1-1-2010

Typical PM targets are not typically better than atypical PM targets: Underlying mechanisms of Prospective Memory retrieval

Ji hae Lee

WASHINGTON UNIVERSITY

Department of Psychology

Typical PM targets are not typically better than atypical PM targets:

Underlying mechanisms of Prospective Memory retrieval

by

Ji hae Lee

A thesis presented to the

Graduate School of Arts and Sciences

of Washington University in

partial fulfillment of the

requirements for the

degree of Master of Arts

December 2010

Saint Louis, Missouri

Abstract

In two experiments, predictions from the discrepancy-plus-search view (e.g., McDaniel & Einstein, 2000) were tested against predictions from the familiarity view (McDaniel, 1995) and the preparatory attentional and memory processes theory (PAM; Smith, 2003). Discrepancy was manipulated by mismatching the actual and the expected category typicality of PM targets while familiarity was manipulated by the category typicality of PM targets alone. Consistent with PAM's prediction, higher PM performance with significant monitoring was found in the conditions where typical category exemplars served as nontargets. While the significant monitoring limited the opportunity for discrepancy to facilitate PM performance, further analyses hinted at a potential effect of discrepancy on PM performance. The implications of the findings are discussed under several theoretical frameworks.

ACKNOWLEDGMENTS

I am very thankful to my advisor, Mark McDaniel, for his tremendous guidance. I am also very grateful for my thesis committee members, Larry Jacoby and Jeff Zacks, for their helpful comments and my colleagues, Michael Scullin and Jill Shelton, for their stimulating discussions.

Ji hae Lee

TABLE OF CONTENTS

INTRODUCTION ……………………………………………………………...… 1

EXPERIMENT 1 ………………………………………………………………… 12

EXPERIMENT 2 ………………………………………………………………… 28

GENERAL DISCUSSION ………………………………………………………. 38

REFERENCES …………………………………………………………………... 45

LIST OF TABLES

TABLE 1: MEAN PM PERFORMANCE IN EXPERIMENT 1 ……………………49

TABLE 2: MEAN REACTION TIMES IN EXPERIMENT 1 ……………………...49

TABLE 3: MEAN PM PERFORMANCE IN EXPERIMENT 2 ……………………49

TABLE 4: MEAN REACTION TIMES IN EXPERIMENT 2 ……………………...49

Typical PM targets are not typically better than atypical PM targets:

Potential mechanisms of Prospective Memory retrieval

Prospective memory (PM) refers to memory for future actions. Real life examples of PM include remembering to pack one's memory stick for a conference, to deliver a message to a colleague, or to pick up cookies for kids on the way home from work. As described by the examples, human life is filled with various PM tasks.

In spite of the importance of successfully performing various PM tasks in everyday life, people often experience failures of PM tasks: we give a presentation without any slides because we forgot to pack the memory stick, and apologize to our colleague because we forgot to deliver the message and so forth. In general, PM tasks are challenging despite the fact that we perform them every day. Typically, a PM task is defined as performing an intended action when a pre-determined stimulus, a PM target, appears to signal the appropriate time to perform the intended action. For example, you may plan to pack your memory stick with data so that you can give it to your friend when you meet him at a conference. At a conference, while you are engaged in a stimulating conversation with other colleagues, your friend comes up to the group and quickly joins the conversation. To perform the PM task of giving the memory stick to your friend, in the absence of explicit retrieval request, you somehow have to recognize that your friend's face is the PM target, signaling the appropriate time to complete the PM task. Since you were not thinking about giving the memory stick to your friend when he joined the conversation, it is challenging to recognize that a stimulus (e.g., your friend's face) presented in the middle of ongoing activities is the PM target. In addition to the recognition of PM target, one has to retrieve the PM intention associated with the PM target to successfully perform the PM task.

Many theorists have proposed potential mechanisms that may support PM performance (e.g., McDaniel, 1995; McDaniel & Einstein, 2000; McDaniel & Einstein, 2007; McDaniel, Guynn, Einstein, & Breneiser, 2004; Smith, 2003; Smith, Hunt, McVay, & McConnell, 2007). Following, I consider three possible mechanisms that may explain how people recognize the PM target and retrieve the PM intention.

The preparatory attentional and memory processes (PAM) theory (Smith, 2003; Smith & Bayen, 2004; Smith et al., 2007) states that PM performance is facilitated by two processes: monitoring processes and memory processes. Smith and colleagues argue that the monitoring processes help one to search for the PM target while the memory processes support recognition of the PM target and retention of the PM intention. Furthermore, they claim that the monitoring processes are resource-consuming and obligatory for successful PM performance. According to the PAM theory, the resource-consuming, obligatory monitoring processes are implicated by the relative slowing down of the ongoing activity during a PM block of trials, in which participants have a PM intention in addition to the ongoing activity, compared to a control block of trials, in which participants have only the task demand for the ongoing activity. The PAM theory terms this relative slowing down as *monitoring cost* and predicts that it should precede any successful PM performance.

Contrary to the PAM theory's (Smith, 2003; Smith & Bayen, 2004; Smith et al., 2007) account of the obligatory monitoring for successful PM performance, the multi-process theory (McDaniel & Einstein, 2000, McDaniel et al., 2004) argues that spontaneous retrieval can support some PM performance in the absence of monitoring cost. According to the multi-process theory, constantly engaging in resource-consuming monitoring processes is not likely to be functional given the limited capacity of our mental resources. McDaniel and colleagues suggest that spontaneous retrieval may facilitate the recognition of PM target and the retrieval of PM intention under some circumstances (e.g. “focal” tasks; Scullin, McDaniel, Shelton, & Lee, 2010), whereas the resource-consuming monitoring may do so under other circumstances (e.g., “nonfocal” tasks; Scullin et al.).

As a potential underlying mechanism of spontaneous retrieval, McDaniel (1995, McDaniel & Einstein, 2000) has suggested the familiarity view. According to the familiarity view, when an item is perceived with high familiarity under the context of a PM task, one may interpret the high familiarity of that item as indicating significance. One explanation for the source of high familiarity of the PM target is a fluency-driven familiarity account (Jacoby & Dallas, 1981; Jacoby, 1983; Jacoby & Whitehouse, 1989). Jacoby and colleagues argue that people may utilize processing fluency of an item as a basis for the familiarity judgment of that item, such that people often judge fluently processed stimuli more familiar than less fluently processed stimuli. Given that PM targets are processed during the instruction for a PM task prior to the PM task itself, when the PM targets are later encountered during the PM task, it is likely that those PM targets could be processed more fluently, either perceptually or conceptually or both, than nontargets (which were not processed during the PM instruction). If so, people may judge the PM target more familiar than the nontargets, that are less fluently processed, during the ongoing activity. Consequently, one might interpret the PM target as bearing significance, which would then initiate a search for the source of the significance. This search then may lead to the recognition of the item as a PM target and the retrieval of PM intention.

Although the familiarity view (McDaniel, 1995) nicely describes how high familiarity of PM target can lead to the recognition of PM target and the retrieval of PM intention, the familiarity view fails to provide a complete account of PM performance. If only a PM target can have high familiarity, people could use the high familiarity as a basis for the significance judgment of the PM target. Going back to the previous example of packing the memory stick, imagine that the memory stick is on the desk with a couple of new toys your child left after his play. In this scenario, your memory stick is an object you encounter every day and you encoded as PM target. Thus, you will find the memory stick highly familiar. On the other hand, your child's new toys, serving as nontargets in this scenario, are not so familiar to you. Being the only object with high familiarity, the high familiarity of your memory stick may signal the significance of the memory stick. While it is possible that for some PM tasks the PM target is the only highly familiar item, often, the PM target is not the only highly familiar item as people continuously process items with varying degrees of familiarity. If the high familiarity of PM target cannot signal significance of the PM target, no search for the source of that significance will be initiated, making the retrieval of PM intention unlikely. Indeed, familiar PM targets do not always lead to higher PM performance than unfamiliar PM targets (e.g., McDaniel & Einstein, 1993).

To account for the lack of diagnosticity of familiarity as a potential source of significance for PM targets (McDaniel, 1995), the discrepancy-plus-search view has been proposed by McDaniel and colleagues (McDaniel & Einstein, 2000; McDaniel et al., 2004). According to Whittlesea and Williams (1998, 2001a, 2001b), people constantly evaluate the quality of their mental processing and make predictions about it based on their knowledge and experience. When the actual processing quality matches with what was expected, no discrepancy is signaled. But, when the actual processing quality mismatches with what was expected, discrepancy is signaled. When discrepancy is signaled, the cognitive system detects it. When discrepancy is detected by the system, an attribution is made to resolve discrepancy. For example, one may have a certain level of expected processing quality for the face of his colleague: higher than that for the face of a stranger but lower than that for the face of his child. If he finds the actual processing fluency of the colleague’s face match to the expected, no discrepancy will be signaled. However, if he finds the actual processing quality of the colleague’s face higher than that of his own child’s face (because he thought about the project he is working on with the colleague), the expected quality will then mismatch; at this point discrepancy will be signaled and an attribution will be made. Whittlesea and Williams suggest that this attribution can vary widely, such that discrepancy can be attributed as indicating familiarity or attractiveness of the discrepant item depending on the context. Many studies have found support for the discrepancy attribution framework by showing that discrepant stimuli are judged to be more familiar (e.g., Whittlelsea & Williams, 1998, 2001a), more preferable (Willems & van der Linden, 2007), more true (Hansen, Dechene, & Wanke, 2008) depending on different task contexts.

