REMEMBERING THE FUTURE: INDIVIDUAL DIFFERENCES IN METACOGNITIVE REPRESENTATION PREDICT PROSPECTIVE MEMORY PERFORMANCE ON TIME-BASED AND EVENT-BASED TASKS IN EARLY CHILDHOOD

by

Kayla B. Causey

A Dissertation Submitted to the Faculty of
The Charles E. Schmidt College of Science
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy

Florida Atlantic University
Boca Raton, FL
December 2010
Copyright by Kayla B. Causey, 2010
REMEMBERING THE FUTURE: INDIVIDUAL DIFFERENCES IN METACOGNITIVE REPRESENTATION PREDICT PROSPECTIVE MEMORY PERFORMANCE ON TIME-BASED AND EVENT-BASED TASKS IN EARLY CHILDHOOD

by

Kayla B. Causey

This dissertation was prepared under the direction of the candidate’s dissertation advisor, Dr. David F. Bjorklund, Department of Psychology, and has been approved by the members of her supervisory committee. It was submitted to the faculty of the Charles E. Schmidt College of Science and was accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

SUPERVISORY COMMITTEE:

David F. Bjorklund, Ph.D.
Dissertation Advisor

Robert Stackman, Ph.D.

Marissa Greif, Ph.D.

Alan Kersten, Ph.D.

Julia Earles, Ph.D.

David L. Wolgin, Ph.D.
Psychology Department Chair

Gary W. Perry, Ph.D.
Dean, Charles E. Schmidt College of Science

Barry T. Rosson, Ph.D.
Dean, Graduate College

November 8, 2010
Date
ACKNOWLEDGEMENTS

I sincerely thank my family and husband for their constant support and confidence. In addition, I wish to acknowledge Aaron Goetz, Chris Chicoyne, Eric Weber, Amanda Farrell, Karen Pinto, and Jaime Sanders, as well as the faculty and staff at the Henderson School and Slattery Preschool, for their assistance in data collection. I also wish to recognize my indebtedness and appreciation to Sam Mueller for programming the CyberCruiser 2.0. Finally, I would like to acknowledge and thank my advisor, Dr. David F. Bjorklund, for allowing me the freedom to explore my own interests, for providing me with so many learning opportunities along the way, and for his unwavering patience and gentle guidance in the process.
ABSTRACT

Author: Kayla Causey
Title: Remembering the Future: Individual Differences in Metacognitive Representation Predict Prospective Memory Performance on Time-based and Event-based tasks in Early Childhood
Institution: Florida Atlantic University
Dissertation Advisor: Dr. David F. Bjorklund
Degree: Doctor of Philosophy
Year: 2010

Prospective memory is remembering to perform an action in the future, such as attending a meeting (a time-based task) or picking up milk at the gas station (an event-based task), and is crucial to achieving goal-directed activities in everyday life. Children who fail to develop prospective memory abilities are likely to experience difficulties interacting with parents, teachers, and peers. To date, research on prospective memory development has been primarily descriptive or focused on underlying executive functioning. This dissertation investigated the developmental relationship between metacognitive representation and prospective memory in preschool and elementary school children and adults. Findings from Study 1 indicated that individual differences in representational ability independently predicted individual differences in 3-year-olds’ performance on event-based tasks that are of low-interest. Qualitative changes are
important to consider when modeling prospective memory development, as with episodic memory. Study 2 presents findings based on a study using the CyberCruiser 2.0, an Xbox-style racing game designed to assess time-based prospective memory. This study confirmed that kindergarten children are capable of completing this time-based prospective memory task but revealed that performance improved with age. Between kindergarten and 2nd grade, children become better aware of their own mental processes and abilities, allowing them to adjust their strategies and perform more comparable to adults. As a result, in this study, younger children tended to overestimate their prospective memory abilities and were less likely to monitor passing “time,” causing them to fail more time-based task trials than older children and adults. Similarly, participants who underestimated the costs of prospective memory failed more time-based tasks relative to those who more accurately assessed these costs. Although this latter relationship was limited to adults, it suggests that a poor metacognitive understanding of the costs of prospective memory may result in missed opportunities to carry out a delayed intention if individuals fail to allocate attentional resources appropriately. These findings have theoretical implications for models of prospective memory and development. Practical implications for educating children are also discussed.
REMEMBERING THE FUTURE: INDIVIDUAL DIFFERENCES IN METACOGNITIVE REPRESENTATION PREDICT PROSPECTIVE MEMORY PERFORMANCE ON TIME-BASED AND EVENT-BASED TASKS IN EARLY CHILDHOOD

List of Tables ......................................................................................................................... ix
List of Figures ........................................................................................................................ xi

Chapter One:

Introduction .............................................................................................................................. 1
Theoretical Models of Prospective Memory ................................................................. 4
Prospective Memory Development ..............................................................................13

Chapter Two:

Study One: Individual differences in representational ability and the prospective memory of three-year-old children .................................................................24
Method .................................................................................................................................26
Results .................................................................................................................................36
Discussion ............................................................................................................................49

Chapter Three:

Introduction to Time-based Prospective Memory ..................................................61
Beyond Executive Functioning ..................................................................................68

Chapter Four:
Study Two: Time-based prospective memory, metacognition, and strategy-use in elementary-aged children and adults ................................................................. 78
Method ........................................................................................................... 79
Results ........................................................................................................... 89
Discussion ...................................................................................................... 112

Chapter Five:

Final Comments and Future Directions ......................................................... 130
Appendix A .................................................................................................... 167
Appendix B .................................................................................................... 168
Appendix C .................................................................................................... 169
References .................................................................................................... 170
LIST OF TABLES

Table 1. Mean Performance on Reminding Tasks for 2-, 3-, and 4-year-olds, from Somerville, Wellman, and Cultice (1983, p. 91) ..........................................................136

Table 2. Results from Guajardo and Best (2000, p. 86) .....................................................137

Table 3. Counterbalancing of Executive Functioning Task Sets with Prospective Memory Tasks Across Session .................................................................................138

Table 4. Correlations and Descriptive Statistics for Performance on Measures of Executive Functioning ........................................................................................................138

Table 5. Correlations for Prospective Memory Performance and Measures of Executive Functioning ........................................................................................................140

Table 6. Factor Loadings and Communalities based on a Principal Components Analysis with Varimax Rotation for Five Items (N = 31).........................................................141

Table 7. Hierarchical Regression Analysis Predicting Individual Differences in Overall Prospective Memory Performance of Children..............................................142

Table 8. Hierarchical Regression Analysis Predicting Individual Differences in the Number of Cues Children Received on Prospective Memory Tasks varying by Interest Level ........................................................................................................143

Table 9. Hierarchical Regression Analysis Predicting Individual Differences in the Number of Cues Children Received on Prospective Memory Tasks varying by Instruction ...........................................................................................................144
Table 10. Hierarchical Regression Analysis Predicting Individual Differences in the Relative Difficulty with which Children Passed each Prospective Memory Task ..................................................................................................................145

Table 11. Mean Missed Refill Opportunities, Presented as a Three-way Interaction Between Gender, Age, and Metacognitive Condition..................................................146

Table 12. Mean Accuracy of Prospective Memory Performance Estimates by Age, Block Order, and Gender ........................................................................................................147

Table 13. Mean Crash Rates for Participants, presented by Block and Metacognitive Presentation Condition ..........................................................................................................148

Table 14. Percent (and number) Estimating Direction of Prospective Memory Cost, by Age Group ..................................................................................................................149

Table 15. Percentage of Participants Using each Strategy in the Final Round of Racing, broken down by Age Group .................................................................150

Table 16. Distribution of Strategy Use across Gender ..............................................................................................................................151

Table 17. Mean Missed refills for each Age Group, presented by Block and Strategy Use ...........................................................................................................................152

Table 18. Mean Accuracy of Participants’ Time Estimation Metacognition, presented by Age and Time of Estimate ................................................................................153
LIST OF FIGURES

Figure 1. Percentage of participants passing each prospective memory task ..............154

Figure 2. Percent within each gender passing prospective memory task ..................155

Figure 3. Mean number of cues (1 – 4) presented by prospective memory task type.
  Error bars represent 5% of the mean. .................................................................156

Figure 4. Mean number of cues received by younger versus older children as a
  function of prospective memory instruction. Error bars represent 95% confidence intervals .................................................................157

Figure 5. Percentage of participants passing each measure of belief/desire reasoning
  (n = 31) ..............................................................................................................158

Figure 6. Design of Study 2 ..................................................................................159

Figure 7. Mean number of prospective memory trials where participant ran out of
  gas, presented by age group. Error bars represent 95% confidence interval......160

Figure 8. Accuracy of estimates of prospective memory performance, presented by
  age group. Error bars represent 95% confidence intervals.................................161

Figure 9. Mean gas-checking frequency during first minute of prospective memory
  block, presented by age. Error bars represent 95% confidence intervals............162

Figure 10. Relationship between participants’ accuracy in predicting prospective
  memory abilities and monitoring .................................................................163

Figure 11. Mean checking frequency at each 5-second interval during the wait period .164
Figure 12. Mean lap time for each block (in seconds) presented by gender and age group.

Figure 13. Partial relationship between adults’ accuracy in predicting the costs to their lap time as a result of the prospective memory task and their prospective memory performance, controlling for KBIT scores.
CHAPTER 1
INTRODUCTION

Prospective memory is remembering to perform an action in the future and is crucial to achieving goal-directed activities in everyday life. Ellis and Kvavilashvili (2000) have outlined prospective memory tasks as requiring (1) a delay between the formation of an intention and the opportunity to carry it out, (2) the absence of an explicit reminder to carry out the task at the appropriate moment, and (3) the need to interrupt one’s ongoing activity in order to carry out the intention. Einstein and McDaniel (1990) further distinguished between two types of prospective memory; Event-based prospective memory entails remembering to perform a particular action when cued by a target event (e.g., giving a message to a colleague when you next see him or her), while time-based prospective memory requires remembering to execute an intended action after the passage of a certain amount of time or at a specific moment (e.g., returning a phone call in five minutes).

