable aspect of the correlational analyses was the presence of near-zero correlations for non-word latencies and PM performance, even in the organizational learning condition. Thus, the interpretation of the slowing in this condition as compared with the control condition as a “check” for PM targets on every trial seems suspect. An adequate explanation as to why the correlations are significant for words, but not for non-words, is not immediately available. Perhaps the degree of monitoring is supported by the presence of certain features in the LDT stimuli that overlap with aspects of the PM targets (e.g., semantic or lexical features). In other words, monitoring would not even be useful if a letter string is not categorized as a word in the first place. Although speculative at this juncture, this argument characterizes the monitoring process as far less general than argued by Smith (1999).
Overview of Experiment 1

The finding of a context-specific LDT slowing for the organizational learning experimental condition, and the speculations concerning the reasons for that slowing, suggested modifications to the procedure. Unfortunately, there is no direct evidence that what caused the slowing was a reallocation of attentional resources to monitor for PM targets. Although PM performance was numerically better for the cues learned with distinctive encoding, as was expected from the retrospective literature, there was no evidence that participants felt differently about the degree of learning in the distinctive versus organizational conditions. Therefore, several aspects of the procedure were modified for Experiment 1 to remedy these problems.

Benjamin, Bjork, & Schwartz (1998) have indicated retrieval fluency as a primary metamnemonic index of perceived memorability. That is, the more quickly an item can be retrieved, the more confident one is that memory for that item will be available later. In their Experiment 1, Benjamin et al. had participants generate answers to a series of general knowledge questions. Instructions indicated that the variable of interest was the time it took to answer each question, and that they should therefore press the ‘enter’ key immediately upon generating an answer to a particular question, but not before. Participants were then asked to predict the likelihood of later free recall for the answer to the question. Higher predictions of recall were associated with faster latencies
to generating an initial answer. Interestingly, the more quickly retrieved answers were associated with poorer objective free recall performance as compared with more slowly retrieved answers.

Benjamin et al. (1998) therefore demonstrated that the speed with which an item can be retrieved from memory leads one to be more confident about the likelihood of subsequent recall, but that those items are actually associated with poorer recall performance. Thus, retrieval fluency does serve as a metamnemonic index, but as shown by Benjamin et al. and further explored in proposed experiment 2, retrieval fluency can sometimes be a misleading index. If retrieval fluency serves as a metacognitive index for one’s monitoring strategy in a PM task, using the same paradigm as in the pilot study, then the more quickly targets are retrieved before the LDT begins, the less monitoring will be engaged during the LDT. That is, fluency of retrieval of target items may lead to more or less monitoring. One could argue that the difference in monitoring observed in the pilot study is simply a result of poorer memory in the organizational group for the PM targets at the beginning of the LDT. However, removing participants who remembered fewer than three of the five target items at initial recall did not change any of the results reported above. Cued-recall retrieval time was included in experiment one as a measure of retrieval fluency.

In addition to attempting to discern the potential role of retrieval fluency, other modifications were made to obtain more direct evidence for a monitoring strategy. A comprehensive post-
experimental questionnaire assessed memory for the PM targets and participants’ subjective awareness concerning their memory for the targets. Recall of the 5 targets was solicited as in the pilot study. However, a more liberal recognition test followed the recall test. Given that recognition of the target as relevant to a delayed intention is presumably what occurs during the ongoing task, a recognition test might be the most appropriate post-experimental way to measure retrospective memory for the PM targets. Furthermore, questions concerning the participants’ awareness of strategies they used were be recorded in hope that the answers to these questions may reveal how participants felt about the importance of the PM task in general (and relative to the LDT), any subjective awareness concerning how memorable they thought the targets were, how often they thought about the delayed intention, and any other strategies they reported using. Data belonging to participants whose pre-LDT recall is less than 80% will be removed from the study. Analyses will be conducted both with and without data belonging to participants whose post-LDT recall or recognition is less than 80%. Finally, a practice 100-item LDT will be given at the outset to establish that no group differences exist prior to the manipulation of interest.
Method for Experiment 1

Participants

144 participants were randomly assigned to either an experimental (i.e. embedded) or control (i.e. delayed) condition. Participants received extra credit in undergraduate psychology courses for their participation.

Materials, Equipment, and Procedures

An initial LDT consisting of 50 words and 50 non-words not used in the target LDT was given before the categorized lists were learned. Otherwise, materials and equipment were the same as in the pilot study, except that the program was changed to record latencies to recall the PM targets during the initial recall of target items, and the “control” items used in the pilot study were excluded, as they are redundant with the other words in the LDT. For initial retrieval, participants were presented with a category cue and a cursor. Retrieval time is defined as the amount of time that passes between cue and cursor presentation and the first key press of the response. Each participant filled out the post-experimental questionnaire (Appendix A) after the LDT and before presentation of the debriefing. Otherwise, the procedure for the experiment was identical to that of the pilot study.
Results and Discussion of Experiment 1

All differences reported were significant at alpha = .05. Data were analyzed only for those participants who recalled at least three of the five target words during the initial recall test (before the second LDT) and who recognized at least four of the five words when tested at the end of the experiment. Additionally, in the experimental conditions, only participants who remembered the content and intent of the PM task and who remembered to carry out the intention at least once during the LDT were included in analyses leaving 64 participants in the PM groups (34 distinctive, 30 organizational) and 57 participants in the control groups (33 distinctive, 24 organizational). Reaction time data on both lexical decision tasks were trimmed so that RTs more than 2.5 standard deviations from each participant’s mean RT were excluded.