According to McDaniel and colleagues (McDaniel & Einstein, 2000; McDaniel et al., 2004), in the context of the PM task discrepancy from an item (e.g., PM target) may signal the significance of that item. Hence, the discrepancy-plus-search view suggests that the discrepancy of a PM target leads to the recognition of PM target and the retrieval of PM intention. For example, you are supposed to send an email to your colleague you met at a conference. Back in your office, you flip through the list of conference attendees, which is full of familiar names, while adding in references to your draft for a book chapter. According to the familiarity view (McDaniel, 1995), given that nontargets (the names of other conference attendees) are highly familiar, familiarity of the PM target (the name of your colleague who asked you to send an email) would not signal any significance of the PM target. However, according to the discrepancy-plus-search view, as long as the actual processing quality mismatches the expected processing quality and signals discrepancy, a PM target surrounded by highly familiar nontargets can be recognized as something significant. Going back to the example, forming the PM intention of sending an email to the colleague may lead you to more fluently process the colleague's name when you see the colleague's name on the list, more so than what you would normally expect from reading the name. This mismatch may then signal discrepancy, which could be attributed to the significance of the item. Once the item is perceived to be significant, the search for the source of that significance will be initiated, possibly leading to the recognition of PM target and the retrieval of PM intention.

Another unique prediction from the discrepancy-plus-search view (McDaniel & Einstein, 2000; McDaniel et al., 2004) is that a PM target with low familiarity can lead to higher PM performance if that PM target elicits discrepancy, even compared to a PM target with high familiarity, if the latter does not elicit any discrepancy. This prediction is in stark contrast with the prediction made by the familiarity view of which states higher PM performance for a PM target with high familiarity than a PM target with low familiarity (McDaniel, 1995). For example, now you are asked to send an email to a foreign scholar you met at a very selective conference. When you flip through the list of conference attendees, it is likely that the foreign scholar's name will read less fluently than other names, as most of them are your collaborators.

According to the familiarity view, given that the nontargets are highly familiar, the relatively unfamiliar PM target, the foreign scholar's name, would not be able to signal the significance of the PM target in this example. However, the discrepancy-plus-search view makes a different prediction. Given that the expected processing quality of the list of conference attendees is set at "easy" after reading many familiar names from the list, when the foreign scholar's unfamiliar name appears, the actual processing quality of "difficult" would mismatch with the expected. This mismatch, then, will signal discrepancy, leading one to make an attribution of significance of the name.

Some studies have found preliminary support for the discrepancy-plus-search view. Guynn and McDaniel (2007) found higher PM performance for participants who were pre-exposed to the PM targets during the study period prior to the PM instruction (pre-exposure group) than for participants who were not pre-exposed to the PM targets prior to the PM instruction (no pre-exposure group). The discrepancy-plus-search view's interpretation of the results is that the pre-exposure of the PM targets led to the much more fluent processing of PM targets compared to that of nontargets. Because participants presumably developed the expected processing quality for the PM targets based on the less fluent processing of nontargets, the processing fluency of PM targets mismatched with expected fluency. This mismatch could have signaled discrepancy, possibly facilitating the PM performance in the pre-exposure group. On the other hand, in the no pre-exposure condition, processing fluency of the PM targets could have been comparable to that of the nontargets, signaling no discrepancy. One shortcoming of this study was that the presumed discrepancy was potentially confounded with familiarity: the familiarity view could explain high PM performance in the pre-exposure condition compared to the no pre-exposure condition because pre-exposure would have increased familiarity.

To further disentangle the discrepancy-plus-search view from the familiarity view, several studies have been conducted (e.g., Brenieser & McDaniel, 2006; Lee & McDaniel, 2010). With an anagram solution task as their ongoing task, Lee and McDaniel manipulated discrepancy by matching or mismatching the actual anagram solution difficulty of PM targets to the expected solution difficulty of PM targets. In the nondiscrepant conditions, the solution difficulty of PM targets matched that of the list of anagrams the PM targets were embedded in (e.g., easy PM targets embedded in the easy list or vice versa). In the discrepant conditions, the solution difficulty of PM targets mismatched that of the list of anagrams in which the PM targets were embedded (e.g., difficult PM targets embedded in the easy list or vice versa). The reasoning was that by solving a list of anagrams with a certain level of solution difficulty (e.g., easy), one would build an expectation about the solution difficulty of the subsequent anagram (easy). If the solution difficulty of PM target (easy) matched the expectation, no discrepancy would be signaled. If the solution difficulty of PM target mismatched, such that a difficult anagram was presented as a PM target after one built an expectation of easy anagrams, discrepancy would be signaled. Supporting the notion that discrepancy facilitates PM performance, Lee and McDaniel found that the PM performance was higher if the solution difficulty of PM targets was different from that of the list than if the difficulty was the same with that of the list. The familiarity view (McDaniel, 1995) predicted higher PM performance for the easy PM targets than the difficult PM targets because more fluent processing of the easy PM targets will presumably be interpreted as indicating high familiarity of those PM targets. However, the familiarity view was not supported as there was no main effect of solution difficulty of PM target. Also, given that relatively high PM performance across all the conditions (*M*s>.80) was not preceded by any significant monitoring cost, no support was found for the PAM theory (which argues for resource-consuming, obligatory monitoring for successful PM performance, Smith, 2003; Smith & Bayen, 2004; Smith et al., 2007).

While the results of previously discussed studies have been suggestive of discrepancy-plus-search processes, there have been only a few studies conducted to investigate the function of discrepancy on PM performance and the findings have been somewhat inconclusive (e.g., Breneiser & McDaniel, 2006; Guynn & McDaniel, 2007, Einstein et al., 2005; Lee & McDaniel, 2010). To obtain more conclusive evidence for the discrepancy-plus-search view on PM, we wanted to extend the findings from Lee and McDaniel. With regard to Lee and McDaniel, there is a possible concern with their use of the anagram solution task as the ongoing activity. Given the relatively longer time frame for solving the anagrams (2-3 sec for the easy anagrams and 4-5 sec for the difficult anagrams), participants could have engaged in monitoring, rather than utilizing discrepancy, to perform the PM task. Even if more monitoring was found in the discrepant conditions, this argument of participants potentially engaging in monitoring is insufficient support for the PAM theory (Smith, 2003, Smith & Bayen, 2004), as the theory fails to explain why participants would have monitored more only in the discrepant conditions. Still, if participants engaged in monitoring in the discrepant conditions during Lee and McDaniel's study, it would be difficult to argue that their anagram paradigm captured the pure influence of discrepancy on PM performance. Thus, we wanted to employ a task that has a relatively short time frame so that the related RT measure is sensitive enough to detect the monitoring cost if participants decide to engage in strategic monitoring. Also, for further generalization of discrepancy, we wanted to manipulate discrepancy by using fluency-driven familiarity other than that of the anagram solution difficulty.

For several reasons we chose a category judgment task as the ongoing activity. People can make category judgments in a matter of hundred milliseconds (msec). Thus, with the much shorter time frame, we have a more sensitive measure for potential monitoring cost. Another factor is the graded structure of categories (e.g., Rosch & Mervis, 1975). One of the classic findings in categorization literature is that exemplars from the same category differ in their category typicality: one may think "water" is a better example of category "liquid" than "blood" with both "water" and "blood" being legitimate examples of the category "liquid". This category typicality difference among exemplars is implicated in differential response times in tasks that require accessing category information, such as a category judgment task. Studies have found that people are faster and more accurate in processing typical exemplars than atypical exemplars (e.g., Collin & Quillian, 1969; Rosch, Simpson, & Miller, 1976). Given that typical and atypical exemplars are processed with different fluency levels, the category judgment task with typical and atypical exemplars seemed suitable to elicit discrepancy for our study.