Prospective memory is the most common memory failure among adults (Crovitz & Daniel, 1984; Kliegel & Martin, 2003) and children (Winograd, 1988) and thus has garnered substantial attention in cognitive psychology (e.g., Kliegel, McDaniel, & Einstein, 2008). Interest in the development of these abilities has also increased among developmental psychologists within the past decade (see Kvavilashvili, Kyle, & Messer, 2008, for a review). While prospective memory is critical to functional life as an adult,
young children must also begin to remember to do things on their own and are sometimes relied on to remind others to complete tasks, particularly once they enter school: remember to do your homework assignment; remember to have your mom sign the permission form; remember to remind your teacher about your allergy. Children who fail to develop prospective memory abilities are likely to experience difficulties interacting with parents, teachers, and peers (McCauley & Levine, 2004; Meacham & Leiman, 1982). Thus, the study of prospective memory development in children has important educational and social implications.

Consider that prospective memory is primarily distinguished from retrospective memory by the absence of an explicit retrieval cue; as Einstein and McDaniel have described, “no one is there to put you in a retrieval mode when the target event occurs… thus, a key question in the prospective memory arena is how, in the absence of a direct request to search memory, the cognitive system supports retrieval of the intended action at the appropriate moment” (Einstein & McDaniel, 2010, p. 1082). The research presented here investigated this question by examining the development of cognitive abilities supporting event- and time-based prospective memory, specifically, the role that metacognitive representations play in prospective memory.

Metacognition, “knowing about knowing,” has received attention regarding clinical applications, particularly among psychologists studying depression (e.g., Teasdale, Moore, Hayhurst, Pope, Williams, & Segal, 2002), anxiety (e.g., Wells, 1995), and attention, as well as its applications in education and instruction (Borkowski, Carr, & Pressely, 1987; Brown, 1987; Carr, Kurtz, Schneider, Turner, & Borkowski, 1989; Roberts & Erdos, 1993; Sternberg, 1986; Van Zile-Tamsen, 1996). Metacognition has
recently become mainstreamed as “mindfulness” and applied in meditation practices and teaching (e.g., Baer, 2003). Collectively, these studies have revealed that the practice of conscious awareness of one’s thought processes improves mood, attention, and concentration, and improved problem solving and critical thinking abilities, among other benefits. This ability also appears to be uniquely human (Metcalfe & Shimamura, 1994). However, the role metacognitive processes play in prospective memory has yet to be investigated.

Like metacognition, the ability to project oneself into the future, to use foresight in planning, and to reconstruct past events is arguably unique to humans and requires a representational capacity (Suddendorf & Busby, 2003; Suddendorf & Corballis, 2007, 2008). Broadly, the ability to represent oneself in past episodes (classic episodic memory) and imagine oneself in future scenarios is referred to as “mental time travel” (Suddendorf & Corballis, 1997). To the extent that prospective memory is a form of mental time travel, it should rely on a similar underlying representational capacity. Examination of the developmental relationship between these processes in humans may reveal how these processes were related in human phylogeny and inform our understanding of mental processes involved in prospective memory, a topic about which there is much debate (e.g., Einstein & McDaniel, 2010; Smith, 2010).

In this dissertation, Study 1 examines individual differences in three-year-old children’s ability to remember versus remind another person to carry out event-based tasks of varying interest level, taking into account children’s executive functioning and representational abilities. Study 2 investigates the influence of metacognitive abilities (e.g., people’s understanding of the costs associated with carrying out a prospective
memory task and how well they expect to perform), motivational factors, and strategic monitoring and strategy use on time-based tasks in early childhood and young adulthood.

In the next section, I outline theoretical accounts of prospective memory and review studies describing its development, focusing primarily on event-based prospective memory, in the context of these theoretical frameworks. Throughout the literature review, I describe empirical questions raised by this literature and describe the hypotheses that follow. In Chapter 2, I address these hypotheses, presenting a study examining hypotheses on the development of prospective memory and representational ability in 3-year-olds. Chapter 3 provides further introduction to time-based prospective memory and reviews recent work on its development. Chapter 4 presents the results of research on the relationship between time-based prospective memory, metacognition, and strategy-use in early elementary-aged children and adults. The final chapter offers final comments and future directions for the study of prospective memory and its development.

**Theoretical Models of Prospective Memory**

Prospective memory is most parsimoniously conceptualized as a three-phase model. During phase one, an intention is formed. Phase two entails the maintenance or recall of this intention, and phase three involves a “switch” from the ongoing task at the appropriate time or event to execute the intention (see, e.g., Ellis, 1996, and also Kliegel, Martin, McDaniel, & Einstein, 2002; McDaniel, Glisky, Rubin, Guynn, & Routhieaux, 1999, for alternative models that include more or fewer “phases”). Research has primarily focused on identifying the processes underlying each of these phases; in fact, much of what was known about prospective memory prior to the “Smith vs. Einstein and
McDaniel era” came from cognitive studies on task-switching and planning. From these fields, two main theoretical accounts of prospective memory emerged.

The executive functioning account of prospective memory is based on research indicating that a limited pool of cognitive resources is recruited for task-switching (i.e., the third phase of prospective memory, described above; Allport, Styles, & Hsei., 1994; Mayr, 2002; Meuter & Allport, 1999; Wong & LeBoe, 2009) and intention maintenance (i.e., phase two, above; Smith, 2003, 2008). Alternatively, the episodic processing account is based on research suggesting that encoding and retrieval processes during intention formation (i.e., phase one) and execution (i.e., during phase three) are most important (Allport & Wylie, 2001; Leboe, Whittlesea, & Milliken, 2005; Rothermund, Wentura, & De Houwer, 2005; Waszak, Hommel, & Allport, 2003; Wylie & Allport, 2000).

According to the executive functioning framework, successful switching from an ongoing task to a prospective memory task is, in part, dependent on self-initiated and effortful monitoring of one’s environment in search of a target event (Burgess & Shallice, 1997; Guynn, 2003; Shallice & Burgess, 1991; Smith, 2003). The mental processes proposed to underlie monitoring are executive functions, those general processes -- including attention, working memory, and inhibition -- that siphon from a limited pool of resources known as the central executive. Because these processes are assumed to draw on limited-capacity resources, performance of ongoing tasks may inhibit monitoring, causing an individual to fail to complete the prospective memory task. Conversely, when attention is directed toward maintaining a prospective memory intention or monitoring
the environment for prospective memory target events, individuals may suffer costs to competing tasks that could be performed in response to the same stimulus.

Following this framework, Smith and colleagues (Smith, 2003; Smith & Bayen, 2004) proposed the preparatory attention model (PAM), arguing that once a person forms a prospective memory intention, he or she puts him or herself in a retrieval mode, activating processes that maintain the intention and monitor the environment for the target event. This level of activation can range from conscious strategic monitoring to preparatory attentional processes that are outside awareness (Smith, Hunt, McVay, & McConnell, 2007). Critically, PAM assumes that prospective memory retrieval can only occur when these executive processes are engaged because they are required to initiate retrospective memory checks to determine whether events in the environment are targets.

The most direct support for this hypothesis comes from work indicating that participants suffer costs to ongoing task performance when in “prospective memory mode” compared to when they are not. Smith (2003) demonstrated that when participants are provided with a prospective memory task, they perform poorly on ongoing tasks that rely on the same limited pool of executive resources. However, when participants are not given the prospective memory intention, their performance on the same ongoing task is not decremented. In Smith’s study, participants performed a lexical decision task 300 ms slower, on average, when given a concurrent prospective memory task compared to when they were not asked to do the prospective memory task. Others (Marsh & Hicks, 1998) have shown that increasing the demands of the ongoing task often lowers prospective memory performance, providing further support that prospective memory relies on a limited pool of resources.
Additional support comes from research citing that adult frontal lobe patients, who have impaired executive functioning, perform poorly on multitask laboratory tests and in their everyday lives (Burgess, Veitch, Costello, & Shallice, 2000). Further support for this relationship is provided when one collectively considers studies demonstrating ongoing working memory and executive functioning development across childhood (Davidson, Amso, Anderson, & Diamond, 2006; Gathercole, Pickering, Ambridge, & Wearing, 2004; Zelazo & Müller, 2002), with studies revealing young children’s marked difficulties on goal-maintenance tasks (Marcovitch, Boseovski, & Knapp, 2007) and those indicating marked declines in children’s prospective memory performance when additional delay is introduced in event-based prospective memory tasks (Kvavilashvili et al., 2008; Rendell, Vella, Kliegel, & Terrett, 2009).

In contrast to the executive functioning model, the episodic processing framework assumes that the mental representations that matter most for prospective memory success are the representations individuals generate at encoding (i.e., when the intention is established). Accordingly, the accessibility of the memory representation for the prospective task influences whether it will be completed and the extent to which it will interfere with an ongoing task. From this basic premise, the likelihood or ease with which individuals switch from an ongoing task to perform a prospective memory task depends on factors known to determine performance in episodic memory tasks, such as the elaborateness of encoding, the familiarity, specificity, and distinctiveness of the target event, the degree of contextual overlap between encoding and retrieval conditions (i.e., the extent to which processes required by these tasks are similar and therefore remain accessible), and the recency of the intention formation. For example, switching from an
ongoing task to a prospective memory task is likely to fail if the processes required for the ongoing task are contextually different from those required for the prospective memory task (Wong & Leboe, 2009).

Versions of this theory have emerged, emphasizing the extent to which remembering is assumed to rely on automatic versus self-initiated retrieval processes. The *simple activation* or *automatic associative activation* model hypothesizes that intention retrieval results from automatic search in response to the target event (Einstein & McDaniel, 1996; Nowinski & Dismukes, 2005), whereas the *familiarity plus search* model posits that the target event automatically initiates a controlled search for the associated action (Einstein & McDaniel, 1996).

McDaniel and Einstein (2000) conceptualized the contextual overlap between prospective memory and ongoing tasks as a *focal/nonfocal* distinction. Prospective memory tasks are *focal* if the ongoing task encourages processing of the cue in a way that closely matches how it was processed at encoding. For example, a prospective memory task would be considered focal if it required participants to respond to the word “cat” during an ongoing lexical decision task because participants are already engaged in processing the meaning of stimuli. However, a prospective memory task requiring participants to respond to a word in the “animal” category would be considered nonfocal because the lexical decision task does not require categorical processing of the stimuli (see Einstein & McDaniel, 2005, for more examples). Following the encoding specificity principle (Tulving, 1983), features of the target event or stimulus that are processed at retrieval need to match those that were processed at encoding for successful retrieval to occur. According to McDaniel and Einstein, this is particularly important in prospective
memory tasks that, unlike cued recall, must be accomplished without prompting to conduct a controlled search.