Lexical decision task data are summarized in Table 3. The data were first entered into separate ANCOVAs for words and for non-words to test the prediction that participants in the experimental conditions would take longer to correctly identify words and non-words than participants in the control conditions (i.e., to determine whether or not monitoring occurred in the experimental condition). Latencies from the initial LDT were used as the covariate. The prediction of strategic monitoring—participants in the organizational condition should monitor more than participants in the distinctive condition—should be observed in a significant
condition (PM/control) x study type (organizational/distinctive) interaction.

Next, correlations were calculated to determine the relationship between retrieval fluency, confidence, PM performance, and LDT latencies. If retrieval fluency does moderate monitoring, there should be a positive correlation between fluency and LDT latencies in both experimental groups. That is, participants who took longer to recall the target items before the LDT began should spend more time monitoring for those targets during the LDT. If retrieval fluency is being used as a metacognitive index, there should be a negative relationship between confidence and retrieval fluency. Additionally, if PM performance depends on the amount of monitoring one engages in, PM performance should correlate positively with LDT latencies.

LDT data for words and non-words were entered into separate 2 x 2 (condition = PM/control, learn = organizational/distinctive) ANCOVAs, using latencies from the initial LDT as the covariate. Unadjusted response latencies are summarized in Table 3, ANCOVA adjusted response latencies are presented in table 4. Participants in the experimental conditions made lexical responses to words more slowly than participants in the control conditions $F(1, 118) = 32.279$, $MSE = 184443$. The same pattern of results was true for non-words $F(1, 118) = 12.906$, $MSE = 148553$. These results replicates Smith’s (1999) finding that monitoring does occur during event based PM tasks. The type of learning task did not produce a main effect.
Table 3.

Response Latencies to LDT by Learn Type and Condition in Experiment 1.

<table>
<thead>
<tr>
<th>Learn Type</th>
<th>Condition</th>
<th>PM</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Words 669.62 (20.11) Nonwords 858.95 (53.14)</td>
<td>Words 632.55 (12.80) Nonwords 782.09 (27.09)</td>
</tr>
<tr>
<td>Distinctive Initial</td>
<td></td>
<td>644.94 (22.83) Nonwords 766.41 26.66</td>
<td>Words 643.70 (32.39) Nonwords 789.18 (66.94)</td>
</tr>
<tr>
<td>Target</td>
<td></td>
<td>741.46 (20.26) Nonwords 797.69 (24.17)</td>
<td>Words 636.94 (16.40) Nonwords 716.16 (29.86)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses represent standard errors.
Table 4.

**ANOVA Adjusted Response Latencies to LDT by Learn Type and Condition in Experiment 1.**

<table>
<thead>
<tr>
<th>Condition</th>
<th>PM</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>Words</td>
<td>Nonwords</td>
</tr>
<tr>
<td>Distinctive</td>
<td>725.60 (11.17)</td>
<td>840.08 (27.34)</td>
</tr>
<tr>
<td>Organizational</td>
<td>741.46 (12.68)</td>
<td>797.70 (13.71)</td>
</tr>
</tbody>
</table>
For words only, the learn by condition interaction was marginally significant $F(1, 118) = 3.50$, $\text{MSE} = 19997$, $p = .06$. Participants in the distinctive experimental condition were expected to show shorter latencies to make their lexical decisions, replicating the pilot study. The marginal study type by condition interaction demonstrates a trend that the extent of monitoring may in fact depend on the learning condition. The pattern of latencies in table 4 suggests that the interaction was driven by greater latencies when PM targets were studied by generating organizational cues compared to latencies in the control condition—that is, the pattern of latencies are consistent with a strategic monitoring hypothesis. When targets were studied by generating distinctive cues, mean latency to LDT responses in the experimental condition was only somewhat greater than in the control condition. Separate ANOVAs were run for each learning condition to confirm the nature of the interaction. Control participants in the distinctive condition correctly identified words more quickly than experimental participants, $F(1, 65) = 30.00$, $\text{MSE} = 2973)$. The same pattern held in the distinctive condition, $F(1, 54) = 23.482$, $\text{MSE} = 6479$. ANOVAs were run on experimental and on control participants, and no main effect of learning was detected for either group. Although the trend is for experimental participants in the distinctive condition to demonstrate less slowing relative to their respective control groups, only marginally significant differences have been detected here, making it impossible to say with certainty that the
organizational and distinctive learning groups engaged in different amounts of monitoring. Neither prospective memory performance nor confidence ratings differed significantly by t-test for the organizational and distinctive learning groups, summarized in table 5. These results did not bear out the prediction that monitoring for the PM targets during the LDT would be influenced by an explicit metacognitive assessment of confidence. Retrieval fluency, measured by the time it took to recall the PM targets before the LDT began, differed significantly between groups, as predicted. Participants in the organizational condition recalled PM targets more slowly than participants in the distinctive condition $t(62) = -2.37$.

Table 5.