More specifically, we could elicit discrepancy by intermixing typical and atypical exemplars. For example, if a participant were to be presented with a list of very typical nontarget exemplars for a category judgment task, one after another, he would experience a certain level of fluency associated with the category judgment task and may expect the next exemplar will be processed at the fluency level he experienced (e.g., very typical, hence, very fluent). If a typical PM target exemplar was next presented, the actual processing quality and the expected would match, signaling no discrepancy (nondiscrepant condition). However, if the same typical PM target exemplar was presented within a list of atypical nontarget exemplars, the processing quality would mismatch with the expected, thereby signaling discrepancy (discrepant condition). This discrepancy then may lead to an attribution of significance of the PM target, in turn stimulating the recognition of the PM target and the retrieval of PM intention associated with the PM target. Another pair of nondiscrepant and discrepant conditions was constructed by using the atypical PM targets and atypical and typical nontargets, respectively. The discrepancy-plus-search view (Einstein & McDaniel, 2000; McDaniel et al., 2004) predicts higher PM performance in the discrepant conditions compared to the nondiscrepant conditions. More specifically, this view predicts the cross-over interaction of the typicality of PM target exemplars and that of nontarget exemplars, such that, PM targets that have the mismatching typicality relative to nontargets will lead to higher PM performance than PM targets that have the matching typicality to nontargets. Furthermore, given that discrepancy-plus-search processes are assumed to support spontaneous retrieval, no monitoring is necessary to facilitate PM performance in the discrepant conditions.

On the other hand, the familiarity view (McDaniel, 1995) predicts no differential performance between the discrepant and the nondiscrepant conditions. According to this view, the fluent processing of typical PM target exemplars is interpreted as high familiarity of those PM target exemplars; thus, high PM performance is expected for the typical PM target exemplars compared to the atypical PM target exemplars. Moreover, the familiarity view also does not predict any significant monitoring cost to precede PM performance as the familiarity view proposes that familiarity is a potential mechanism that supports spontaneous retrieval.

Contrary to the familiarity (McDaniel, 1995) and the discrepancy-plus-search views (Einstein & McDaniel, 2000; McDaniel et al., 2004), the PAM theory (Smith, 2003; Smith & Bayen, 2004; Smith et al., 2007) predicts that a significant monitoring cost will be observed for PM blocks of trials. Also, it further predicts no differential PM performance between typical versus atypical PM targets. It is possible for PAM theory to argue that typical PM target exemplars will lead to higher PM performance than atypical PM target exemplars as the recognition of the former may be easier. However, when only a small number of PM targets are used, it is unlikely that the difficulty in the recognition and monitoring processes will differ between the PM target exemplars with different typicality. Furthermore, the PAM theory predicts that PM performance for the same PM targets should not differ based on the typicality of nontarget exemplars.

Experiment 1

In this experiment, we manipulated the typicality of PM target exemplars and nontarget exemplars to elicit discrepancy. While we wanted to maximize the discrepancy by increasing the difference between the different category typicality levels, we tried to make sure that the difference was not too obvious to participants. Our concern was that if the difference is too obvious, participants may recognize discrepancy is experienced because the PM target exemplar has a category typicality different from nontargets. If so, discrepancy may not lead to the attribution of significance of PM target. Hence, rather than selecting the most typical and atypical exemplars from a category, for our atypical exemplars, we used the exemplars from the third, instead of the fourth, quartile of typical exemplars of a category.

In addition to manipulating discrepancy, we implemented two features to control for possible ceiling effects in PM performance. In PM laboratory paradigms, PM performance is often at the ceiling hindering the observation and the interpretation of influence of any IVs on PM performance. To avoid this potential ceiling, first, we chose three PM targets. We reasoned that, with three PM targets, we may potentially increase the difficulty of the PM task. Secondly, we constructed the PM instruction in a manner that did not specify the context in which PM targets would appear. Marsh, Hicks and Bink (2006) reported selective monitoring behaviors when the PM context was specified. When Marsh and colleagues told their participants that PM targets will appear only during the second lexical decision task, their participants showed monitoring cost only during that second lexical decision task. Though the PM instruction was provided prior to all lexical decision tasks, these participants did not show any monitoring cost during the first lexical decision task when compared to the participants in the control group who were not asked to perform any PM task. We reasoned that if the context in which PM targets appear is not specified, participant cannot selectively engage in monitoring. For example, upon receiving the PM instruction not specifying the PM context, people may initiate monitoring. And suppose that PM targets appear in the latter of the two distinctive tasks following the PM instruction, such as a lexical decision task and a category judgment task. By the time participants perform the category judgment task, they may be less likely to stay engaged in any monitoring that was initiated upon receiving the PM instruction and maintained during the lexical decision task. The more likely people disengage from monitoring, the more likely discrepancy may facilitate PM performance. Thus, as a distractor task, a lexical decision task always followed the PM instruction, preceding the category judgment task in which PM targets appeared. Also, to measure monitoring cost, participants will receive a PM block and a control block of lexical decision and category judgment tasks. By comparing the mean response time of category judgments in the PM block to that in the control block, we may examine monitoring cost associated with PM task.

Methods

Participants and design. Seventy eight participants were recruited from the Washington University in St. Louis community and participated in the experiment in exchange for a partial course credit or monetary compensation. The experiment was a 2 x 2 x 2 mixed factorial design, with the category typicality of the PM target exemplar (PM target typicality, typical vs. atypical) and the category typicality of the nontarget exemplar (nontarget typicality, typical vs. atypical) as between subjects factors and the block type (PM vs. control) as a within subjects factor. The presentation order of block was counterbalanced so that a half of the participants first received the PM block and the other half first received the control block.

Materials. Two separate sets of stimuli were used, one for the lexical decision task and another for the category decision task. For the lexical decision task (which was used as a distractor task) a set of six hundred items was used (for further information, see Exp. 4 in Scullin et al., 2010). The set of six hundred items was divided into two subsets for a lexical decision task in each block, the PM and the control block. The order in which each subset was assigned to a certain block was counterbalanced. A half of each subset was words and the other half was nonwords. All of the items were 4-8 letters long and pronounceable.

For the category decision task, 49 categories were selected from Van Overschelde, Rawson, and Dunlosky's (2004) updated and expanded Battig and Montague's norm (1969): 3 categories for the PM targets and 46 categories for the nontargets were used. From each of the 49 categories, all of the exemplars were ordered according to their category typicality and divided at the median typicality into the "typical subset" and the "atypical subset". Then, the four most typical exemplars from the typical subset were picked to serve as "typical exemplars". The four most typical exemplars from the atypical subset were picked to serve as "atypical exemplars". Pilot studies were conducted to validate this manipulation of category typicality of PM target and nontarget exemplars and found that people were faster at judging the "typical exemplar" belongs to a given category than the "atypical exemplar" and rating the "typical exemplars" more typical of a certain category than the "atypical exemplars". All of the exemplars were nouns. The mean length and the mean log-transformed Hyperspace Analogue to Language (HAL) frequency were 5.55 letters and 9.24 for the typical nontarget exemplars and 5.60 letters and 8.55 for the atypical nontarget exemplars, respectively.

From each of the three PM target categories (musical instrument, transportation vehicle, relative), we picked one PM target, for a total of three PM targets. The most typical exemplar among the "typical subset" from each of the PM target category was drawn to serve as "typical PM target exemplars" (drum, bus, uncle). The most atypical exemplars among the "atypical subset" from the same categories were drawn to serve as "atypical PM target exemplars" (harp, taxi, grandparent). The mean length and the mean HAL frequency were 4 letters and 9.60 for the typical PM targets and 6.33 letters and 6.86 for the atypical PM targets, respectively. Each of the three PM targets was randomly presented on the 31st, the 55th, and the 84th trial in the PM block.

The rest 46 (out of 49) categories were used to generate nontargets. All four exemplars from the "typical subset" served as "typical nontarget exemplars" and all four exemplars from the "atypical subset" served as "atypical nontarget exemplars". For each nontarget exemplar type (either typical or atypical), a half of the four exemplars remained in the initial pairing of the category they belonged to while the other half of the four exemplars were randomly paired with other categories (e.g., for typical exemplars from the category "fruit": apple-fruit, banana-fruit, cherry-furniture, and orange-metal). The order of this paring was counterbalanced. A total of 184 nontarget exemplars were then divided into two sets of 92 trials, one set for the PM block and one for the control block.

Procedure. Participants were tested in groups of 1-6. Participants sat in front of a computer monitor and were provided a keyboard for their response. All materials were presented in 18 point Times New Roman font type and in black on a white background. After signing the consent form approved by the Institutional Review Board at Washington University in St. Louis, participants were given the instructions for the lexical decision task and the category judgment task. A few practice trials followed each instruction. The lexical decision task was used (1) to make it vague in which context the PM target would appear, and (2) to have some time interval between the PM instruction and the actual PM task. For the lexical decision task, participants were instructed to press the *y* key if the string of letters on the screen was a real word and the *n* key if it was not a real word. A fixation point was presented during the interval between trials. For each trial, the string of letters stayed on the monitor until participants made their response.

For the category judgment task, participants were asked to press the *y* key if the word on the center of the screen belonged to the category presented above the word. Participants were told to press the *n* key if the word did not belong to the presented category. Each pairing of the word and the category was presented until participants made their response. In addition to always presenting the category above the given word, for each trial, the category was presented with a set of colons surrounding it from both sides (e.g., :Fruit:) so that it was clear for the participants which stimulus was the category and which one was the exemplar they needed to categorize.