By reconceptualizing the focal/nonfocal distinction of prospective memory tasks, Einstein and McDaniel were able to put forth a multiprocess framework, arguing that both episodic and executive functioning processes can accomplish prospective memory, dependent on the circumstances. Like the PAM theory, the multiprocess theory assumes that participants can engage in preparatory attentional processes, monitoring the environment for target events. In addition, based on participants’ anecdotal reports that the prospective memory intention “popped” into mind, the multiprocess framework also assumes that prospective memory can be achieved by spontaneous retrieval processes that may be less costly, dependent on the circumstances. This hypothesis is derived from the perspective that prospective memory is important for human functioning, but that monitoring and relying on preparatory attentional processes can be costly, and therefore, it is adaptive to be able to rely on multiple processes for prospective memory retrieval that are context-specific (Einstein & McDaniel, 2010). For example, when the prospective memory task is focal, meaning the overlap between retrieval and encoding contexts is sufficient, retrieval of the prospective memory intention can occur in the absence of preparatory attentional processes.

Moreover, because there may be a bias toward spontaneous retrieval, if participants’ encoding of the intention is sufficiently deep to form a good association between the target cue and the intended action (e.g., if the prospective memory task was simulated or imagined), spontaneous retrieval of the intention may occur in the absence of preparatory attentional processes, even when tasks are nonfocal. In this case, the
representation remains accessible enough to overcome the interference of the ongoing
task and achieve spontaneous activation by the target event or stimulus, which can occur
automatically or as the result of a *familiarity plus search* process. Other factors that may
influence reliance on spontaneous retrieval include the relative importance of the
prospective and ongoing tasks and the personal inclinations of the participant
(dependencies reviewed in McDaniel & Einstein, 2007), including metamemory
assumptions (Marsh, Cook, & Hicks, 2006).

Support for the multiprocess framework comes from consistent observations of no
significant costs and accurate prospective memory performance when the prospective
memory task involves a single focal target (Einstein et al., 2005). Moreover, when
Einstein, McDaniel, and colleagues (Einstein et al., 2005, Experiment 5) asked
participants to suspend the prospective memory task while completing a lexical decision
task (to use Smith’s terminology, when they took participants out of “prospective
memory mode”) but presented a prospective memory target, participants’ response times
slowed to prospective memory targets (albeit by only 55 ms). These findings suggest that
the intention was reflexively retrieved when the target was presented, in the absence of
preparatory attentional processes.

Additional support for the multiprocess framework comes from work on aging and
prospective memory (Rendell, McDaniel, Forbes, & Einstein, 2007) demonstrating that
age-related differences in prospective memory were slight when the target event was
focal (favoring younger adults), but more pronounced when the target event was not focal
to the ongoing activity. These findings suggest that the processes underlying performance
on prospective memory tasks involving focal versus nonfocal targets are differentially
susceptible to age-related declines; declines associated with nonfocal targets are due to declines in the resources needed for monitoring, such as executive functioning (Craik, 1986; see also age-related declines in speed of processing, Salthouse, 1991; 1996), whereas the lack of age-related declines for focal tasks indicates reliance on a different set of processes, ones that do not decline with age.

Thus, our theoretical understanding of the processes underlying prospective memory is developing at a rapid pace. While the multiprocess framework has helped to streamline an overall theoretical account of prospective memory, identifying the circumstances under which the executive function or the episodic processing framework is best suited to explain task-switching costs and prospective memory performance still poses somewhat of a challenge. These two theoretical approaches often make quite similar predictions and are both equally capable of accounting for the basic phenomenon of prospective memory failure and costs to ongoing tasks. That prospective memory suffers from a bit of an identity crisis only perpetuates this challenge. Some have argued prospective memory is no different than episodic memory (Roediger, 1996) and that attempting to distinguish it is futile (Crowder, 1996). Graf and Uttl (2001) attempted to address this issue, promoting a “divide and conquer strategy” (p. 438) to the study of prospective memory that mirrors the approach adopted to study retrospective memory, and advising researchers to propose distinct subdomains (cf. episodic and semantic memory) and identify precisely which subdomain is targeted by each investigation.

In this effort, Graf and Uttl (2001) have regarded monitoring (i.e., vigilance) and “prospective memory proper” as part of a continuum of possible prospective memory functions. At one end of this continuum – monitoring – the prospective task dominates
working memory and conscious awareness during the retention interval. At the other end—the prospective memory proper end—the prospective memory task is out of working memory during the retention interval and conscious awareness is focused on competing activities. Graf and Uttl proposed that what varies along the continuum is the proportion of available processing resources that are allocated to the prospective task during the retention interval. For monitoring (i.e., vigilance), all or most of the available resources are allocated to the prospective task, whereas for prospective memory proper, all or most of the available resources are allocated to competing activities. Other tasks, such as remembering to change a sign on the door, may fall between these extremes, recruiting a more equal allocation of resources to prospective and other activities. In this way, prospective memory proper is further distinguished from retrospective memory in that a plan or intention does not remain active in working memory throughout the delay interval. Researchers are then charged with describing and demonstrating what factors influence where a task falls on the prospective memory continuum.

Investigations of prospective memory development may inform the working definition of prospective memory by addressing what develops to permit prospective memory: executive functions or representational access? According to the executive function models, prospective memory development should depend on the development of underlying executive abilities. In contrast, the models that include episodic processing components (i.e., the multiprocess framework and episodic processing model) would assume that prospective memory development is largely the function of processes underlying episodic memory, namely, representational access. However, a continuum
approach (Graf & Uttl, 2001) may be most informative in generating hypotheses regarding the contexts in which different processes and resource allocations are observed.

**Prospective Memory Development**

Somerville, Wellman, and Cultice (1983) were the first to empirically document event-based prospective memory in young children. In a naturalistic setting, a child’s caretaker instructed the child to remind him or her of a task to be completed at a specified time in the future. This task required children as young as 2-years-old to later recognize the appropriate target event and then recall and perform the task. For example, children were told, “Remind me to buy milk when we go to the store tomorrow.” Somerville and colleagues varied these tasks in two ways: (1) the interest to the child (high versus low interest, as determined by the caretaker), and (2) the delay between the instruction (e.g., “Remind me to buy milk”) and the cue to remind (short delays were less than five minutes, long delays were morning to afternoon or evening to next morning). Somerville et al. found that 2-year-olds failed to spontaneously remind their caretakers of low-interest events after long delays, but succeeded 80% of the time when the event was of high interest and after only short delays. While age differences in the overall performance of 2-, 3-, and 4-year olds were nonsignificant, the patterns revealed that older children were more likely than younger children to engage in spontaneous and prompted reminding in the low-interest, long-delay condition (see Table 1).

Sommerville and colleagues noted it was unclear, however, whether these differences (although nonsignificant) were due to 2-year-olds’ inability to recognize cues to remind or their inability to recall the information specified for reminding. The differences in performance on low-interest tasks as a function of age might also reflect
age differences in representational abilities; as Somerville and colleagues described, the social nature of the task not only requires that children understand another person does not remember something and needs reminding (a failure to reason about others’ beliefs as differing from one’s own), but the low-interest task also requires that children understand another’s desires as perhaps differing from their own. These seminal findings provoked further research on prospective memory in early childhood.

Contrary to Somerville et al., some of this research has revealed substantial development of prospective memory abilities. In efforts to investigate the development of prospective memory and its relationship to retrospective memory in a relatively controlled setting, Guajardo and Best (2000) compared performance across 3- and 5-year-olds on both laboratory computer-based prospective memory tasks and naturalistic prospective memory tasks. In the computer-based task, 3- to 5-year-old children were presented with random pictures for five seconds with a one second inter-stimulus interval (ISI) and were instructed to remember as many pictures as they could for later recall. To measure prospective memory, the researchers also told the children to press the space bar every time they saw a target picture (e.g., turtle). Half of these children completed the task with a picture of the target object (a turtle, for instance) taped to the computer screen, intended to serve as an external cue. Children’s retrospective memory was also measured as the number of items they were able to later recall. The naturalistic tasks involved a short-delay task of 20 minutes (children remembering to ask for a sticker and close the door at the completion of the computer task) and a long-delay task of 24 or 72 hours (children remembering to return the picture they received at the end of the first session and to ask for a pencil to take home).
Results of the computer-based prospective memory task revealed developmental differences in prospective memory performance; 5-year-olds remembered to press the key more often than 3-year-olds when the target appeared on the screen. Moreover, 5-year-olds were more likely than 3-year-olds to perform the naturalistic prospective memory task for both short delays (20 to 30 minutes) and long delays (24 to 72 hours). Only 25% of the 3-year-olds remembered to close the door and 52% remembered to ask for the sticker, whereas the corresponding percentages were 75% and 83%, respectively, for the 5-year-olds. Three-year-olds also required more prompting to complete tasks, on average, compared to 5-year-olds, providing further evidence that performance improved with age.

There was also more variation across tasks for 3-year-olds than 5-year-olds. Three-year-olds were almost twice as likely to remember to get a sticker as they were to remember to close the door, return a picture, or ask for a picture to take home (see Table 2, below).

Not only did the 5-year-olds demonstrate better prospective memory performance, but their retrospective memory performance was also significantly better; 5-year-olds recalled more pictures than 3-year-olds, and when asked, “What did I (the experimenter) ask you to remember to do when you saw the turtle?”, 81% to 88% of the 5-year-olds recalled the task, but only 52% to 69% of 3-year-olds remembered. However, separate correlations for each age group revealed that the 3-year-olds' retrospective (free recall) and prospective memory performances were significantly correlated, $r = .49, p < .001$, but the 5-year-olds' were not, $r = .18, p > .05$. 