**Prospective Memory Performance and Confidence Ratings in Experiment 1.**

<table>
<thead>
<tr>
<th>Learn type</th>
<th>Performance</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinct</td>
<td>0.70(0.05)</td>
<td>76.17(3.68)</td>
</tr>
<tr>
<td>Organiz.</td>
<td>0.64(0.05)</td>
<td>69.17(3.32)</td>
</tr>
</tbody>
</table>

**Correlational Analyses**

Correlations were calculated to further examine the idea that different processes were engaged depending on how one studied the PM targets. All correlations for the distinctive condition are presented in Table 6. Correlations for the organizational condition are presented in Table 7. If metacognition plays a role in
determining the quantity and/or quantity of monitoring during the LDT, the correlations should shed some light on that idea. For all correlations reported here, the same pattern was found when correlations were calculated using ANCOVA adjusted means or the raw means. Correlations of prospective memory performance and response time on the LDT for each learning group were calculated, as well as correlations between confidence and LDT latencies (Table 6). None of the correlations reached significance. Additionally, correlations between LDT latencies and post-experimental difficulty ratings were calculated, and again none reached significance.

The pre- and post-LDT assessments of difficulty were not significantly correlated with LDT latency for either of the experimental groups, suggesting that one’s confidence in ability to recognize the PM targets was not related to the quantity of observed monitoring. Next, the idea that confidence and PM performance were tested. It was predicted that higher confidence would be associated with higher PM performance. The correlation between pre-LDT confidence and PM performance was significant in the distinctive group (r = .40). The correlation between post-LDT confidence and PM performance was only marginally significant in the organizational and distinctive groups (r = -.33, p = .07, r = -.33, p = .06, respectively). Thus, initial confidence was not a reliable predictor of subsequent PM performance.
Table 6.

Correlations in the Distinctive Experimental Condition in Experiment 1.

<table>
<thead>
<tr>
<th></th>
<th>Pre-Con</th>
<th>Post-Con</th>
<th>Fluency</th>
<th>PM Perf.</th>
<th>Pre Recl</th>
<th>Post Recl</th>
<th>Recog</th>
<th>WordRT</th>
<th>NonWrdRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Con</td>
<td>1.00</td>
<td>-0.49</td>
<td>0.36</td>
<td>0.40*</td>
<td>0.38*</td>
<td>0.32+</td>
<td>0.12</td>
<td>-0.24</td>
<td>-0.20</td>
</tr>
<tr>
<td>Post-Con</td>
<td></td>
<td>1.00</td>
<td>-0.07</td>
<td>-0.33+</td>
<td>-0.20</td>
<td>-0.02</td>
<td>-0.08</td>
<td>0.05</td>
<td>0.07</td>
</tr>
<tr>
<td>Fluency</td>
<td></td>
<td></td>
<td>1.00</td>
<td>-0.11</td>
<td>-0.35*</td>
<td>-0.33+</td>
<td>0.03</td>
<td>0.41*</td>
<td>0.25</td>
</tr>
<tr>
<td>PM Perf</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.29</td>
<td>0.31+</td>
<td>-0.05</td>
<td>0.10</td>
<td>-0.27</td>
</tr>
<tr>
<td>Pre Recl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.22</td>
<td>-0.20</td>
<td>-0.34*</td>
<td>-0.28</td>
</tr>
<tr>
<td>Post Recl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>-0.48*</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>Recog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>0.27</td>
<td>0.14</td>
</tr>
<tr>
<td>WordRT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>-0.39*</td>
</tr>
<tr>
<td>NonWrdRT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 7.

Correlations in the Organizational Experimental Condition in Experiment 1.

<table>
<thead>
<tr>
<th>Pre-Con</th>
<th>Post-Con</th>
<th>Fluency</th>
<th>PM Perf</th>
<th>Pre Recl</th>
<th>Post Recl</th>
<th>Recog</th>
<th>WordRT</th>
<th>NonWrdRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Con</td>
<td>1.00</td>
<td>-0.46</td>
<td>0.16</td>
<td>0.15</td>
<td>0.04</td>
<td>0.29</td>
<td>0.08</td>
<td>0.24</td>
</tr>
<tr>
<td>Post-Con</td>
<td>1.00</td>
<td>-0.21</td>
<td>-0.33+</td>
<td>0.018</td>
<td>0.06</td>
<td>-0.11</td>
<td>-0.22</td>
<td>-0.361*</td>
</tr>
<tr>
<td>Fluency</td>
<td>1.00</td>
<td>0.11</td>
<td>-0.10</td>
<td>0.20</td>
<td>0.05</td>
<td>-0.21</td>
<td>-0.17</td>
<td></td>
</tr>
<tr>
<td>PM Perf</td>
<td>1.00</td>
<td>0.30</td>
<td>-0.05</td>
<td>0.12</td>
<td>0.03</td>
<td>-0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre Recl</td>
<td>1.00</td>
<td>0.08</td>
<td>0.16</td>
<td>0.21</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post Recl</td>
<td>1.00</td>
<td>0.380*</td>
<td>-0.19</td>
<td>-0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recog</td>
<td>1.00</td>
<td>0.19</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WordRT</td>
<td>1.00</td>
<td>-0.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NonWrdRT</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In the distinctive learning condition, PM performance was significantly correlated with retrieval fluency (i.e. the time it took to recall the PM target items following the study portion of the experiment; \( r = .40 \) and \( .36 \), respectively), but not in the organizational condition (\( r = .15 \) and \( .11 \), respectively). If retrieval fluency is being used as a primary metamnemonic index in generating confidence ratings, then a relationship of fluency and confidence is expected. In the distinctive learning condition, although confidence was correlated with the number of items recalled following the study portion of the experiment (\( r = .38 \)), confidence did not significantly correlate with retrieval fluency (\( r = -.241 \)). This suggests that, at least in the distinctive learning condition, the number of items recalled may play a more significant role in making confidence judgments than does retrieval fluency.