After receiving the initial instructions and completing the practice trials for each task, participants were presented with the first set of lexical decision task and category judgment task. Before the first set started, the PM instruction was provided to only the participants who performed the PM block first. The participants were told to press the *q* key during the experiment if they saw either “harp”, “taxi”, or “grandparent”, if they were in the atypical PM targets conditions, or “drum”, “bus”, or “uncle”, if they were in the typical PM target conditions. The same PM instruction was provided upon the completion of the first set of lexical decision task and category judgment task for the participants who performed the control block first.

After receiving the PM instruction, participants who performed the PM block first summarized the instructions they received on a piece of paper. Participants who received the control block first were also asked to summarize the appropriate instructions at this point. Participants were asked to summarize instructions for other tasks (e. g., lexical decision task) at the appropriate times throughout the experiment and the experimenter always checked the accuracy of summarized instructions, including the PM targets. After summarizing the PM instruction, participants solved the lexical decision task for about five minutes. Upon the completion of the lexical decision task, participants performed the category judgment task. Participants who received the control block first responded to 92 trials of the category judgment task while participants who performed the PM block first responded 95 trials of the category judgment task, including 3 PM target trials. For each participant, upon the completion of each task, the computer told participants that the task was over and the next task was to be started.

When the first set of lexical decision task and category judgment task was over, participants who performed the PM block first were told that "there would be no secondary task that requires pressing the *q* key or looking for any specific words (the PM targets)". Participants who performed the control block first received the PM instruction upon the completion of the first set of tasks.

After completing the first set of tasks and the processing of appropriate instructions, participants solved a second set of lexical decision and category judgment tasks. After the second category judgment task, participants were given the post-test questionnaire to check their retrospective memory for the PM targets. Upon the completion of the questionnaire, participants were debriefed and excused.

Results

Participants' responses along with the Reaction Times (RTs) to the PM target exemplars and the nontarget exemplars during the category decision tasks were recorded. The recorded responses and the RTs to the PM target and the nontarget exemplars were averaged and analyzed. An alpha level of .05 was set for statistical significance for all statistical analyses, unless noted otherwise. Performance on the lexical decision task was recorded but was not analyzed.

PM performance. The accuracy of PM performance was obtained by computing the proportion of correct PM response for the three PM targets. The correct PM response was pressing the *q* key upon the presentation of the PM target exemplars. A 2 x 2 x 2 between subjects measures Analysis of Variance (ANOVA) was conducted on the mean PM performance with the PM target typicality (typical vs. atypical), the nontarget typicality (typical vs. atypical), and the block order (PM block first vs. control block first) as between subjects factors. A significant main effect of nontarget typicality was found, $F(1, 70) = 4.0$, $MSE = .12$, in that the participants were more likely to perform the PM task when the nontargets were typical exemplars (M=.70) than when the nontargets were atypical (M=.54). Neither the main effect of PM target typicality (F<1) nor the interaction of PM target typicality by nontarget typicality (F=1.17) was significant (see Table 1 for the means).

There was a significant two-way interaction of PM target typicality and block order, $F(1,70)= 8.70$, $MSE= .12$. Post-hoc comparisons revealed that the PM performance for the typical PM targets (M=.80) was higher than for the atypical PM targets (M=.48) in the control block first condition, $F(1,70)=7.71$, MSE=.12, while the PM performance for the atypical PM targets (M=.67) was only nominally greater than for the typical PM targets (M=.52) in the PM block first condition, $F(1,70)=1.89$, MSE=.12, p=.17. No other effects or interactions were significant.

Category judgment reaction times. We analyzed the RTs to the category judgments to see if participants were slowed down on the category judgment task when the PM task was present. PM target exemplars were excluded from this analysis.

The RTs were trimmed following the methods used by Einstein et al's (2005). The RTs from only the correctly judged exemplars in each block (PM vs. control) were averaged. Also, RTs that were two standard deviations smaller or greater than the individual means were removed (see Table 2 for the means). Then, the averaged RTs were entered into appropriate ANOVAs.

First, to validate our experimental manipulation of typicality, we took only the RTs from the control block from the group who received the control block first and entered the RTs into a 2 x 2 x 2 mixed repeated measures ANOVA with the response type (yes vs. no) as a within subjects factor and the PM target typicality and the nontarget typicality (typical vs. atypical) as between subjects factors. There was a main effect of response type, in that yes responses (1073 msec) were quicker than the no responses (1113 msec), $F(1, 33) = 7.14$, $MSE=4226.78$. More importantly, there was a significant main effect of nontarget typicality, $F(1,33)=6.86$, $MSE=4226.78$, in that participants took longer to make category judgments for the atypical nontargets (1174 msec) than typical nontargets (1012 msec), validating our manipulation and replicating the classic typicality effect (e.g., Rosch, Simpson, & Miller, 1976). Neither the main effect of PM target typicality ($F=1.17$) nor other interactions were significant (Fs < 1).

Next, we entered the RTs into a 2 x 2 x 2 x 2 x 2 mixed repeated measures ANOVA with the response type (yes vs. no) and the block type (PM vs. control) as within subjects factors and the PM target typicality (typical vs. atypical), the nontarget typicality (typical vs. atypical), and the block order (PM block first vs. control block first) as between subjects factors. Again, there was a significant main effect of response type, $F(1,70) = 28.05$, $MSE=5790.13$, in that participants were slower in making the no response (1161 msec) than the yes response (1121 msec). More critical to the present study, there was a significant main effect of block type, $F(1, 70) = 44.27$, MSE=5790.13, such that participants were slower at making the category judgments in the PM block (1184 msec) than in the control block (1097 msec). This reflects that having to perform the PM task interfered with the category decision task. There was a significant interaction of block type by nontarget typicality, $F(1, 70) = 10.16$, MSE= 5790.13. Following post-hoc comparisons found that the interaction was caused by the greater difference in RTs between the two nontarget conditions (*M*s= 1152 and 1041 msec for atypical nontargets and typical nontargets, respectively) on the control blocks relative to the PM blocks (*M*s= 1198 and 1171, respectively).

To identify any potential monitoring cost directly associated with PM performance in each condition, planned comparisons were made between the PM block and the control block from each condition. Among the four conditions, except the (discrepant) condition with the typical PM targets and the atypical nontargets ($F(1,70)=2.63$), all three conditions showed significant monitoring: $F(1,70)=4.67$, MSE=5790.13 for the nondiscrepant condition with atypical nontargets, $F(1,70)=14.30$, MSE=5790.13 for the discrepant condition with typical nontargets, and $F(1, 70)= 47.42$, MSE=5790.13 for the nondiscrepant condition with typical nontargets. Post-hoc comparisons revealed that the monitoring cost for the nondiscrepant condition with typical nontargets (170 msec) was significantly higher than that for the discrepant condition with atypical nontargets (40 msec), $F(1,70)=29.19$, MSE=5790.13 while the monitoring cost for the discrepant condition with typical nontargets (91 msec) was only nominally higher than that for the nondiscrepant condition with atypical nontargets (52 msec), $F(1,70)=2.63$, *MSE*=5790.13, *p*=.11. Together with the higher PM performance in the typical nontarget conditions compared to the atypical nontarget conditions, the monitoring seems to be more prominent in the conditions where higher PM performance was observed.

Furthermore, the omnibus ANOVA revealed a significant main effect of PM target type, *F*(1,70)=11.78, *MSE*=5790.13, suggesting that people were on average faster at making the category judgment on nontargets if the PM targets were atypical (*M*=1072 msec) than typical (*M*=1208 msec), regardless of the typicality of the nontarget exemplars themselves. This main effect of PM target typicality regardless of nontarget typicality suggested that having to perform the PM task might influence the RTs on the category judgment. The main effect of nontarget typicality was only marginal (*F*(1,70)=3.01, *MSE*= 5790.13, *p*=.087), showing that participants took longer to respond to atypical nontarget exemplars (*M*=1175 msec) than to typical nontarget exemplars (*M*=1106 msec). This marginally significant effect of nontarget typicality on RTs also seemed to suggest that performing a PM task might affect the RTs rather than to reflect unsuccessful manipulation of category typicality.

Category judgment accuracy. Accuracy of the category judgment task was calculated by computing the proportion of correct judgments for each response type (yes vs. no) in each block (PM vs. control). Again, PM target exemplars were excluded in this analysis. Accuracy of the category judgment task was entered into a 2 x 2 x 2 x 2 x 2 mixed repeated measures ANOVA with the response type (yes vs. no) and the block type (PM vs. control) as within subjects factors and the PM target typicality and the nontarget typicality (typical vs. atypical), and the block order (PM block first vs. control block first) as between subjects factors. There was a significant main effect of response type, $F(1, 70) = 8.42$, $MSE = .002$, showing that participants were more accurate in making the no responses (M=.93) to the category judgment task than the yes responses (M=.91). There was a significant interaction of response type by nontarget typicality, $F(1, 70) = 17.64$, MSE=.002. Post-hoc comparisons revealed that the category judgment accuracy was higher for the no response (M=.95) than the yes response (M=.90) for the atypical nontargets ($F(1,70)=23.41$, MSE=.002) while the accuracy did not differ as a function of response type for the typical nontargets (Ms=.92 for no responses and .91 for yes responses). There was no significant effect of block type, $F<1$, in that performing the PM task did not interfere with the accuracy of the category judgment task (Ms=.92). There was a significant interaction of block type by block order, $F(1,70) = 6.16$, MSE=.002, showing the accuracy for the PM blocks differed as a function of block order (Ms=.90 for the control block first condition and .93 for the PM block first condition) while the accuracy for the control blocks did not differ as a function of block order (Ms=.92).