15
Supporting the patterns observed by Somerville et al. (1983), these findings suggest significant improvements in prospective memory abilities with age, possibly as a function of improvements in retrospective memory. While these results indicate that the underpinnings of prospective memory are available to 3-year-olds (see also Wang, Kliegel, Liu, & Yang, 2008, for converging evidence), these findings do not make clear what changes take place to account for the better performance of 5-year-olds.

Guajardo and Best (2000) proposed several plausible explanations for the better performance of older children. First, 5-year-olds were more likely than 3-year-olds to select a strategic approach for remembering to press the key when the target picture appeared in the computer task. For instance, when asked if they did anything to help them remember the task, older children stated they looked at the picture of the item (taped onto the computer). Although the presence of an external cue did not improve children’s performance on the computer task, their selection of a strategic response (e.g., looking at the cue) was related to their performance on the prospective memory task. Thus, the success of 5-year-olds may be related to their better understanding of strategies and a metacognitive understanding of their own limitations, relative to 3-year-olds. This interpretation is consistent with data showing that correlations between strategies and retrospective memory performance increase steadily as a function of age (Schneider & Weinert, 1995).

A second proposal was consistent with executive functioning models of prospective memory, suggesting that differences in working memory development might have accounted for differences between 3- and 5-year-olds’ prospective memory performance. The relationship between working memory and prospective memory performance has
been examined in the adult literature (e.g., Marsh & Hicks, 1998) and suggests that central executive processing is critical for effective completion of prospective tasks. Guajardo and Best (2000) suggested that the observed age differences in prospective memory performance in 3- to 5-year-olds may be due, in part, to developmental differences in the central executive; 5-year-olds processing capacity (see Case, 1995) and increased processing speed (Kail, 1991; Hitch & Towse, 1995; Miller & Vernon, 1996) perhaps enabled them to better plan and monitor their performance.

Perhaps the most interesting finding from Guajardo and Best (2000) concerns the relationship between retrospective and prospective memory performance. Einstein and McDaniel (1990), Brandimonte and Passolunghi (1994), and Kidder, Park, Hertzog, and Morrell (1997) have failed to demonstrate a significant relationship between these two forms of memory in adults and children. The findings of Guajardo and Best seem to demonstrate that these two forms of memory are related in 3-year-olds, but not 5-year-olds. It is likely, however, that this relationship is mediated by central executive functioning.

Einstein and McDaniel (1990) have noted that these two forms of memory may be related when the retrospective aspect of the prospective task is challenging. Three-year-olds may have found this task more difficult than did 5-year-olds.

Also, because the retrospective memory task (remembering the objects) was the ongoing task, 3-year-olds may have been more susceptible to mutual interference between prospective memory and retrospective memory, resulting in a spurious correlation between performance on the retrospective and prospective memory tasks in
this age group (see Kvavilashvili & Fisher, 2007; McDaniel & Einstein, 2000; Smith, 2003, for converging evidence regarding mutual interference).

However, while 5-year-olds could retrospectively recall the instructions for the computer-based task, 3-year-olds seemed to rely solely on its implicit retrieval because even when they completed the task, one-third or more could not verbalize the instructions. This explanation might similarly account for 3-year-olds’ poorer performance on the door, picture, and pencil tasks relative to the sticker task. Although Guajardo and Best did not report 3-year-olds’ retrospective memory for the naturalistic task instructions, younger children may have been unmotivated to sufficiently encode these tasks.

Only recently has there been a study investigating the cognitive processes underlying event-based prospective memory in children (Smith, Bayen, & Martin, 2010). Using a multinomial processing tree, Smith and colleagues compared the prospective memory performance of 7- and 10-year-olds and adults. Adults performed better than 10-year-olds, who performed better than 7-year-olds, when asked to press the space bar when a previously viewed picture appeared (prospective memory task) during a color-matching task (ongoing task). Smith and colleagues sought to examine whether these age differences were due to the development of processes underlying the prospective or retrospective components of prospective memory. The prospective component, according to Einstein and McDaniel (1996), refers to remembering “that something must be done” while the retrospective component refers to remembering when it must be done and what the intended action is. Findings indicated that adults differed from children on the prospective component of the task (measured as costs to the ongoing task), meaning their
ability to remember that they must press a button when the picture appears. However, the age-related difference in prospective memory performance between 7- and 10-years-olds was not due to differences in the prospective component but to differences in the ability to correctly discriminate targets (previously viewed pictures) from foils (pictures not previously reviewed), that is, the retrospective component. Smith et al. were able to draw these conclusions because the formal model allowed them to control for differences related to the prospective component (i.e., ongoing task costs) and isolate changes in the ability of 7- and 10-year-olds children to remember when something must be done. These findings suggest that younger children relied more on familiarity processes and less on resource-demanding processes. If future investigations can better distinguish when prospective memory relies on familiarity and when it relies on resource-demanding processes, then we are one step closer to understanding age-appropriate prospective memory tasks.

In summary, the research I have described so far suggests that developmental differences may be observed for some instances of prospective memory, but not others, and provides support for the multiprocess framework. Children as young as 3-years-old may perform comparably to older children on tasks of high but not low interest and on those tasks that rely on familiarity more than vigilant monitoring. Based on these findings, considering the multiprocess framework in light of Graf and Uttl’s conceptualization of a prospective memory continuum, one would hypothesize that because low-interest tasks may not be as strongly encoded or represented in memory, success on these tasks may be more effortful, requiring reliance on executive functions that develop with age, as well as more effortful encoding and retrieval strategies. In
contrast, high-interest tasks are likely to be encoded deeply and are less reliant on age-related executive functioning.

Researchers have been preoccupied with determining where prospective memory tasks should be placed in the realm of memory: How is it different from encoding, rehearsal, and retrieval? When is it different? To some degree, as a result, the uniquely human ability to represent intentions and place them into different temporal localities has been ignored as a potentially key component of prospective memory.

**Representational abilities**

I propose that, like episodic memory, prospective memory requires the ability to represent future and past intentions and understand that these intentions can differ across time. Psychologists have adopted the term “theory of mind” to describe the ability to reason about mental states of the self and others as differing across time and individuals (e.g., Baron-Cohen, 1995). Thus, children who lack a theory of mind likely demonstrate prospective memory deficits. Although research on event-based prospective memory development demonstrates that its trajectory resembles that of theory of mind (i.e., a substantial increase in abilities observed between the third and fifth years; Langer, 1986), little has been done to explicitly investigate this connection. However, it has been noted that the emergent capacity for symbolic representation, followed by the reflective coordination of symbolic coordination, between the third and fifth years yields success in a variety of domains. At this age, children first succeed at pantomime tasks (e.g., Dick, Overton, & Kovacs, 2005), rule-use tasks like the dimensional change card sort task (DCCS; Zelazo, Frye, & Rapus, 1996), and appearance-reality and false-belief tasks (Flavell, 1986; Gopnik & Astington, 1988; Taylor & Flavell, 1984; Wimmer & Perner,
the possibility of a unifying underlying mechanism, referred to as “representational ability” hereafter.

Evidence for the relationship between representational ability and prospective memory is circumstantial. Research has demonstrated that high-functioning children diagnosed with Autism Spectral Disorder (HF-ASD) often experience impairments when creating and activating delayed intentions, (i.e., prospective memory, see Mackinlay, Charman, & Karmiloff-Smith, 2006). This observation, coupled with evidence of autistic children’s impairments on tasks involving theory of mind (Baron-Cohen, 2005) and self-reference (Lombardo, Barnes, Wheelwright, & Baron-Cohen, 2007) supports the link between prospective memory and representational abilities. Various aspects of children’s future-oriented behavior, including anticipating future states/events, delaying gratification, planning, and engaging in acts of prospective remembering, develop coherently and substantially during the preschool years, supporting the notion that children undergo a change in their underlying temporal representational abilities (Atance & Jackson, 2009; for converging evidence, see Moore, Barresi, & Thompson, 1998).

Study 1 examines the question; what is the role of representational ability in early prospective memory development? I hypothesized that representational ability is necessary for reasoning about intentions of the self and others across time, and will predict developmental and individual differences in performance on prospective memory tasks that require encoding and retrieval. Representational deficits likely account for differences in 2-year-olds’ poorer performance on tasks of low-interest relative to tasks of high-interest in Somerville et al.’s study (1983). Specifically, when tasks are of low-
interest and young children’s executive functioning is insufficient for monitoring, those who do not possess representational abilities required for encoding and retrieval processes will fail to remember to carry out a task that requires representation of a noncurrent intentional stance.

**Motivational Factors**

Goal-based models of motivation define goals as mental representations of desired and undesired future states (see Penningroth & Scott, 2008). Penningroth and Scott (2008) have proposed that prospective memory representations should be added to existing goal-based cognitive network models, such as goal-systems theory (e.g., Kruglanski, 1996; Kruglanski et al., 2002). In their view, superordinate goals and motivations (e.g., “go on a fieldtrip with my class”) serve as knowledge structures with associative links to prospective memories (e.g., “get parent to sign permission slip”). Individuals plan and implement activities, or means, to reach their goals, and their ability to carry out each of these subordinate activities, in effort to achieve the superordinate goal, relies, in part, on prospective memory.

Some recent promising research has supported this model, indicating that superordinate goals generate “top-down” activation of subordinate goals, influencing performance on prospective memory tasks in the real world. For instance, Penningroth and Scott (2007) reviewed evidence that individuals rate prospective memory tasks related to goals as more important than prospective memory tasks that are not goal-related (Penningroth, 2005a), that people use strategies more frequently to carry out tasks rated relatively important (Penningroth, 2005b), and that task importance increases the
accessibility of some prospective memory tasks during the delay, leading to their earlier retrieval (Penningroth, 2006).