If retrieval fluency has heuristic value in allocating capacity to monitoring, at least at an implicit level, a positive relationship between retrieval fluency and LDT latencies should have been observed. Retrieval fluency did correlate with latencies to words in the distinctive condition (\( r = .41 \)). This pattern was not replicated in the organizational learning condition, where confidence did not significantly correlate with number of items initially recalled or with retrieval fluency (\( r = .04 \) and \( .16 \), respectively), nor did retrieval fluency correlate with LDT word responses (\( r = -.211 \)).
Overview of Experiment 2

If retrieval fluency is being used strategically as a metacognitive indicator of future noticing and responding to the PM target later in the experiment, then it should be possible to create a situation in which retrieval fluency would actually be misleading. The paradigm used by Benjamin et al. (1998), and discussed above, was used here. Benjamin et al. predicted that when the retrieval used to make a metacognitive judgment does not match the sort of retrieval that will actually occur at test, that retrieval will not be diagnostic of future performance. When participants retrieve an answer to a general knowledge question, that retrieval is from semantic memory. Prior research has shown that retrieval from episodic memory accurately predicts the likelihood of subsequent retrieval of the same information from episodic memory (Nelson & Dunlosky, 1991). Benjamin et al. predicted and confirmed that retrieval time from semantic memory does not accurately predict retrieval of the same information from episodic memory. The same reasoning was applied here to a PM task. When fluency of retrieval from semantic memory is used to predict future recognizability of PM targets, that fluency should be misleading as an index of subsequent episodic retrieval, and also of eventual PM performance.

In experiment 2, participants answered general knowledge questions with the goal of remembering the last five answers for a later prospective memory task. More difficult questions should take
longer to answer and lead to greater slowing on the ongoing LDT task during the PM phase of the experiment.
Method for Experiment 2

Participants

102 participants were randomly assigned to either an experimental (i.e. embedded) or control (i.e. delayed) condition. Participants received extra credit in undergraduate psychology courses for their participation.

Materials and Equipment

Twenty-five general knowledge questions (15 moderately easy filler questions, 5 moderately difficult, and 5 easy; Appendix B) were drawn from norms collected by Nelson and Narens (1980). Questions defined as easy were those answered with 89.6% accuracy or greater. The probability of correctly answering a moderately easy question ranged from .752-.870, and for moderately difficult questions, the probability of coming up with the correct answer ranged from .593-.733. Items used as PM cues were controlled for syllable length.

For each set of 20 questions, the first 15 questions were the same moderately easy questions for all participants. The last five questions were the target questions, and were either five easy questions or five moderately difficult questions. The answers to those questions served as the PM cues. The target questions were always the last five questions presented, and were presented with a reminder that these items were the ones that should be remembered for use during the next phase of the experiment. General knowledge
questions were presented and response latencies recorded using
SuperLab Pro Experimental Lab Software (Cedrus Corporation, 1999).

The PM portion of the experiment was the same LDT as in
experiment 1.

Procedures

Individuals participated one or two at a time and were randomly
assigned to one of four groups (easy-PM group, moderately difficult-
PM group, and so forth). The beginning of the experiment was
identical to that of experiment one: participants began the study
by completing a 100 item LDT included to measure baseline LDT
performance. Instructions for the next part of the study were like
those used by Benjamin et al. (1998, experiment 1). Participants
were informed that they would be asked 20 trivia questions and that
the time it took to answer each question was of primary interest to
the experimenter. Participants were told to press the space bar as
soon as they knew the answer to the question presented on the
screen, and not before. Upon pressing the space bar, a screen
appeared instructing participants to write down the answer to the
question on the answer sheet provided. Once the participant wrote
down the answer, they pressed the space bar again and the correct
answer to the question was presented on the screen. Participants
pressed the space bar again to continue on to the next question.
Instructions for the PM task and the LDT were provided before the
study portion of the experiment began. Participants in the
experimental conditions were instructed to remember the answers to
the last five questions for a later task — the LDT — and that whenever they encountered any of those answers during the LDT they should press the forward slash key (‘/’) after making their lexical decision response, and before pressing the space bar to continue to the next trial. The last five questions were indicated to the participant by changing the text color of the question from white to green. Individuals in the control condition were instructed that they should try to remember the answers to the last five questions for a later memory test.

Participants were discouraged from asking questions once the experiment began in order to decrease the likelihood of questions that serve as external reminders of the delayed intention. Because instructions for the LDT were provided at the beginning of the experiment, minimal instructions were provided immediately before the LDT began. Once the experiment was complete, participants answered the same post-test questionnaire as in experiment 1. They were then debriefed, given credit, and dismissed.
Results and Discussion of Experiment 2

Data were analyzed in much the same way as they were in experiment one. First, it was important to determine whether or not the manipulation of general knowledge question difficulty was effective. Accuracy in answering the questions was compared for participants in the easy and the moderately difficult groups. It was also expected that the manipulation would moderate perceived difficulty. Comparisons of pre-LDT confidence ratings were compared for experimental participants receiving each type of question. As in experiment one, words and non-words were first entered into separate 2 x 2 ANCOVAs (condition x question difficulty). A main effect of condition (experimental/control) was predicted, driven by slower responses on the LDT. The interaction between condition and question difficulty was also expected to reach significance—indicating that more monitoring occurred in the moderately difficulty condition relative to the easy condition.