Individual differences analysis. Given that significant monitoring may limit the effect of discrepancy on PM performance, we wanted to see whether the effect of discrepancy would emerge for individuals for whom monitoring was minimal. One approach to eliminate monitoring is selectively looking at PM performance from the individuals that exhibit no or relatively less monitoring (Einstein et al., 2005, Experiment 4). We used the modified version of the procedure Einstein and colleagues used. Einstein and colleagues used the difference in mean RTs between the PM block and the control block (after adjusting for the speedup of the 2nd block due to practice) to compute the monitoring index for an individual. Given the possibility of discrepancy emerging on selective PM trials (e.g., later trials in the experiment), we wanted to compute separate monitoring indexes diagnostic of PM performance for each PM trial for each individual. Recently, some studies suggested that the functional monitoring, the relative slowing down of trials proximal to the PM target, may be a more sensitive measure of monitoring than the overall monitoring, the relative slowing down in the PM block compared to the control block as a whole (Scullin, McDaniel, & Einstein, 2010; Scullin et al., 2010). Thus, we took the 5 nontarget trials preceding each PM target from the PM block and the matching trials from the control block. Then, we took the difference between the 5 trials from the PM block and the 5 trials from the control block, after adjusting for the average speed up from the block 1 to the block 2 (the block 1 and 2 could have been either the PM or the control block based on the counterbalancing order). We then used this difference value to categorize an individual either as engaged in monitoring (monitors) or not engaged in monitoring (non-monitors). While there is no clear consensus on the magnitude of slowing down that defines functional monitoring, we set the arbitrary criterion of less than 30 msec difference between the PM block and the control block as indicating no monitoring.

Out of 78 participants, 40 individuals were categorized as non-monitors as they exhibited less than 30 msec monitoring for the 1st PM trial. Thirty-two and 39 individuals were categorized as non-monitors for showing less than 30 msec monitoring for the second and the third PM trial, respectively. From those who were categorized as non-monitors, 37 people showed less than 30 msec monitoring for more than one trial and 24 people showed less than 30 msec monitoring for only one trial. A close inspection of PM performance from the individuals with less than 30 msec monitoring for each PM trial as a function of discrepancy (discrepant vs. nondiscrepant conditions, collapsed across typical and atypical PM targets) revealed nominally higher PM performance in the discrepant condition for all three PM trials (*M*s= .61 for the first, .73 for the second, and .67 for the third PM trials) than that in the nondiscrepant condition (*M*s=.47, .50, .56, respectively). These differences and the averaged PM performance across trials (*M*s= .65 for the discrepant condition and .49 for the nondiscrepant condition), however, were not statistically significant (Fs < 1.89) according to the planned comparisons conducted.

Discussion

To summarize the results from the Experiment 1, first, we found that people are more likely to perform the PM task if their nontarget exemplars were typical than atypical, regardless of the PM target typicality. These results are not readily explained by any of the current theories of PM retrieval, most of which focus on the property of the PM target rather than that of nontargets (McDaniel, 1995; McDaniel & Einstein, 2000; McDaniel et al., 2004; Smith, 2003). The familiarity view (McDaniel) predicts higher PM performance for PM targets with high typicality while the discrepancy-plus-search view (McDaniel & Einstein) predicts higher PM performance for PM targets with the mismatching typicality to the typicality of the nontargets. Failures to find any significant effect of PM target typicality and interaction of PM target typicality by nontarget typicality does not support either the familiarity view or the discrepancy-plus-search view (Einstein & McDaniel; McDaniel et al.).

Rather, two patterns of the results seem to be consistent with the PAM theory (Smith, 2003; Smith & Bayen, 2004; Smith et al., 2007). One piece of support for the PAM theory is that there was no effect of PM target typicality on PM performance. While the difficulty of recognition check for PM targets with different typicality may differ, the PAM theory would possibly not expect that PM target typicality would influence PM performance given the small number (three) of PM targets used in Experiment 1. No substantial difference in recognition check difficulty then could have led to no difference in PM performance as a function of PM target typicality.

Another pattern that is consistent with the PAM theory (Smith, 2003; Smith & Bayen, 2004; Smith et al., 2007) is the finding of PM performance being associated with monitoring. Significantly more monitoring associated with high PM performance in the typical nontarget conditions seems to be consistent with the PAM theory, which argues that monitoring is necessary for successful PM performance (e.g., Smith). While the PAM theory does not specify in which circumstances people engage in monitoring, an explanation can be derived from the theory to account for the data. Given that monitoring is resource-consuming, it is possible that engaging in a task demanding less resources, such as making a category judgment for typical exemplars, participants might be more likely to engage in monitoring. More monitoring with the relatively spared resources in the typical nontarget condition would enhance PM performance in that condition. Participants in the atypical nontarget conditions, on the other hand, may have had less resources available given this category judgment task was more demanding (as evidence by longer RTs in categorizing atypical exemplars). Thus, these participants may have had less monitoring. In return, these participants' PM performance was lower compared to that of the participants in the typical nontarget conditions who could monitor more. Related to this possibility, Marsh and Hicks (1998) reported that their participants' PM performance was lower if they had to simultaneously perform a task that is more resource-demanding (e.g., random number generation at a fast pace) than a task that is less resource-demanding task (e.g., random number generation at a slow pace). Although Marsh and Hicks did not have the RT measures with which to compare to our results, it seems possible that when their participants had to engage in a more resource-demanding task, participants might have been less likely to engage in monitoring, which resulted in a lower level of PM performance.

Despite the absence of clear evidence for the discrepancy-plus-search view, it seemed premature to completely rule it out. Rather, we reasoned that the monitoring could have eliminated the chance for discrepancy to facilitate PM performance in Experiment 1. Indeed, when we selectively looked at the people with no/minimal monitoring we found that these participants' PM performance was nominally higher in the discrepant conditions than in the nondiscrepant conditions. Hence, the results suggest that discrepancy-plus-search processes can support PM performance when one is not monitoring. However, the results also suggest that, when a person constantly looks for the PM target during the ongoing activity, he might not need to use a mechanism that spontaneously leads one to recognize the PM target and retrieve the PM intention.

One possible reason for the significant monitoring in Experiment 1 is the number of PM targets we used. Cohen, Jaudas, and Gollwitzer (2008) have found that, when people have to remember more than one PM target, they are likely to engage in monitoring behaviors by showing linearly increasing monitoring cost for increasing number of PM target. Also, Einstein and colleagues (2005, Experiment 3) found significantly more monitoring when people have to remember 6 PM targets than 1 PM target. Thus, while we succeeded in controlling for the potential ceiling in Experiment 1 by using three PM targets, by doing so, we might have encouraged monitoring behaviors in our participants. Hence, in Experiment 2, we attempted to eliminate the monitoring by implementing some changes, such as decreasing the number of PM targets.

Experiment 2

Given that the significant monitoring may have interfered with the effect of discrepancy on PM performance to emerge in Experiment 1, the following changes were made for Experiment 2. First, we reduced the number of PM targets from three to two. We wanted more than one PM target trial to allow us to observe the effect of discrepancy. Yet, we were reluctant to repeatedly present the same PM target multiple times because the repetition may alter the integrity of category typicality of atypical PM targets and subsequently discrepancy experienced with those PM targets. For example, when an atypical PM target is presented repeatedly, participants might find the repeated atypical PM target as typical over time, which may then prohibit participants to find that repeated atypical PM target discrepant from the list of typical nontarget exemplars. Also, we noticed that while having to remember three PM targets stimulate significant monitoring, having to remember two targets stimulated only marginally significant monitoring in Cohen and colleagues' study (2008). Together with Lee and McDaniel's (2010) finding that showed no monitoring with two PM targets, we reasoned that it might be possible that two PM targets would not stimulate any significant monitoring.