Other studies suggest that the relative importance of the ongoing and prospective memory tasks predict performance on each (e.g., Ceci & Bronfenbrenner, 1985; Guajardo & Best (2000); Somerville et al., 1983; see also Penningroth & Scott, 2007). Kliegel, Martin, McDaniel, and Einstien, (2001, 2004) argued that the greater importance of a task improves performance by motivating participants to adopt strategies so that attention can be allocated elsewhere. In addition, I argue that motivation improves prospective memory insofar as it reduces reliance on limited cognitive resources. Consistent with this argument, Kliegel et al. (2001) reported that importance had an effect on time-based prospective memory performance, but not event-based prospective memory performance, presumably because the former requires more vigilant monitoring and is more taxing on cognitive resources. That is, Kliegel et al. assumed that tasks deemed important encourage more strategic monitoring, and, as such, the perceived importance of a task could have little influence over a task that relied little on monitoring, but that was automatically retrieved by the target event. Consistent with this interpretation, Kliegel et al. (2004) found that task importance did influence event-based prospective memory performance if the task was relatively monitoring-demanding.

As such, Study 1 also examines how children’s interest in the prospective memory task interacts with representational abilities and executive functioning to predict event-based prospective memory performance. I predicted that individual differences in executive functioning and representational ability would predict individual differences in children’s performance on event-based prospective memory tasks of low-interest.

23
CHAPTER TWO

STUDY 1: INDIVIDUAL DIFFERENCES IN REPRESENTATIONAL ABILITY AND THE PROSPECTIVE MEMORY OF THREE-YEAR-OLD CHILDREN

Following the approach used by Smith et al. (2010), Study 1 explores whether the development of prospective memory across the third year is the result of executive functioning and/or representational ability development. By asking 3-year-old children to complete tasks of high or low interest that require them to carry out the intention themselves or to remind the experimenter to complete the task, I examine how these factors affect performance in preschool children, aiming to better understand how these factors interact to influence where each task falls along the Graf and Uttl prospective memory continuum.

Because interesting tasks warrant more attention, I propose they require less effort at encoding to establish a strong association to a target event and are more likely to rely on automatic retrieval than monitoring processes. If high-interest tasks do not require preparatory attention, 3-year-olds’ performance on these tasks will not be related to executive functioning because remembering these tasks will not require effortful monitoring. However, when a high-interest task requires reminding the experimenter to carry out the task, representational abilities may be recruited to reason about the experimenter’s memory.
In contrast, I argue the nature of low-interest tasks makes them more susceptible to forgetting. As such, individual differences in children’s ability to remember *that* the low-interest task exists should be related to their ability to represent the intention to carry out the task. Their ability to remember *when* the task must be carried out should be related to the extent to which they are monitoring the environment for a cue. As such, I predict that individual differences in executive functioning and representational ability will mediate the relationship between age and prospective memory performance on tasks of low-interest. Representational abilities should be more influential in mediating performance when 3-year-olds are required to remind an adult to carry out the task of low interest because this task involves not only representing an intention, but reasoning about whether another person remembers the intention.

This study explores the unique impacts of representational abilities and executive functioning on individual differences in 3-year-olds’ event-based prospective memory. Because event-based prospective memory develops across the third year, this age group presents an opportunity for studying *what* develops in children to enable prospective memory. For these reasons, mediation analyses are invoked to explore the extent to which individual differences mediate age-related differences in prospective memory across the third year. In cases where a statistical relationship between age and prospective memory performance is not observed, age is controlled as an extraneous variable while the unique contributions of individual differences in executive functioning and representational ability are examined.
Method

Participants

Thirty-two 3-year-old children, recruited from preschools in the South Florida area (males = 12), participated following parental consent and verbal assent. One male was excluded as an outlier because of his age (59 months old), resulting in 31 preschoolers with a mean age = 43 months (ranging 35 to 53 months, SD = 5.4 months). Mean age for male participants was 43.08 months, range = 14 months, and the mean age for females was 43.1 months, range = 18 months. At the end of each of two sessions, participants chose a sticker and were thanked for their participation. Children were excluded from analyses on an item-by-item basis if they declined to participate in a task (Simon Says, n = 1; digit span, n = 1; backward trail span, n = 3; Stroop, n = 3) or because participants incorrectly answered knowledge questions used to confirm their understanding (dimensional change card sort, n = 1). On two occasions, the parent interrupted the session before the prospective memory task could be carried out (High-interest remember condition, n = 1; Low-interest “remind me” condition PM, n = 2). All participants were tested in the afternoon.

Design

Prospective memory tasks. The design was a 2 x 2 within-subjects factorial varying interest level (high or low) and level of metacognition (“remember” to retrieve item by ones’ self or “remind me” to retrieve item). Participants were given instructions for a high-interest “remember” and low-interest “remind me” prospective memory tasks in one session and instructions for a high-interest “remember” and low-interest “remind me” tasks in the other session. Sessions were held approximately one week apart.
Prospective memory tasks were counterbalanced across two sessions so that half of participants received the high-interest “remind me” / low-interest “remember” combination in the first session, while the other half received these instructions in the second session. This counterbalancing is presented in Table 3. Order of task instructions was also counterbalanced across participants, but for simplicity this is not depicted in the table below. For example, half of participants who received the high-interest “remind me” task with the low-interest “remember” task in session 1 would hear the instruction for the low-interest “remember” task followed by instruction for the high-interest “remind me” task at the beginning of the session, while the other half would receive the high-interest “remind me” instructions first followed immediately by the low-interest “remember” task instructions.

Executive functioning tasks. Executive functioning tasks were presented in a fixed order, designed to maintain the child’s interest across the session. Three tasks were administered in each session. Set A included the digit span, Simon Says, and N-back tasks, whereas Set B included the forward and backward trail making, day/night Stroop, and dimensional change card sort (DCCS) tasks. Presentation of each set was counterbalanced across sessions and prospective memory tasks as demonstrated in Table 3.

Belief/Desire tasks. Each child was asked to complete either a set of belief tasks or a set of desire tasks in each session. The presentation of these tasks was counterbalanced across session and combination of executive functioning and prospective memory tasks, depicted in Table 3. As a result, eight possible presentation order combinations were formed and children were randomly assigned to each group.
Materials and Procedure

Children were tested individually across two sessions in a small room on their preschool campus. In each session, the child was asked to sit at a small table across from the experimenter. An unbiased observer was present at all sessions to record children’s responses and was seated out of view of the child. Once the child’s verbal consent was obtained, the observer began timing the session and the experimenter provided the child with two prospective memory instructions, followed by three executive functioning tasks and then the belief or desire tasks.

Prospective memory tasks. To assess children’s ability to remember tasks of high-interest, participants were asked to retrieve a sticker at the end of the session. To assess children’s ability to remember tasks of low-interest, participants were asked to change a sign on the door of the experiment room.

In every session, a small blue polka-dot bag containing stickers was placed a few feet from the door. The bag was located behind the experimenter and in the background of the participants’ left visual field. A double-sided 8”x10” sign was hung on the door, also behind the experimenter and in the background of the participants’ right visual field. On one side, the sign was green and displayed a printed smiley face with the word “OK” printed in large black ink. On the other side, the sign was red and displayed a large hand with the palm facing forward, below the word “STOP.” Upon entering the room with the child, the experimenter referenced the sign stating, “Oh, I have to flip this sign over to the red side so that others will know we are using this room. That way, they know not to come in.” The door was then positioned so that it was partially ajar and the sign was still visible from where the child was seated.
Prospective memory instructions were presented first. To assess children’s ability to “remind me,” the experimenter turned to point to the appropriate object and said, “[Oh/Also], when we’re all done, will you remind me to [get a sticker for you out of that blue bag (high-interest) / change the sign on the door back to the green side (low-interest)]?” For these tasks, the object was placed out of reach of the child (i.e., the sticker bag was placed high on a filing cabinet and the sign was hung high on the door). This was done following piloting so that, if children remembered the task, they would be encouraged to follow through with reminding the experimenter, rather than carrying out the task on their own. For “remember” tasks, the experimenter said, “[Oh/Also], when we’re all done, will you remember to [get a sticker out of that blue bag (high-interest) / change the sign on the door back to the green side (low-interest)]? You don’t have to ask me, you can just do it yourself.” For these tasks, the object was placed at a height the child could reach (i.e., the sign was hung on the door handle and the stickers were set on a small chair). After giving each instruction, the experimenter checked children’s comprehension and encouraged them to process or plan the task by following up with, “Can you [help me] remember to do that?” waiting for a response, and then asking, “How will you remember to do that?”

Once all other tasks were completed, the experimenter thanked the child for his or her participation, and asked if he or she had any questions before saying, “Okay, we are all done!” (prospective memory cue 1). At this point, the timer was stopped and the experimenter began to rise from the table, pretending to be distracted by papers to allow the child the opportunity to carry out the prospective memory tasks. If, after a brief pause, the child did not carry out the task, the experimenter stated, “Okay, I’ll take you back to
your class now…” (prospective memory cue 2) and escorted the child toward the door. If the child failed to carry out either or both of the prospective memory tasks once passing through the doorway, the experimenter paused at the door after him or her, in between the sign and sticker bag and asked, “Did we forget anything?” while looking puzzled. If the child still failed to carry out a task, he or she was given a sticker for their participation, thanked, and walked back to class.

Dependent variables were children’s success at carrying out the task (pass/fail) and the number of cues needed before carrying out the task.

**Digit-span task** (Wechsler). The experimenter verbally presented digits at a rate of one per second, and the child was asked to repeat them verbatim. An example was demonstrated for each participant. Two trials were administered for each length, beginning with two digits. The task ended once children made errors on both trials for a given digit span. Longest digit span served as the performance measure of short-term memory.

**Forward trail making.** This task was a modified version of the paradigm described in Salthouse et al. (2000), used here to assess short-term memory. An 8”x10” card was presented illustrating nine familiar animals with monosyllabic names, placed pseudo-randomly, and printed in color (see Appendix A). Participants were asked to name each animal to ensure familiarity. Instructions indicated that the experimenter would point to a series of animals and then the child should point to the same animals in the same order. After an example and a practice round, the experimenter began by saying aloud the names of two animals as she pointed to them. Similar to the digit-span task, two
trials were conducted for each trail length, and the task ended after the child erred on both trials of a given length. Longest trail span served as the performance measure.

**Backward trail making.** Using the same card as in forward trail making, the child was instructed that the game was going to change and the child was to point to animals in the opposite order as the experimenter. After an example and a practice round, the experimenter began saying aloud the names of two animals as she pointed to them. The child was then told, “Now it’s your turn. Remember to go backward.” The task continued in this manner until the child erred on both trials for a given length. Longest backward trail span served as the performance measure of working memory.