Following the ANCOVAs, correlations were calculated to determine the relationship between retrieval fluency (the time it took participants to bring to mind answers to the target questions), confidence, PM performance, and LDT latencies. As in experiment one, if retrieval fluency does moderate monitoring, there should be a positive correlation between fluency and LDT latencies in both experimental groups. That is, participants who took longer to recall the target items before the LDT began should spend more time monitoring for those targets during the LDT. If retrieval fluency
is being used as a metacognitive index, there should be a negative relationship between confidence and retrieval fluency. Additionally, if PM performance depends on the amount of monitoring one engages in, PM performance should correlate positively with LDT latencies.

All differences reported were significant at alpha = .05. Data were analyzed only for those participants who recognized at least four of the five words when tested at the end of the experiment. As in experiment one, only participants who remembered the content and intent of the PM task and who remembered to carry out the intention at least once during the LDT were included in analyses leaving 42 participants in the PM condition (23 received the moderately difficult target questions, 21 received the easy target questions). In the control group, 46 participants remained (21 moderately difficult, 25 easy). Reaction time data on both lexical decision tasks were trimmed so that RTs more than 2.5 standard deviations from each participant’s mean RT were excluded.

The manipulation of general knowledge question difficulty was effective: moderately difficult target questions were correctly answered less frequently than easy target questions, $t(88) = -6.246$, $m = 3.1$ and $4.6$, respectively. Given the tendency for information that takes longer to initially retrieve to be better remembered on a later test (Gardiner, Craik, and Bleasdale, 1973), it was predicted that individuals in the moderately-difficult condition would have better recall and recognition performance on the post-test.
questionnaire. The results reported here equated recall and recognition performance by including only those participants who recalled three or more target items and recognized four or more target items, making comparison of groups on recall and recognition performance uninformative. When all of the data are included, neither mean recall nor recognition performance differed between the easy and moderately difficult groups $t(99) = -.014$, $t(99) = 1.44$, respectively. However, real and perceived difficulty of target questions was effectively manipulated, as demonstrated by lower rate of correct initial responding to general knowledge questions and by the marginally significant difference in confidence. Experimental participants answering easy target questions gave marginally higher confidence ratings, $t(42) = -1.95$, $p = .06$. PM performance, summarized in table 8, did not differ for the experimental groups, $t(42) = -.264$.

LDT data for words and non-words were entered into separate 2 x 2 (condition = PM/control, question = easy/moderately difficult) ANCOVAs, using latencies from the initial LDT as the covariate. Raw means for LDT performance are presented in table 9, and ANCOVA adjusted response latencies are presented in table 10. The experimental group correctly identified words more slowly than the control group $F(1, 83) = 38.69$, $MSE = 369752$. The same pattern was also true of non-words $F(1, 81) = 10.56$, $MSE = 118848$. The type of question did not produce a main effect on the speed of lexical responding. The condition by question interaction did not approach
Table 8.

Prospective Memory Performance and Confidence Ratings in Experiment 2.

<table>
<thead>
<tr>
<th>Learn type</th>
<th>Condition</th>
<th>Performance</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mod. Diff.</td>
<td>PM</td>
<td>0.62(0.05)</td>
<td>65.38(4.40)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td></td>
<td>73.50(5.00)</td>
</tr>
<tr>
<td>Easy</td>
<td>PM</td>
<td>0.63(0.07)</td>
<td>79.13(3.49)</td>
</tr>
<tr>
<td></td>
<td>Control</td>
<td></td>
<td>72.00(3.43)</td>
</tr>
</tbody>
</table>

significance, suggesting that the manipulation of general knowledge question difficulty did not moderate the amount of capacity dedicated to monitoring for the PM cues during the LDT ($F_s < 1.2$).

Correlational Analyses

All correlations in the easy experimental group are reported in Table 11. Correlations for the moderately difficult experimental group are reported in Table 12. In order to determine the relationship between PM performance and monitoring, correlations of prospective memory performance and response time on the LDT for each learning group were calculated. None reached significance. Correlations between confidence and LDT latencies, which were predicted to be negatively related. None of the correlations reached significance.
Table 9.

Response Latencies to LDT by Learn Type and Condition in Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>Condition</th>
<th>PM</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Words</td>
<td>Nonwords</td>
</tr>
<tr>
<td>Mod. Diff.</td>
<td>Initial</td>
<td>669.79 (20.61)</td>
<td>923.59 (58.98)</td>
</tr>
<tr>
<td></td>
<td>Target</td>
<td>827.64 (30.09)</td>
<td>902.80 (38.85)</td>
</tr>
<tr>
<td>Easy</td>
<td>Initial</td>
<td>681.10 (30.01)</td>
<td>852.78 (64.73)</td>
</tr>
<tr>
<td></td>
<td>Target</td>
<td>777.00 (29.52)</td>
<td>826.79 (42.34)</td>
</tr>
</tbody>
</table>

Note: Numbers in parentheses represent standard errors.
Table 10.