Secondly, we changed the PM instruction so that it specified the context in which PM targets to appear: While not specifying the context in which PM targets will appear could have reduced the chance of monitoring when PM targets appear in Experiment 1, it could also have led participants to guess the context in which the PM targets would appear. After solving the first set of tasks, participants might have figured out the structure of the task set and tried to guess when the PM targets might appear. For example, after completing the first set of tasks, consisted of the lexical decision task and the category judgment task, participant may expect to have another lexical decision task followed by another category judgment task. If they do not encounter any PM targets during the second lexical decision task (after encountering no PM target during the first set of tasks), participants may assume that the PM targets will appear during the second category judgment task if they will ever appear. With this reasoning, some participants may have engaged in the monitoring on the onset of the second category judgment task. The higher PM performance for the typical PM targets when the control block was first presented in Experiment 1 hinted at this possibility and we wanted to eliminate this possible monitoring strategy. We reasoned that by specifying the PM context, we may reduce the possibility of people using the contextual information to regulate their monitoring behavior. Also to further investigate any strategic regulation of monitoring behaviors, such as deciding to monitor if a task is perceived to be easy (as hinted in Experiment 1 by faster RTs to typical nontargets), we added in the monitoring strategy questionnaire at the end of the experiment.

In addition to specifying the PM context, we increased the total number of nontargets during the PM block (from 92 trials in Experiment 1 to 148 in Experiment 2). More importantly, we increased the number of nontarget trials preceding the first PM target (from 30 trials in Experiment 1 to 103 trials in Experiment 2). Scullin and colleagues (2010) successfully eliminated monitoring by presenting 500 nontarget trials prior to the first PM target presentation. Because the numbers of normed categories and their exemplars are limited and repeating those items, especially the atypical exemplars, would hinder the manipulation of discrepancy, the present stimulus set did not allow us to match the number of trials preceding the first PM target to that of Scullin and colleagues' study. However, by increasing the number of nontargets preceding the first PM target, we hoped that participants might disengage from the monitoring they may have initiated at the beginning of the category judgment task, possibly prior to the presentation of first PM target. Furthermore, we reasoned that increasing the number of nontargets preceding the PM target may increase the magnitude of discrepancy experienced.

Methods

Participants and design. A total of 88 participants from the Washington University in St. Louis community participated in the experiment in exchange of a partial course credit or monetary compensation. The design was the same as Experiment 1, a 2 x 2 x 2 mixed design with the PM target typicality and the nontarget typicality (typical vs. atypical) as between subjects factors and the block type (PM vs. control) as a within subjects factor.

Materials and procedure. Materials for Experiment 2 were the same as in Experiment 1, except some minor changes described below. First, the number of PM targets was reduced from three to two. PM targets for the typical PM target conditions were "drum" and "bus" while PM targets for the atypical PM target conditions were "harp" and "taxi". The mean length and the mean HAL frequency were 3.5 letters and 9.76 for the typical PM targets and 4 letters and 7.67 for the atypical PM targets, respectively. Also, to increase the number of nontargets in the PM block, the control block was shortened from total of 92 trials to 36 trials. 56 trials from the control block were added into the PM block, making the PM block to have the total of 148 nontarget trials. The numbers of trials in the corresponding lexical decision tasks were also altered accordingly.

Procedure for Experiment 2 had several minor changes as well. First, the shortened control block always preceded the PM block. Because we were limited in the number of stimuli we can use under this paradigm, we could not counterbalance the presentation order of the PM block and the control block with the same number of trials in each block to calculate monitoring cost. Thus, by always presenting the control block prior to the PM block, we could measure the baseline of each participant’s category judgment task independent of any PM intention. After performing a minute of the lexical decision task, participants performed 36 trials of the category decision task (the control block). Then, participants were presented with the PM instruction with the specification of the context in which the PM targets will appear. After each instruction for a task, participants were told to write down the instruction to make sure that they understood the instruction. Participants turned in their instruction summary and then upon the examination of the summary by the experimenter, they were told to press the *1* key to proceed. Then, for 5 minutes, participants performed another round of the lexical decision task. Upon the completion of the lexical decision task, participants solved 150 trials of the category judgment task in which the PM targets were presented on the 104^{th} and the 140^{th} trials. At the end of all the tasks, participants received a questionnaire on potential strategy use on monitoring behaviors during the experiment.

Results

The same response measures, such as the key responses and the RTs, were recorded and analyzed. An alpha level of .05 was set for statistical significance for all statistical analyses, unless noted otherwise. Performance on the lexical decision task, which was used as a distractor task, was not analyzed.

PM performance. The accuracy of PM performance was derived by computing the proportion of correct PM response for the two PM targets. The correct PM response was pressing the *q* key upon the presentation of a PM target exemplar. A 2 x 2 between subjects ANOVA with the PM target typicality and the nontarget typicality (typical vs. atypical) as between subjects factors was conducted on the mean PM performance (see Table 3 for the means). There was no main effect of PM target typicality or nontarget typicality (Fs < 1). Also, while the nominal difference was in the direction predicted as a priori by the discrepancy-plus-search view, such that the mean PM performance in the discrepant conditions (M=.77) was higher than that in the nondiscrepant conditions (M=.73), there was no significant interaction of PM target typicality by nontarget typicality (F < 1).

While the PM performance difference between the discrepant conditions and the nondiscrepant conditions was only nominal, an inspection of the data led us to reason that the effect of discrepancy might be different from the first PM target to the second PM target for the following reason. At the beginning of the task in which the PM target is to appear, people may engage in monitoring. Based on the finding from Experiment 1 showing more monitoring in the typical nontarget conditions, it could be possible that the participants in the less resource-demanding conditions (e.g., typical nontarget conditions) engage in the monitoring more so than the participants in the relatively more resource-demanding conditions (e.g., atypical nontarget conditions). This differential engagement in monitoring then may lead to differential PM performance: more monitoring may lead to higher PM performance in the typical nontarget conditions. After responding to the first PM target, participants may disengage from the monitoring, thinking the chance of encountering another PM target is low. If participants were to disengage from the monitoring, then PM performance could be supported by the discrepancy-plus-search processes. Consequently, discrepancy may then affect PM performance primarily for the second PM target.

To test this idea, we conducted two separate between subjects ANOVAs on mean PM performance: one for the first PM target and another for the second PM target. For both ANOVAs, the PM target typicality and the nontarget typicality (typical vs. atypical) were entered as between subjects factors. For the first PM target, while there were no significant main effects or interaction (Fs < 1), there was a nominal difference between the conditions with atypical nontargets (M=.68) and the conditions with typical nontargets (M=.82). Notice this nominal difference may reflect the possible selective monitoring behavior as described above. For the second PM target, there were no significant main effects or interaction (Fs <1). Yet, interestingly enough, there was a nominal difference on mean PM performance as a function of discrepancy: PM performance for the second PM target was nominally higher in the discrepant conditions (M=.80) than in the nondiscrepant conditions (M=.70).

Category judgment reaction times. To investigate any potential slowing down of the category judgment task due to the PM intention, RTs to make the category judgment task were analyzed. The trimming was done in the same way as described in Experiment 1. Averaged RTs were entered into a 2 x 2 x 2 x 2 mixed measures ANOVA with the response type (yes vs. no) and the block type (PM vs. control) as within subjects factors and the PM target and the nontarget typicality (typical vs.

atypical) as between subjects factors (see Table 4 for the mean). There was a main effect of response type, (F(1, 84)=35.35, MSE=6412.50), in that participants were faster in correctly judging an exemplar belonging to a given category (1082 msec), thus making a yes response, than judging a nonexemplar not belonging to the given category (1141 msec). Also, there was a significant main effect of block type, F(1,84)=196.91, *MSE*=6412.50, in that participants were faster in making the category judgments in the control block (1036 msec) than in the PM block (1188 msec). Also, there was a main effect of nontarget typicality, F(1,84)=14.17, *MSE*= 6412.50, showing faster reaction times for the typical nontargets (1036 msec) than for the atypical nontargets (1187 msec). Note that this result, again, reflects that the experimental manipulation of category typicality was successful. No other effects or interactions were significant.

Planned comparisons were made to look at the monitoring cost for each condition. All of the four conditions were found to have a significant monitoring cost, the smallest F value being F(1,84)=29.52, *MSE*=6412.50. Post-hoc comparisons revealed that the monitoring cost for the discrepant condition with typical nontargets (184 msec) was significantly higher than that for the nondiscrepant condition with atypical nontargets (132 msec), F(1,84)=4.64, *MSE*=6412.50. The monitoring cost for the discrepant condition with atypical nontargets (138 msec) did not differ significantly from that of the nondiscrepant condition with typical nontargets (155 msec), F <1.

In addition to the overall monitoring measures, we wanted to see if the functional monitoring cost differed as a function of discrepancy. Furthermore, we were curious if the functional monitoring cost may show differential monitoring strategies engaged for the first and the second PM targets that showed differential data patterns. Following Scullin et al.'s (2010) procedure, we entered averaged RTs from the five trials preceding each PM target and corresponding trials from the control block (the last five of 36 trials) into a 3 x 2 x 2 repeated measures ANOVA with the trial type (control, first PM, and second PM) as a within subjects factor and the PM target typicality and the nontarget typicality (typical vs. atypical) as between subjects factors. There was a main effect of nontarget typicality, $F(1,84)=6.95$, $MSE=108198.85$, showing that participants were faster at making category judgments for the typical nontarget exemplars (1043 msec) than atypical nontarget exemplars (1149 msec). Also, there was a main effect of trial type, $F(1,168)=28.58$, $MSE=24955.27$. Planned comparisons revealed higher RTs for the five trials preceding the first and the second PM trials (Ms= 1144 and 1152 msec, respectively) compared to the control trials (M=992 msec), $F(1, 168)=40.74$ and 45.14, MSE= 24955.27, respectively. These results indicate that having the PM intention slowed down the ongoing activity in the PM block compared to the control block. No other effects were significant.