“**Simon Says**” (Strommen, 1973). Children were told they were to play a game like Simon Says. Children were asked to point to different body parts, then the experimenter introduced them to two finger puppet mice, one named “Simon” who was “a very friendly mouse,” and another named “Oscar” who was “a mean mouse.” Children were instructed that whenever Simon told them to do something, they should do it, but when Oscar told them to do something they should “just ignore him - don’t do it.” After practicing following one order from Simon and ignoring one from Oscar, 18 test trials were administered, six of which required children to inhibit their response to Oscar’s instructions. If children made any move to follow Oscar’s instruction, they were given a score of zero for that trial and told, “Remember, don’t do what Oscar says.”

The number of trials correctly ignored served as the performance measure (out of six). To confirm that children understood the rule, after the task, each was asked, “When Oscar tells you to do something, should you do it?” followed by, “When Simon tells you to do something, should you do it?” All children correctly recalled the rules. Only one
child refused to participate in this task and was eliminated from further analysis involving this measure.

**Day/Night Stroop** (Gerstadt, Hong, & Diamond, 1994). Children were presented with two 8”x 10” cards, each with 20 pictures of suns and moons. A baseline was established by asking children to say, “day” when the experimenter pointed to the sun and to say, “night” when the experimenter pointed to the moon. The experimenter pointed to each image, moving from left to right, top to bottom. Next, children were shown the second card and told, “Now this will be a little bit tricky, because I would like you to do the opposite. When you see the Sun I want you to say, ‘night’ and when you see the moon I want you to say, ‘day.’” The experimenter then continued in pointing to each image in the same manner as for card 1. Children’s errors were corrected for both cards. The differences in errors for card 1 and card 2 served as the performance measures of inhibition.

**N-back book.** The n-back task was adapted from Mantyla et al., 2007, who used it to assess updating, attention, and working memory. In standard n-back tasks (Gevins & Cutillo, 1993), participants are asked to monitor the identity or location of a series of stimuli and to indicate when the currently presented stimulus is the same as the one presented n trials previously. A 38-page book was constructed using 8”x 10” colored construction paper. One of eight different shapes (a triangle, rectangle, star, square, diamond, arrow, circle, and house; all approximately 3”x 3”) was alternately presented on each page. There were eight instances where a shape was repeated, for example, where two pages in a row presented a square. The number of “filler” pages between repeating-shape pages (those presenting alternating shapes) varied from one to three. Participants
were instructed they would see each shape in the book, one at a time, and were to tap the
book page if two shapes were repeated. The experimenter then demonstrated using the
first four pages, where a triangle was presented on both the third and fourth pages, by
tapping the fourth page. If children failed after a practice trial, the experimenter corrected
them and reviewed the rules before continuing. Pages were turned at an even pace (2–3 s
per item). Scores reflected the total number of correctly identified repeat shapes minus
the sum of the number of missed targets (omissions) and false responses (commissions),
ranging from zero to eight.

**Dimensional change card sort** (DCCS; adapted from Frye et al., 1995). Children
were instructed to sort cards that varied on two dimensions, shape and color. The task
consisted of two phases, a pre-switch phase and a post-switch phase. During the pre-
switch phase, children were shown two standard cards that differed on both dimensions
(one red triangle and one green circle). The experimenter labeled each card in terms of
both dimensions (e.g., “Here is a green circle”). The first target dimension (e.g., shape)
was then highlighted and children were given a pair of rules for sorting, “We are going to
play a game. This is the shape game. The shape game is different from the color game. In
the shape game, all the circles go in this box, and all the triangles go in that box. We
don’t put any circles in that box. No way. We put all the triangles over here and only
circles go over there. If it is a circle, then it goes here. If it is a triangle, then it goes there.
This is the shape game.” The experimenter then sorted two cards (e.g., a red circle and a
green triangle) to demonstrate what the child was required to do. A series of five pre-
switch cards was presented for sorting, one at a time in the following order: green circle,
red triangle, green triangle, red circle, green circle. On each trial, the experimenter stated
the relevant rules explicitly (e.g., “If it is a circle, then put it here, but if it is a triangle, put it there”), labeled the card in terms of both relevant dimensions (e.g., “Here is a red circle”), and asked the participant: “Where does this go?” Children were given feedback regarding the correctness of their responses: When children placed a card into one of the two trays, the experimenter said, “Yes, that’s right” or “No, that’s not correct—remember the rules” and continued to the next trial. After children correctly sorted five consecutive cards according to one dimension, the experimenter continued to the next phase of the card sort.

In the post-switch phase, the experimenter administered five post-switch test trials, requiring children to sort a total of five test cards according to the alternate dimension (e.g., color). Prior to the post-switch phase, children were told, “Now we are going to switch. We are not going to play the shape game any more. We are going to play the color game. When there is a red card, you have to put it in this box, but whenever there’s a green card, then it goes in that box. We don’t put red cards in that box. No way. We put all green cards over here and only when there’s a red card does it go over there. If there’s a red card, then it goes here. If there is a green card, then it goes there. This is the color game.” As in the pre-switch test trials, the relevant rules were stated on each trial and cards were labeled as both relevant dimensions. Cards were presented in the following order: red circle, green circle, red triangle, green triangle, red circle. Children were not corrected on these trials; when children placed a card into one of the two trays, the experimenter simply stated, “Okay let’s do another,” and continued to the next card. The dependent variable was the number of correct post-switch sorts for cards 1, 4, and 5 (range: 0-3). These trials require the child to negotiate conflicting dimensions while trials
2 and 3 do not discriminate among children who are following the new rule from those who are not; dimensions are not conflicting on these trials, but can be sorted correctly following either rule. Children’s knowledge of post-switch rules was assessed at the end. Only one child failed to correctly recall the rules and her scores were omitted from further analyses involving this task.

**Belief tasks** (adapted from Gopnik & Astington, 1988). Belief tasks involved a representational change and false-belief task. Children were presented with a crayon box filled with candles rather than crayons. Each child was shown the box and asked, “What do you think is in this box?” After revealing the candles, the researcher assessed the child’s knowledge of his or her own belief by asking: “What did you think was in this box before?” A score of 1 was given if the child answered, “candles.” After the candles were replaced and the box closed, the researcher assessed the child’s understanding of another’s false beliefs: “[name of a peer, e.g., Sarah] hasn’t seen what is in this box. What do you think she will think is inside this box when she first sees it?” A score of 1 was given if the child answered, “candles.”

**Desire tasks.** Children’s understanding of their own desires was measured using a task adapted from Moore, Baressi, and Thompson (1998, modeled after Gopnik & Slaughter, 1991). Children were shown an orange and yellow box and asked, “Which container would you like to open - the orange one or the yellow one?” After selecting a container, the child was allowed to briefly handle the object inside (a toy cow or a toy pig). The toy was then replaced inside the container and lid returned, and the child was asked, “Now which box do you wish to open?” All children elected to open the box not previously opened. Before the child could open the second box, he or she was asked,
“When I first asked you, before we opened any box, which one did you want to open then - the orange one or the yellow one?” A response was considered correct if the child named the container that was first opened.

The child was then allowed to open the remaining box and handle the toy inside. After the toy and lid were replaced, the child was told, “[name of a different peer, e.g., Justin] was here earlier. He only got to open the orange box [for example]. What box do you think Justin will want to open when he comes back - the orange one or the yellow one?” A response was considered correct if the child named the other box (in this case, yellow). A score of 1 was given for each correct response on the desire tasks.

Sessions were videotaped and later coded by two separate raters, blind to hypotheses. Inter-rater agreement was 99% and differences were resolved by discussion.

Results

All effects are reported as significant at $p < .05$, unless otherwise indicated. Effect sizes are reported as partial eta squared. Means are reported as estimated marginal, and root mean square errors are reported for each test ($RMSE$).

Preliminary analyses

Presentation order was not related to the total number of prospective memory tasks passed, performance on belief/desire reasoning tasks, or any measures of executive functioning. Session time was not related to whether a participant passed individual prospective memory tasks. Age and gender effects are described as they pertain to each of these variables in the respective sections, below.

Prospective Memory
Overall prospective memory performance was calculated as the total number of prospective memory tasks a participant completed across the two sessions (total possible = 4). Children completed over half of the four prospective memory tasks, on average ($M = 2.45$, $SD = 0.18$, $n = 31$).

The percent of participants passing individual prospective memory tasks is presented in Figure 1. The majority of participants remembered to get a sticker for themselves (i.e., high-interest “remember” task, $n = 26$, 89.6% remembered), but forgot to remind the experimenter to change the sign on the door (low-interest “remind me” task, $n = 10$, 33.3% remembered). Cochran’s test confirmed the difference was significant, $\chi^2(3) = 32.83$.

Chi-square tests revealed that presentation order was significantly related to performance on the high-interest “remember” task, $\chi^2(7) = 21.67$. A closer examination revealed that the only three children who failed to remember the high-interest “remember” task were in presentation order four and received this task in the second session. These children were relatively young, falling below the median age (41, 35, and 40 months). Given that other children who received the same task in the first and second sessions passed it, this effect may be confounded by the age of participants in this group and can be attributed to possible fatigue of these younger participants. However, session time was not statistically related to whether a participant passed or failed a prospective memory task; high-interest “remember,” $r_{pb} = .13$; high-interest “remind me,” $r_{pb} = -.05$; low-interest “remember,” $r_{pb} = -.05$; low-interest “remind me,” $r_{pb} = -.12$, ns.

**Age and gender effects.** The correlation between age and overall prospective memory performance, $r = .27$, was not statistically significant, possibly due to the
limited age range in the small sample. Participants who passed the low-interest “remember” task were slightly older ($M = 45.1$ mo., $SD = 5.9$, $n = 14$) than participants who failed ($M = 41.4$ mo., $SD = 4.4$, $n = 17$), but this relationship was only marginally significant, $r_{pb} = .35$, $p = .05$. Age was not related to performance on the other three prospective memory tasks.