**ANCOVA Adjusted Response Latencies to LDT by Learn Type and Condition in Experiment 2.**

<table>
<thead>
<tr>
<th>Question</th>
<th>Words</th>
<th>Nonwords</th>
<th>Words</th>
<th>Nonwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM</td>
<td>PM</td>
<td>PM</td>
<td>PM</td>
<td>PM</td>
</tr>
<tr>
<td>Moderately Diff.</td>
<td>827.64 (15.22)</td>
<td>902.80 (30.98)</td>
<td>672.13 (21.15)</td>
<td>795.41 (38.67)</td>
</tr>
<tr>
<td>Easy</td>
<td>765.70 (19.80)</td>
<td>801.92 (34.57)</td>
<td>630.73 (10.62)</td>
<td>713.86 (15.37)</td>
</tr>
</tbody>
</table>
Table 11.

Correlations in the Easy Experimental Condition in Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>Pre-Con</th>
<th>Post-Con</th>
<th>Fluency</th>
<th>Num. Corr</th>
<th>PM Perf</th>
<th>Post Recl</th>
<th>Recog</th>
<th>WordRT</th>
<th>NonWrdRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Con</td>
<td>1.00</td>
<td>-0.62</td>
<td>0.18</td>
<td>-0.21</td>
<td>-0.06</td>
<td>0.25</td>
<td>-0.13</td>
<td>-0.15</td>
<td></td>
</tr>
<tr>
<td>Post-Con</td>
<td>1.00</td>
<td>-0.09</td>
<td>0.19</td>
<td>-0.07</td>
<td>-0.08</td>
<td>-0.13</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>1.00</td>
<td>-0.09</td>
<td>-0.14</td>
<td>0.05</td>
<td>0.51</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num. Corr</td>
<td>1.00</td>
<td>-0.08</td>
<td>-0.19</td>
<td>0.13</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM Perf</td>
<td>1.00</td>
<td>-0.08</td>
<td></td>
<td>0.55</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post Recl</td>
<td>1.00</td>
<td></td>
<td></td>
<td>0.02</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recog</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WordRT</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td>.73*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NonWrdRT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 12.

Correlations in the Moderately Difficult Experimental Condition in Experiment 2.

<table>
<thead>
<tr>
<th></th>
<th>Pre-Con</th>
<th>Post-Con</th>
<th>Fluency</th>
<th>Num.Corr</th>
<th>PM Perf</th>
<th>Post Recl</th>
<th>Recog</th>
<th>WordRT</th>
<th>NonWrdRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Con</td>
<td>1.00</td>
<td>-0.44</td>
<td>0.00</td>
<td>0.21</td>
<td>0.21</td>
<td>-0.12</td>
<td>0.21</td>
<td>0.29</td>
<td>0.09</td>
</tr>
<tr>
<td>Post-Con</td>
<td>1.00</td>
<td>-0.15</td>
<td>0.07</td>
<td>-0.42+</td>
<td>-0.42+</td>
<td>-0.19</td>
<td>-0.25</td>
<td>-0.12</td>
<td></td>
</tr>
<tr>
<td>Fluency</td>
<td>1.00</td>
<td>-0.29</td>
<td>0.03</td>
<td>-0.08</td>
<td>0.11</td>
<td>0.14-0.61*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num.Corr</td>
<td>1.00</td>
<td>-0.10</td>
<td>-0.05</td>
<td>0.04</td>
<td>0.08</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM Perf</td>
<td>1.00</td>
<td>0.25</td>
<td>-0.07</td>
<td>0.34</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post Recl</td>
<td>1.00</td>
<td>-0.08</td>
<td>-0.14</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recog</td>
<td>1.00</td>
<td>-0.19</td>
<td>-0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WordRT</td>
<td>1.00</td>
<td>0.64*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NonWrdRT</td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Additionally, correlations between LDT latencies and post-experimental difficulty ratings were calculated, and again none reached significance.

In the easy condition, pre- and post- LDT confidence judgments (i.e. question #5, Appendix A) correlated with one another (r = -.617), but not in the moderately difficult condition (r = -.276). One’s confidence in ability to recognize the PM targets was predicted to be negatively related to observed monitoring on the LDT but in fact neither of the confidence ratings in either condition were correlated with the speed of LDT responding, with one exception: in the easy condition, PM performance and RT to correctly respond to words on the LDT (r = .55). PM performance was not related in either condition to speed of LDT responding. For participants in the experimental condition receiving easy general knowledge questions, the time it took to answer the target questions was positively correlated with the ANCOVA adjusted LDT latencies for words and non-words (r = .51 and .51), and was also correlated with the number of target questions answered correctly (r = .72). None of these correlations approached significance in the moderately difficult condition.
General Discussion

The experiments reported here are interesting in that they shed some light on the role of monitoring in PM. The results of the pilot study provided some evidence that not only does monitoring occur in an event based PM task, but that the monitoring engaged may strategically depend on the memorability of the target items, and perhaps even the type of non-target item encountered. With this in mind, the first experiment was designed to test the role of metacognition in allocating capacity to monitoring, specifically the influence of retrieval fluency of PM target items. The results were consistent with the DARC model (Smith, 1999) in that it appears participants devoted capacity to the PM task by monitoring for the PM target before the target occurred.

The data presented in experiment one, although not entirely consistent with the pilot study results, leave the door open to the possibility of strategic monitoring in event-based PM tasks. These results are especially useful in that they provide even more support for the idea that event-based PM tasks are not automatic. The presence of a marginal study type by condition interaction suggests that further investigation of the possibility of strategic monitoring is warranted. It may be that the addition of the confidence rating after the initial recall of the PM targets and before the beginning of the LDT caused participants to refocus attention on making the PM response. That is, the confidence rating could increase the amount of capacity dedicated to monitoring
relative to what would have been allocated without the confidence rating. It is also possible that the interaction present in the pilot study could have been driven by group differences that went unmeasured (e.g. practice LDT).