Further comparisons found that the functional monitoring cost did not differ as a function of discrepancy for the first and the second PM targets, or as a function of nontarget typicality for the first PM target, Fs< 1.55. However, the functional monitoring cost for the second PM trial was marginally higher in the typical nontarget conditions (M=192 msec) than in the atypical nontarget conditions (M=128 msec), $F(1,168)=3.61$, $MSE=24955.27$, $p=.06$.

Category judgment accuracy. Accuracy of the category judgment task was computed by taking the proportion of correct judgments for each response type (yes vs. no) in each block (PM vs. control). PM target exemplars were excluded in this analysis. Accuracy of the category judgment task was entered into a 2 x 2 x 2 x 2 repeated measures ANOVA with the response type (yes vs. no) and the block type (PM vs. control) as within subjects factors and the PM target typicality and the nontarget typicality (typical vs. atypical) as between subjects factors. There was a significant effect of block type, $F(1,84)=25.66$, $MSE=.002$, in that participants correctly judged more exemplars in the control block ($M=.96$) than in the PM block ($M=.94$), suggesting that having PM intention may hurt the performance accuracy of ongoing activity. Also, there was a significant main effect of nontarget typicality, $F(1,84)=5.68$, $MSE=.002$, in that participants made more correct category decisions for the typical nontarget exemplars ($M=.96$) than for the atypical nontarget exemplars ($M=.94$). This effect, again, proves that our manipulation of exemplar typicality was successful. No other effects or interactions were significant.

Individual differences analysis. With the same reasoning and procedure used in Experiment 1, we computed an individual's monitoring index for each PM target. We used less than 200 msec difference between the PM block and the control block as an arbitrary cut off point in categorizing people exhibiting less monitoring than the others. We could not use the same cut off point used in Experiment 1 because the average individual monitoring index was higher in Experiment 2 than in Experiment 1. For the first PM trial, a total of 12 people were categorized as exhibiting less monitoring: eight people from the discrepant conditions and four from the nondiscrepant conditions. For the second PM trial, a total of ten people were categorized as exhibiting less monitoring: four from the discrepant conditions and six from the nondiscrepant conditions. Only two people showed monitoring less than 200 msec for both PM trials. The mean PM performance for the first PM target was the same between the discrepant and the nondiscrepant conditions (*M*s=.75). The mean PM performance for the second PM was nominally higher in the discrepant condition (*M*=1.0) than in the nondiscrepant condition (*M*=.50).

Discussion

In Experiment 2, we found a number of nominal, but interesting, data patterns. First, we found further support for selective monitoring behaviors found in Experiment 1. There was a higher PM performance in the typical nontarget conditions (*M*=.82) than in the atypical nontarget conditions (*M*=.68) for the first PM target. The functional monitoring cost for the first PM trial did not differ as a function of nontarget typicality, possibly because both discrepant and nondiscrepant conditions already exhibited significant monitoring. Still, higher overall monitoring cost for the typical nontarget conditions, together with the higher PM performance in the typical nontarget conditions, suggests that the property of nontargets (e.g., perceived task difficulty induced by processing typical/atypical category exemplars) might be a factor that people utilize to regulate their monitoring behaviors.

Secondly, we found some support showing discrepancy facilitating PM performance. There was a nominal difference on PM performance for the second PM target as a function of discrepancy (*M*=.80 for the discrepant conditions and *M*=.70 for the nondiscrepant conditions). Also, while only a very small sample size was used (total of ten participants in the discrepant and the nondiscrepant conditions), the individual differences analysis showed that the mean PM performance for the second PM trial in the discrepant conditions was higher than that in the nondiscrepant conditions when we selectively looked at the participants with less monitoring. Given that these patterns were predicted as a priori by the discrepancy-plus-search view (McDaniel & Einstein, 2000; McDaniel et al., 2004), we reject the assumption that the unsuccessful manipulation of discrepancy led to the nominal difference between the discrepant conditions and the nondiscrepant conditions. Instead, we suggest another factor might have been responsible for this nominal difference. While certain features of Experiment 2 were implemented to eliminate any monitoring (e.g., smaller number of PM targets), we failed to minimize the monitoring throughout the experiment. Subsequently, the presence of monitoring could have limited the full emergence of discrepancy processes, leading only to nominally different PM performance between the discrepant and the nondiscrepant conditions. This failure to eliminate monitoring precludes any decisive conclusions about the role of discrepancy on PM performance.

General Discussion

By manipulating the category typicality of PM target and nontarget exemplars, we attempted to extend Lee and McDaniel's (2010) finding supporting discrepancy-plus-search processes in PM performance. The experimental manipulation of category typicality was successful, as indicated by the classic finding of a typicality effect (e.g., shorter RTs to typical exemplars than to atypical exemplars during a task requiring category information; Rosch, Simpson, & Miller, 1976). Unfortunately, however, we failed to eliminate monitoring that can mask the full emergence of discrepancy in PM performance. Below we discuss the findings of the study, possible shortcomings of the data, and suggestions for future studies.

In Experiment 1, we found higher PM performance in the typical nontarget conditions (M=.70) than in the atypical nontarget conditions (M=.54), independent of PM target typicality. Also, the higher PM performance was preceded by significant higher monitoring cost. This finding of higher PM performance in the typical nontarget conditions in the presence of significant monitoring was partially replicated in Experiment 2 for the first PM target. These data patterns seem to suggest that the property (e.g., the ease of processing) of nontargets may have determined PM performance, possibly through regulating monitoring behaviors.

Although the PAM theory (e.g., Smith, 2003) does not explicitly predict higher PM performance in the typical nontarget conditions, one may develop an explanation within the PAM theory for the data patterns just described. Given the shorter RTs to typical nontarget exemplars than to atypical nontarget exemplars during the control block in Experiment 1 and 2, along with the classic findings on the typicality effect, one may assume that resource demands for the processing of the typical nontargets were less than that for the processing of the atypical nontargets. Given participants in the typical nontarget conditions had more resources available than participants in the atypical nontarget conditions, the PAM theory may suggest that the higher PM performance in the former could have been facilitated by monitoring afforded by spared resources. In line with this idea, studies have found higher PM performance in the condition where participants are asked to perform a task that is relatively less demanding compared to PM performance in the condition where participants are asked to perform a task that is relatively more demanding (e.g., Marsh & Hicks, 1998). One interesting question to follow-up is whether or not people can recognize that they have spared resources to engage in resource-consuming monitoring behaviors to perform a PM task in a less resource-demanding setting. It is possible that people in the typical nontarget exemplar conditions voluntarily engaged in monitoring after they realized the task was easy. Another possibility is that spared resources in the typical nontarget exemplar conditions allowed monitoring to be efficient in that condition. In the former case, one may emphasize the importance of understanding the task demand of ongoing activities while in the latter one may emphasize the importance of decreasing the task demand of ongoing activities on performing a PM task.

In addition to the question how people may capitalize on the spared resources when performing a PM task, another factor to consider for future study is whether individual differences play a role in the use of monitoring strategies. Particularly in Experiment 1, we observed that individuals engaged differently in monitoring behaviors: some participants exhibited less than 30 msec monitoring for all three PM trials and others exhibited significant monitoring for all trials while the rest of the participants varied in their monitoring behaviors. Although it is premature to argue that these data indicate stable individual differences in engaging in monitoring behaviors, we suggest that, at least for the people who showed consistently less/more monitoring over three PM trials, some individual differences could have led them to employ a particular strategy to monitor or not to monitor. Unfortunately, we could not further test this idea as the paradigm was not designed to investigate the individual difference factors and how they may influence monitoring strategies, and subsequently, PM performance in different settings. Future studies are needed to better understand this relationship between individual difference factors and the use of monitoring strategies.

In Experiment 1 and 2, we failed to find significant support for the discrepancy-plus-search view (Einstein & McDaniel, 2000; McDaniel et al., 2004). One possibility for the lack of a significant effect of discrepancy on PM performance is that experimental manipulation of category typicality of nontargets and PM targets was ineffective. This is unlikely for the following reasons. First, pilot studies conducted prior to Experiment 1 and 2, using the same stimuli set as Experiment 1 and 2, found that participants rated the typical exemplars more typical than the atypical exemplars. Second, RT to make the category judgments in Experiment 1 and 2, along with the response in the pilot studies, showed shorter RTs toward the typical exemplars than the atypical exemplars. Replicating the classical typicality effect (e. g., Rosch, Simpson, & Miller, 1976), these RT measures indicate that the experimental manipulation was effective in inducing different levels of typicality between two kinds of exemplars (typical vs. atypical). Lastly, differential PM performance as a function of nontarget typicality (higher PM performance in the typical nontarget conditions than in the atypical nontarget conditions) also suggests that the experimental manipulation differentiated conditions (typical vs. atypical nontarget conditions). Thus, the failure to find an effect of discrepancy on PM performance was likely not because the manipulation of typicality was too weak.