Females remembered to carry out more prospective memory tasks than males, even after controlling for age (estimated marginal means: $M_{girls} = 2.84$, $SE = 0.24$, CI[2.35,3.40]; $M_{boys} = 1.92$, $SE = .31$, CI[1.29, 2.54]), $F(1, 30) = 6.28$, $RMSE = 1.06$, $\eta^2_p = .17$. As depicted in Figure 2, girls passed each of the prospective memory tasks at a higher rate than boys. Contrasts between boys and girls were performed separately for each condition. Due to small sample sizes, 2-sided Fisher’s exact tests are reported. Girls were slightly more likely than boys to pass “remember” tasks, $\chi^2(2) = 5.63$, $p = .09$. Boys and girls did not differ significantly in the number of high-interest, low-interest, or “remind” tasks passed. When each task was examined separately, the difference in the number of prospective memory tasks passed between girls and boys on the high-interest remind task approached significance, $\chi^2(1) = 4.28$, $p = .06$. Contrasts failed to reach significance, however, for the high-interest remember, low-interest remind, and low-interest remind tasks, all $\chi^2(1) \leq 3.21$, $ns$.

**Prospective Memory Cues**

The number of cues a participant required to pass each task was used to gauge his or her recognition of prospective memory target events. Recall that participants received an initial cue, “Okay, we’re all done!” followed by a second cue, “Okay I’ll take you back to your class now!” if they failed to complete the task after a brief pause.
Participants were assigned a score from one to three to indicate the number of cues they received prior to passing a task. Participants were assigned a score of “4” if they did not recall the prospective memory task after the third cue. Figure 3 presents the mean number of cues participants received per prospective memory task.

These scores were subjected to a 2 (interest level: high vs. low) x 2 (instruction: remember vs. remind) x 2 (gender) mixed ANCOVA with age as the covariate, interest and instruction as within-subjects variables, and gender as a between-subjects variable.

There was a significant main effect of interest, $F(1, 27) = 5.41$, $RMSE = .90$, $\eta^2_p = .17$. Children required fewer reminders to pass tasks of high-interest ($M = 2.21$, $SE = .17$, CI[1.87, 2.55]) compared to tasks of low-interest ($M = 3.35$, $SE = .13$, CI[3.08, 3.61]). There was also a significant main effect of instruction, $F(1, 27) = 4.59$, $\eta^2_p = .15$. Children required fewer cues to remember tasks on their own ($M = 2.66$, $SE = .14$, CI[2.37, 2.94]) than they did to remind the experimenter to carry out the task ($M = 2.90$, $SE = .16$, CI[2.59, 3.20]). This analysis also produced a significant interaction between the covariate age and instruction, $F(1, 27) = 5.59$, $\eta^2_p = .17$. This interaction is presented in Figure 4. A closer examination of the parameter estimates revealed that older participants received fewer reminders in the low-interest “remember” task compared to younger participants, $b = -.10$, $t(30) = 5.10$. No other main effects or interactions were observed.

**Executive Functioning**

Table 4 presents means and standard deviations (on the diagonal) for each measure of executive functioning and correlation coefficients (above the diagonal). Significant correlations are flagged. Simon Says, 1-back, backward trail span, and dimensional change card sort (DCCS) tasks were distributed non-normally. Simon Says
and DCCS tasks were bimodally distributed, while the backward trail span and 1-back was positively skewed (participants tended to perform poorly on these tasks), even after log correction. Review of the literature revealed that preschoolers’ performance on the DCCS and Simon Says tasks are frequently bimodal (e.g., Carlson, 2005; Frye et al., 1995; Waxman & Namy, 1997). The 1-back task is a modified version of the n-back task used to assess working memory in older participants and so no comparable data exist. The same is true for the backward trail span task, another working memory task modified from tasks used for older participants. Carlson (2005) reduced these variables to binary scores (i.e., pass/fail) for the purposes of analysis, but cited this as a potential limitation because of the loss of sensitivity. For this reason, and due to the small sample size here, these variables were not reduced to binomial scores, but treated non-parametrically as ordinal ranked data; Kendall’s tau-b is appropriate for examining correlations involving ranked data, especially with many ties, and is reported for these relationships (in bold), and Pearson’s r is reported for all others. Age correlations are presented in the bottom row.

Examination of Table 4 reveals a strong positive correlation between performance on Stroop and Simon Says tasks, known to measure inhibitory capacity. Digit-span and Forward trail span, two tasks measuring short-term memory, were also positively related. Notably, performance on the 1-back was related to digit-span performance as well.

Older children performed better on the forward trail span and 1-back tasks, and age was marginally related to backward trail span performance. It is important to note that it remains possible that age effects might be significant if younger and/or older children were also tested, as was done in some of the original studies using these tasks.
(e.g., the Day/Night task; Gerstadt et al., 1994). For example, when Carlson (2005) conducted post hoc analyses (Tukey’s Honestly Significant Difference) on various executive functioning tasks, she failed to find significant difference between young 3- (36–41 months) and older 3-year-olds (42–47 months) but reported significant improvement from young and older 3-year-olds to young 4-year-olds, and again from young 4- to older 4- and 5-year-olds.

Boys and girls did not differ significantly on any measures of executive functioning.

**Executive functioning and prospective memory.** The relationship between prospective memory and executive functioning performance is summarized in Table 5. This relationship was examined for overall prospective memory performance first; the total number of prospective memory trials a participant passed was positively related to performance on the forward trail making task, the DCCS, Simon Says, and the 1-back task.

Next, the number of cues (1 through 3) that participants required to pass a particular task was correlated with individual measures of executive functioning. Participants were assigned a score of “4” if they failed to complete the task after the third explicit cue. As such, negative correlations between executive functioning and number of cues would indicate that participants with better executive functioning required fewer cues to complete the task. The number of cues needed to pass the high-interest “remember” task was not significantly related to any measure of executive functioning. The number of cues required to complete the high-interest “remind me” task was negatively related to the 1-back and digit-span tasks. The number of cues required to pass
the low-interest “remember” task was also negatively related to performance on the 1-back, as well as the forward trail span, digit-span, and Simon Says tasks. Finally, the number of cues required to carry out the most difficult, low-interest “remind me” task was strongly related to performance on the DCCS, but only approached significance in the relationship to Simon Says. Those participants who required fewer cues to carry out the intention performed better on tasks of executive functioning. These findings suggest that monitoring was related to performance on these tasks.

Belief/Desire Reasoning

Children passed a mean of 1.13 out of the four belief/desire reasoning tasks \((n = 31, \text{SD} = 1.09)\). Boys and girls did not differ in the total number of tasks passed. Age was marginally related to the sum of tasks passed (out of four), \(r = .35, p = .05\).

The percentage of children passing each of the belief-self, belief-other, desire-self, and desire-other tasks is presented in Figure 5. Cochran’s test failed to indicate a significant difference in the proportion of participants passing each task, \(\chi^2(3) = 3.46, ns\).

Belief/desire reasoning and prospective memory success. Those children who succeeded in reasoning about the beliefs and desires of themselves and others were more likely to remember to carry out a delayed intention; there was a significant, strong positive correlation between the number of belief/desire tasks that children passed and the number of prospective memory tasks passed, \(r = .42\). This relationship remained marginally significant after partialling out age, \(r = .37 (p = .05)\), suggesting that individual differences in children’s prospective memory performance was uniquely related to individual differences in belief/desire reasoning.
The number of cues a participant received for each prospective memory task was correlated with the sum of belief/desire tasks passed. Participants who passed more belief/desire reasoning tasks required fewer cues to complete the low-interest “remind” task, $r = -.44$. This relationship remained after partialling out age ($r = -.44$). No other correlations were significant. These findings suggest that participants with better theory of mind were better able to recognize the target event as a cue to remind the experimenter to change the sign on the door.

**Individual Differences in Prospective Memory Performance**

**Summary of age and gender effects.** With a sample of 31 three-year-old children, age was not significantly related to the total number of prospective memory tasks participants passed or to performance on the “interest” factor. Age was related to performance on the “instruction” factor of prospective memory; older participants received fewer reminders in the low-interest “remember” task compared to younger participants. Age was also positively related to performance on the 1-back, forward trail span, and backward trail span executive functioning tasks. Finally, age was positively, albeit marginally, related to the number of belief/desire reasoning tasks a participant performed.

Gender was related the total number of prospective memory tasks a participant passed, independent of age; girls remembered to carry out more tasks than boys. However, these contrasts did not reach statistical significance when compared for each individual prospective memory task or the number of cues participants’ received on any prospective memory factor. Boys and girls did not differ on measures of executive functioning or belief/desire reasoning.
**Factor Analysis.** The next step in exploring how individual differences in executive functioning and belief/desire reasoning related to individual differences in children’s prospective memory performance was to factor analyze individual measures of executive functioning and overall performance on belief/desire reasoning tasks. When all measures of executive functioning, described above, and the sum of belief/desire reasoning tasks were input into a principal components analysis, the resulting Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .46, below the recommended value of .6. In addition, the Bartlett’s test of sphericity was not significant, $\chi^2(28) = 36.12$. This issue was explored further using multiple regression analyses, revealing that multicollinearity existed between the digit span, forward trail span, and backward trail span tasks. Because multicollinearity increases the standard error of factor loadings, making them less reliable and difficult to interpret, particularly when sample sizes are small, the digit span and backward trail span tasks were eliminated from further factor analyses, and only forward trail span scores were included. The forward trail span task was chosen because participants seemed to have the best understanding of this task yet it produced the greatest variability in scores, compared to the digit span and backward trail span tasks, and of the three, this task demonstrated the consistently strongest relationship to prospective memory performance.

These methods yielded six variables, of which the factorability was examined. The KMO adequacy was .65, and Bartlett’s test of sphericity approached significance, $\chi^2(15) = 22.30, p = .10$. Additionally, the communalities were all above .3, further suggesting that each item shared some common variance with other items. Given these
overall indicators, factor analysis with varimax rotation was conducted on the six items. Principal components analysis was used because the primary purpose was to identify and compute composite scores for the factors underlying executive functioning and representational ability.