Although the results observed in experiment two differed from those predicted, they do provide yet another piece of evidence that monitoring does occur during event-based PM tasks. The expected question by condition interaction did not approach significance. The manipulation of perceived difficulty was somewhat effective (i.e. the difference between confidence ratings for participants in the easy and the moderately difficult question groups was marginally significant), and participants were in fact less likely to answer correctly the moderately difficult general knowledge questions. As in experiment one, the results presented here are consistent with the DARC model (Smith, 1999). Slowing occurred in both experimental groups, at an apparently equivalent rate. However, as in experiment one, the use of the confidence ratings may have influenced monitoring strategy.

The predictive value of retrieval fluency in retrospective memory has been tested by Benjamin et al. (1998), wherein the speed with which an item was retrieved had heuristic value in predicting the likelihood of future recall of studied items. They showed that relying on retrieval fluency is misleading in a situation where the type of retrieval used in making a metacognitive judgment (i.e. retrieval from semantic memory) does not match the type of retrieval
that is being predicted (i.e. retrieval from episodic memory). The studies presented here attempted to extend those findings to PM by showing an example of an instance in which fluency can be a reliable predictor of future memory performance (experiment 1) and an instance in which fluency is a misleading index of performance (experiment 2). Although objective retrieval fluency was shown to predict slowing during the LDT for the distinctive group in experiment 1, and for participants in the easy question group in experiment 2, the correlation of retrieval fluency and LDT speed did not reach significance across all experimental groups. Objective retrieval fluency was not proven here to be directly predictive of monitoring, but failure to find an effect of retrieval fluency on confidence and/or monitoring does not rule out a metacognitive component of resource allocation in PM monitoring. The results also demonstrate the need to consider how people regulate and control their environment with regard to PM. Retrieval fluency and item memorability as examined in this study are only two potential characteristics that people might examine to predict future performance. Others include cue familiarity, frequency, specificity, and salience, just to name a few. These have been studied in the past, but only with regard to their direct effect on PM performance. Only a small amount of work has been done to establish how people actively evaluate such characteristics to modify and regulate their environment. Furthermore, the diagnosticity of these characteristics may be poorly learned and
their predictive value may be context-specific (cf. Benjamin, et al., 1998).

These results are valuable in that they cannot be explained by the \textit{Noticing + Search} model, nor can they be explained by Goschke and Kuhl’s model. Event-based PM performance cannot be explained by models that suggest that event-based PM cues are noticed automatically, without monitoring. The results observed in both experiments reported here are not consistent with the \textit{Noticing + Search} model proposed by Einstein and McDaniel (1996). In their model, cognitive capacity is not required during the period between establishing the intention and encountering the target event in the environment. Goschke and Kuhl’s (1996) model would suggest that because the targets and the PM responses are simple and well specified, allocation of cognitive resources to monitoring should not be necessary. Realization of the intention should rely solely on the increased level of activation associated with the encoded intention and action schema. It appears that event-based PM is more complex than has been suggested by Goschke and Kuhl—even simple and well defined cues may require monitoring.

None of results presented here are incompatible with Ellis's framework. One possibility is that the PM tasks were more akin to vigilance tasks than to a delayed intention. In Smith’s (1999) experiments, instructions were presented in such a way as to make the LDT akin to a vigilance task. Participants were instructed that memorization of the target words and making the appropriate PM
responses was extremely important. In the pilot study, instructions were written and reinforced in such a way as to make the LDT the primary task. This difference in instructions could cause the LDT in the pilot study to be interpreted less as a vigilance task, leading the participant to believe that monitoring, or at least consistent, vigilant monitoring, may not be necessary. The addition of the confidence rating in each experiment could have served as a reminder to the participant that they should make a special response to the PM targets, turning the task into a vigilance task—the confidence rating may have served to make the PM task appear more important.

The multiprocess view offered by McDaniel and Einstein (2001) could provide a post hoc explanation of the pilot data and the experimental data. Their framework predicts that controlled processes will be involved in noticing the PM target depending on the target distinctiveness, association of the target with the action, and the importance of the task. In the pilot study, the distinctiveness of the targets was manipulated in what might be considered a low-importance task. When the targets were high in distinctiveness (i.e. better learned), less capacity was deemed necessary for noticing those targets. When the targets were low in distinctiveness, participants may have felt that monitoring was necessary. In the experiments above, the addition of the confidence rating served to increase the perceived importance of the task, leading to a perceived necessity for monitoring. McDaniel and
Einstein do not suggest that monitoring is always necessary in an event-based PM task. Instead, a number of factors influence monitoring.

This paper has explored some of the metacognitive factors that influence monitoring in an event-based PM task. Although the potential influences of interest here, confidence and retrieval fluency, did not emerge as reliable predictors of monitoring, retrieval fluency did correlate significantly with monitoring when the targets were more memorable in experiment one, and when questions were easy in experiment two. That is, the longer it took for the targets to be retrieved, the more time one spent monitoring during the LDT. This suggests that retrieval fluency may be a relevant predictor of monitoring.