While the experimental manipulation of typicality was successful, it remains possible that the processing difference between typical versus atypical PM targets was not substantial enough to induce a discrepancy. However, a number of data patterns from Experiment 1 and 2 leads us to consider another possibility. First, there was a nominal difference in PM performance as a function of discrepancy in Experiment 2, particularly for the second PM target (*M*s=.80 and .70 for the discrepant and the nondiscrepant conditions, respectively). Given that the aforementioned pattern of higher PM performance in the discrepant conditions was predicted a priori by the discrepancy-plus-search view (McDaniel et al., 2004), we find it unlikely that the nominal difference was random. Rather, we entertain the possibility that the monitoring might have minimized the chance for the discrepancy-plus-search processes to support any PM performance.

According to the multi-process theory (Einstein et al., 2005), discrepancy-plus-search processes support the recognition of PM target and the retrieval of PM intention when one is not engaged in the monitoring processes. Given the presence of significant monitoring in Experiment 1 and 2 (as indicated by cost), participants could have performed the PM task without the support from the discrepancy-plus-search processes. When participants disengage from monitoring, however, discrepancy-plus-search processes may emerge to facilitate the recognition of PM target and the retrieval of PM intention. The nominally higher PM performance in the discrepant conditions for the second PM target in Experiment 2 may reflect this possibility. Also reflecting this possibility was the higher PM performance in the discrepant conditions than in the nondiscrepant conditions when we selectively looked at the people with less monitoring (particularly in Experiment 1).

Though we implemented some features to minimize monitoring, we could not eliminate monitoring in the present paradigm. Though we decreased the number of PM targets from three in Experiment 1 to two in Experiment 2, monitoring was still present. Also, presenting more than 100 nontarget trials prior to the first PM target presentation in Experiment 2 failed to eliminate monitoring. To control for monitoring in a future study using this paradigm, we plan to limit the number of PM targets to one and to increase the number of nontarget trials preceding the first PM target presentation. Cohen and colleagues'(2008) finding of marginally significant monitoring cost even with two PM targets suggests that the use of two different PM targets might have been sufficient to encourage people to monitor in the present paradigm. By using one PM target for several target trials, one may decrease the likelihood of people engaging in monitoring behaviors while controlling for the ceiling on PM performance. Additionally, by presenting many nontarget trials prior to the first PM target, one may minimize the likelihood of people staying engaged in monitoring as in Scullin and colleagues' (2010) study, which showed minimal monitoring for the PM target presented following 500 nontarget trials. Given the limited number of normed categories and exemplars, the present paradigm could have only 103 nontarget trials precede the first PM target, which is a far smaller number Scullin and colleagues used. Moreover, the present study could not increase the number of nontarget trials by repetition because the repetition of atypical exemplars may easily compromise the manipulation of discrepancy in certain conditions: presenting a typical PM target after presenting a list of "repeated" atypical exemplars may no longer serve as a discrepant condition given that the repetition may alter the ease of processing repeated atypical exemplars. For one possible way to increase the number of nontargets preceding the first PM target without repeating the same nontarget exemplars, one may begin the PM block with moderately typical exemplars for a number of trials, and then, present typical or atypical nontargets to induce discrepancy. By doing so, we may lower the likelihood of people staying engaged in monitoring until the presentation of first PM target without contaminating the integrity of category typicality of exemplars by repetition.

A PM target event is never presented in isolation, but is processed under the influence of many different factors (e.g., resource demanded by an ongoing activity) in real life settings. Yet, often, many theories make predictions, with the primary focus on the property of PM target. Perhaps more fruitful predictions may be made if some of the existing theories start to incorporate how the property of PM targets interacts with other factors that may influence PM performance, such as the property of nontargets or individual differences in employing monitoring strategies.

References

Battig, W. R, & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monographs, 80,* 1-46.

Breneiser, J., & McDaniel, M. A. (2006). Discrepancy processes in prospective memory retrieval. *Psychonomic Bulletin & Review, 13(5),* 837-841.

Cohen, A., Jaudas, A., & Gollwitzer, P. M. (2008). Number of cues influencs the cost of remembering to remember. *Memory & Cognition, 36,* 149-156.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior, 8,* 240-247.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. Journal of Experimental Psychology: *Learning, Memory, & Cognition, 16,* 717–726.

Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., & Breneiser, J. (2005). Multiple processes in prospective memory retrieval: Factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General, 134,* 327-342

Guynn, M., J., & McDaniel., M., A. (2007). Target preexposure eliminates the effect of distraction on event-based prospective memory. *Psychonomic Bulletin & Review, 14,* 484-488.

Jacoby, L., L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 21-38.

Jacoby, L., L. & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 3*, 306–340.

Jacoby, L., L. & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, *118*, 126–135.

Lee, J. & McDaniel, M. A, (2010). Wait, there is something special about you: Discrepancy-plus-search view on the retrieval of Prospective Memory intention. Unpublished manuscript.

Marsh, R. L., Hicks, J. L., & Bink, M. L. (1998). The activation of completed, uncompleted, and partially completed intentions. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *24*, 350-361.

McDaniel, M. A. (1995). Prospective memory: Progress and processes. In D. L. Medin (Ed.), *The psychology of learning and motivation, 33*, 191-221. San Diego: Academic Press.

McDaniel, M. A., & Einstein, G. O. (1993). The importance of cue familiarity and cue distinctiveness in prospective memory. *Memory, 1*, 23–41.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology, 14*, 127-S144.

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field*. Thousand Oaks, CA: Sage.

McDaniel, M. A., Guynn, M. J., Einstein, G. O., & Breneiser, J. E. (2004). Cue focused and automatic-associative processes in prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 605–614.

Marsh, R. L., & Hicks, J. L. (1998). Event-based prospective memory and executive control of working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 24, 336-349.

Marsh, R. L., Hicks, J. L., & Cook, G. I. (2006). Task interference from prospective memories covaries with contextual associations of fulfilling them. *Memory & Cognition, 34*, 1037-1045.

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573-605.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 491-502.

Scullin, M.K., McDaniel, M.A., & Einstein, G.O. (2010). Control of cost in prospective memory: Evidence for spontaneous retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 190-203

Scullin, M.K., McDaniel, M. A., Shelton, J. T., & Lee, J. (2010). Focal/nonfocal cue effects in prospective memory: Monitoring difficulty or different retrieval processes? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36(3)*, 736-749.

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, and Cognition, 29,* 347–361.

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 756–777.

Smith, R. E., Hunt, R. R., McVay, J. C., & McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 734-746.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289–335.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? The unexpected basis of feelings of familiarity. *Acta Psychologica, 98,* 141-166.

Whittlesea, B. W. A., & Williams, L. D. (2001a). The discrepancyattribution hypothesis: I. The heuristic basis of feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 3-13.

Whittlesea, B. W. A., & Williams, L. D. (2001b). The discrepancyattribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 14-33.

Table 1. Mean proportion of correct PM performance as a function of the PM target typicality and the notarget typicality. Standard deviations in parenthesis.

	Typical PM target	Atypical PM target
Typical nontarget	.77 (.27)	.62 (.39)
Atypical nontarget	.51 (.37)	.53 (.41)

Table 2. Mean Reaction Times (in milliseconds) for the PM Block and the control block as a function of the PM target typicality and the nontarget typicality. Standard deviations in parenthesis. T- referring to Typical and A- referring to Atypical. –list referring to nontargets.

	T-target in T-list	A-target in T-list	T-target in A-list	A-Target in A-list
PM Block	1273 (221)	1070 (148)	1249 (223)	1147 (159)
Control Block	1103 (179)	979 (130)	1209 (181)	1095 (191)

Table 3. Mean proportion of correct PM performance as a function of the PM target typicality and the notarget typicality. Standard deviation in parenthesis.

	Typical PM target	Atypical PM target
Typical nontarget	.75 (.40)	.80 (.33)
Atypical nontarget	.77 (.25)	.71 (.37)

Table 4. Mean Reaction Times (in milliseconds) for the PM Block and the control block as a function of the PM target typicality and the nontarget typicality. Standard deviations in parenthesis. T- referring to Typical and A- referring to Atypical. –list referring to nontargets.

	T-target in T-list	A-target in T-list	T-target in A-list	A-Target in A-list
PM Block	1098 (135)	1121 (169)	1285 (227)	1248 (282)
Control Block	943 (101)	983 (151)	1101 (208)	1116 (247)