Time-based and event-based PM tasks have been dichotomized on the basis of the necessity of self-initiated retrieval in time-based tasks on the one hand (Einstein et al., 1998), as compared with the more automatic retrieval of event-based targets on the other hand (Einstein & McDaniel, 1996). McDaniel and Einstein (in press) suggest that controlled and automatic processes can both play a role in PM, whether independently or simultaneously. As such, viewing time-based and event-based PM tasks as opposite ends of a "monitoring" continuum may be a more realistic perspective. These results contribute to the growing body of data that are contrary to the notion that monitoring is necessary only in time-based PM tasks and will hopefully stimulate further investigation into what
factors, including metacognitive factors, influence the contributions of controlled and automatic processing.
References


58


End Notes

1. In experiment one, 22 participants were removed from analyses for failure to recall three or more of the PM targets before the LDT began. In the experimental condition, 7 distinctive and 9 organizational participants were excluded. In the control condition, 2 distinctive and 5 organizational participants were excluded.

2. Including all participants in experiment one caused the marginal interaction to fail to approach significance, and caused the pattern of correlations to change slightly--the marginal correlation between post-LDT confidence and PM performance did not approach significance when participants were excluded.

3. In experiment two, 7 experimental participants were removed from analyses for failure to make at least one PM response (4 moderately difficult, 3 easy). One participant, in the moderately difficult control condition was excluded for failure to recognize 4 or more of the PM targets on the post-test questionnaire. Two more experimental participants were excluded due to a program malfunction (one moderately difficult, one easy). One control participant (moderately difficult) and one experimental participant (easy) were excluded for failure to follow experimenter instructions.

4. Including all participants in experiment two did not change the outcome of the analyses.
Appendix A

Post Experimental Questionnaire

(Side 1)

1. Did you remember to look for the sports (weapons)?

2. What were you supposed to do when you saw one of the 5 sports (weapons)?

3. Please write down the five sports (weapons) you were instructed to look for.

<Turn page>

(Side 2)

4. Circle the five sports (weapons) you were instructed to look for:

   Bomb       Knife
   Stick      Chain
   Rifle      Missile
   Club       Gun
   Whip       Pistol

5. Before you started the word/non-word task, how difficult did you think it would be to notice the five words? (circle the appropriate response).

   0 1--------2--------3--------4--------5--------6--------7
   Easy-------->-------->-------->-------->-------->----Very Difficult

   If you do not remember or did not think about the difficulty of the task, circle zero.

6. In responding to the 5 sports, did you use any of the following strategies? (circle one)

   a) Just knew I would recognize the 5 words and make the response.
   b) Reminded myself throughout the word/non-word task
c) AFTER word/non-word response, checked whether or not letter string was a sport (weapon).

d) BEFORE word/non-word response, checked whether or not letter string was a sport.

e) none
f) other (please write down the strategy you used):

7. Did you think about responding to sports more after you saw the first sport?

Please write down any other comments you have about how you remembered to respond to sports during the word/non-word task.
Appendix B

General Knowledge Questions for Experiment 2

Answers follow each question in all capital letters. Numbers in parentheses are probability of recall.

Easy

1. What is the name of the horse-like animal with black and white stripes? ZEBRA (.970)
2. What is the name of the molten rock that runs down the side of a volcano during an eruption? LAVA (.915)
3. What sport uses the terms “Gutter” and “Alley”? BOWLING (.896)
4. What is the name of a dried grape? RAISIN (.896)
5. What is the sport associated with Wimbledon? TENNIS (.896)

Moderately Difficult

1. In which type of ski race does the downhill skier make sharp turns around poles? SLALOM (.726)
2. What is the name of the navigation instrument used at sea to plot position relative to the magnetic north pole? COMPASS (.685)
3. What is the name of the lightest wood known? BALSA (.619)
4. What is the name of the liquid portion of whole blood? PLASMA (.607)
5. What is the name of the crime in which a person purposely betrays his country? TREASON (.593)
1. Which precious gem is red? RUBY (.870)
2. What is the name of an airplane without an engine? GLIDER (.856)
3. What is the name of the rubber object that is hit back and forth by hockey players? PUCK (.852)
4. What is the name of the remains of plants and animals that are found in stone? FOSSILS (.852)
5. What is the name for a medical doctor who specializes in cutting the body? SURGEON (.844)
6. What is the name of an inability to sleep? INSOMNIA (.837)
7. What is the name of the spear-like object that is thrown during a track meet? JAVELIN (.833)
8. What is the name of the ship that carried the pilgrims to America in 1620? MAYFLOWER (.822)
9. What is the term for hitting a volleyball down hard into the opponent’s court? SPIKE (.819)
10. What is the name of the severe headache that returns periodically and often is accompanied by nausea? MIGRAINE (.807)
11. What is the last name of the author who wrote “Romeo and Juliet”? SHAKESPEARE (.796)
12. What is the name of the bird that cannot fly and is the largest on earth? OSTRICH (.770)
13. What is the name of the thick layer of fat on a whale? BLUBBER (.767)
14. What was the name of the supposedly unsinkable ship that sunk on its maiden voyage in 1912? TITANIC (.763)

15. Which type of snake do Asian snake-charmers use? COBRA (.752)
Vita

Dale W. Cockman was born in Greensboro, North Carolina, in 1976. He received his Bachelor of Arts in psychology from the University of North Carolina at Greensboro in 1998. The author then enrolled for one year as a non-matriculating graduate student at the University of North Carolina at Greensboro during the 1998-1999 academic year. Dale enrolled in the cognitive psychology at Louisiana State University in 1999 and defended this thesis in June 2001. He is now the Manager of Information Technology for The National Tree Trust in Washington, D.C.