The Simon Says, Stroop, 1-back, and forward trail span loaded heavily on the first factor (Eigen value = 1.96), while the DCCS and overall belief/desire reasoning scores loaded heavily on the second factor (Eigen value = 1.60). The two factors explained 59.23% of the variance. All factors had a primary loading above .5 and only one factor had a cross-loading above .3 (1-back), however this factor had a strong primary loading of .62. When the internal consistency of these measures was examined using Cronbach’s alpha, it was revealed that the moderate alpha of the executive function factor (.56) could be improved (to .64) by elimination of the 1-back task. Subsequent principal components analysis revealed that removal of this variable improved the overall factorability of the remaining five (KMO = .647; Bartlett’s test of sphericity, $\chi^2 (10) = 21.281, p < .05$) and raised the subject-to-variable ratio above the recommended .5 (STV = 6.2). The resulting final structure is presented in Table 6 and accounted for 67.44% of the variance. The factor including the Simon Says, Stroop, and forward trail span tasks was labeled “executive functioning” because these tasks have been shown to reliably measure inhibition and short-term memory. The second factor, comprised of belief/desire reasoning and the DCCS was labeled “representational ability.” Kloo, Perner, and Giritzer (2010) also found these tasks were related across the third year of development and put forth the "redescription" hypothesis: By age 4, children understand that one and
the same thing can be described differently under different perspectives, a cornerstone for object-based set-shifting, as is required by the DCCS task. Similarly, Muller, Zelazo, and Imrisek (2005) found that performance on the DCCS predicted false-belief reasoning, and proposed that understanding false belief reflects a general understanding of representation, prepositional negation, and the ability to use higher-order rules. Dick, Overton, and Kovacs (2005) added that this relationship may reflect children’s developing reflective competence in coordinating symbolic representation, that is, the ability to reflect on conscious thought.

Standardized scores on the five variables were weighted (multiplied by their primary factor loadings). The composite scores for the two factors were then created from the mean of the weighted standardized scores on items that had their primary loading on each factor.

**Regression analyses.** To examine the unique contributions of representational ability and executive functioning to prospective memory capacities, individual differences in overall prospective memory performance were examined in a regression model. Although age was not significantly related to overall prospective memory performance, any shared variance was eliminated in the first step, followed by gender. Executive functioning was forced into the second step, followed by representational ability in the third and final step. Examination of Table 7 reveals that the inclusion of executive functioning and representational ability in a model of prospective memory partially mediated the effects of gender; when these individual differences were added, the effect of gender diminished to marginal significance.
Difference scores in the number of cues received (one through four) for tasks of each type were calculated next for each factor. For instance, “interest” scores were calculated as the mean number of cues received for high-interest tasks minus the mean number of cues received for low-interest tasks. “Instruction” cues were calculated as the mean number of cues received for “remember” tasks minus the mean number of cues received for “remind” tasks.

Because age was not related to the number of cues received in “interest” conditions, the variable was regressed out during the first step of the hierarchical regression analysis, followed by gender. Executive functioning was entered in the third step, and finally, representational ability in the fourth and final step. This analysis is presented in Table 8. Individual differences in representational ability uniquely accounted for differences in the number of cues participants received for low-interest prospective memory tasks. Those participants who were able to remember low-interest tasks as easily as they remembered high-interest tasks had better representational ability than those who required more cues for low-interest relative to high-interest tasks. This relationship was independent of age, gender, and executive functioning.

Because age interacted with the number of cues received in the “instruction” conditions, difference scores were regressed onto gender in the first step, followed by executive functioning in the second step, and representational ability in the third step. Any remaining unique contributions of age were examined in the fourth and final step. Examination of Table 9 reveals that individual differences in gender, representational ability, and executive functioning failed to mediate developmental differences in the number of cues a participant received on “remember” versus “remind” tasks. The
negative relationship between age and the relative number of cues needed for the
“remind” tasks relative to “remember” tasks indicates that older participants required
more cues to complete the “remind” task compared to the “remember” task, whereas
there was no difference for younger participants.

Although the ANCOVA of prospective memory cues did not reveal a significant
interaction between instruction and interest once age was controlled for, trends in
performance suggested that children found the high-interest “remember” task to be
easiest, followed by the high-interest “remind,” low-interest “remember,” and finally, the
low-interest “remind” task (see figures 1 and 3). To allow for a within-subjects
examination of individual differences in the difficulty with which children passed tasks,
the number of cues a child received during the high-interest “remember” task served as
the baseline score from which the number of cues for the other three tasks were
subtracted. These three difference scores were then entered into hierarchical regression
analyses, as above, to examine the unique effects of age, gender, executive functioning,
and representational ability. These results are presented in Table 10, and indicated that
individual differences in children’s scores on the representational ability factor positively
predicted differences in the number of reminders needed to pass the low-interest
“remind” task, relative to the number of reminders needed to pass the high-interest
“remember” task. Performance on the low-interest “remind” task was more similar to
performance on the high-interest “remember” task for 3-year-olds with better
representational ability; they required relatively fewer reminders to carry out the task than
did three-year-olds with poorer representational ability, after controlling for age, gender,
and executive functioning. While this effect was significant, the overall regression model was not, however. No other effects or models were significant.

**Discussion**

The findings of the present study indicate that children as young as 3 years of age are able to use prospective memory in naturalistic settings. Consistent with previous findings (Guajardo & Best, 2000; Sommerville et al., 1983), not only did 3-year-olds perform familiar prospective memory tasks, such as asking for something they like, but some also remembered to carry out a task that was not particularly interesting, and even remembered to remind an adult to do so. However, these tasks were not carried out with equivalent ease. Like 3-year-olds in Guajardo and Best’s (2000) study, children's incentive to complete the to-be-remembered task played a role in their performance. Children were more likely to remember to ask for the sticker than they were to remember to change the sign on the door. This finding suggests that the level of interest in receiving the sticker facilitated their prospective memory for this task.

Analyses of the more sensitive measure, the number of cues a participant required to complete tasks, confirmed that 3-year-olds required more prompting to change the sign on the door than they did to retrieve a sticker. These patterns suggest that intentions to carry out low-interest tasks were less salient in participants’ minds while high-interest tasks were more easily encoded, maintained, or retrieved following implicit cues. However, these findings alone do not specify whether intention formation, maintenance, and/or retrieval processes benefited.
According to PAM, participants required fewer cues to remember the high-interest tasks because these tasks engendered greater preparatory attention. Recall that Smith and colleagues (Smith, 2003; Smith & Bayen, 2004) proposed that preparatory attentional mechanisms rely on a limited pool of cognitive resources, known as executive functions. Therefore, 3-year-olds with better executive functioning should have required fewer cues to remember to retrieve the sticker. However, individual differences in executive functioning were not related to the number of prompts 3-year-olds needed to remember the sticker. Rather, the high-interest task seemed to reduce the need for preparatory attentional monitoring and maintenance. These findings appear consistent with the multiprocess framework, suggesting that success in this task did not require preparatory attention, but was achieved through spontaneous retrieval when encountering the target event. Kliegel et al. (2001, 2004) implied that event-based tasks that do not rely on monitoring would not benefit from motivational redirection. These findings clarify the direction of this link. While tasks requiring monitoring likely benefit from motivational prompts, it is more parsimonious to argue that when the inherent nature of the task promotes a certain degree of motivation, an individual’s need to invoke preparatory intentional mechanisms is diminished. In this way, performance on event-based tasks benefits from motivational factors insofar as a participants’ motivation reduces his or her need to maintain an intention in working memory.

According to the multiprocess framework, a representation of the intention must be retrieved upon encountering the target event. As such, individual differences in representational ability should be related to the ease with which children carry out the task, measured by the number of cues they require to do so. This was not the case,
however, children’s performance on the DCCS and belief-desire tasks, thought to support representational ability, was not related to the number of prompts a child received when remembering to receive a sticker.

So what accounts for 3-year-olds’ ability to remember tasks of high-interest? One might argue that it was the bag of stickers’ visibility throughout the experiment that accounted for children’s lack of reliance on representational ability. However, the validity of this explanation is unlikely given that the sign on the door (the low-interest task) was also visible throughout the entire session, yet children’s ability to remember to change it was related to representational ability.

First, consider that the relationship between the encoding and retrieval processes supporting retrospective memory and prospective memory performance is not clearly supported in the literature. Einstein and colleagues (e.g., Einstein & McDaniel, 1996b; Einstein & McDaniel, 2005; Einstein et al., 2005) have often demonstrated that prospective memory does not require preparatory attentional processes, and conclude, almost by default, that spontaneous retrieval processes are responsible for success. However, working with adults and older children, Einstein and McDaniel (1990), Brandimonte and Passolunghi (1994), and Kidder et al. (1997) showed that these two forms of memory are not related. In contrast, Guajardo and Best (2000) demonstrated that retrospective and prospective memory are related for 3-year-olds. Based upon studies with adults, Einstein and McDaniel (1990) proposed that these two forms of memory would be related when the retrospective aspect of the prospective task is challenging. In their studies, the retrospective component of the prospective memory task is usually minimally challenging (remembering what has to be done). It is the prospective
component of the task (remembering when to perform the action) that produces variability in performance. Guajardo and Best (2000) interpreted that the relationship between 3-year-olds' performance on the retrospective and prospective memory tasks might be related because memory tasks, in general, are relatively difficult for younger children. This interpretation may also account for the results here.

The task of getting a sticker may have been on the minds of 3-year-olds throughout the sessions, suggesting it was a current intention, whereas changing the sign on the door was a future intention. There was no desire or reason to change the sign before the session ended, however, 3-year-olds likely desired a sticker from the moment they encoded the intention. Preschoolers were possibly so intrinsically driven to achieve the task that representing this intention was unnecessary. Similar arguments have been made against claims of mental time travel in non-human animals when tasks involve food (e.g., Suddendorf & Busby, 2003). It is possible that the intention to obtain the sticker was an active intention throughout the session, short-circuiting retrieval processes, reducing the demands on the prospective component of the task, and undermining the need to maintain a representation of a future intention. As such, it is likely that an individuals' ability to carry out a high-interest task in the future is a function of inhibitory capacity, rather than prospective memory.

It follows that, if 3-year-olds were preoccupied by the prospective memory task of getting a sticker, they might have performed worse on the executive functioning and belief-desire reasoning measures conducted during the delay. Unfortunately, the pairing of the high-interest and low-interest prospective memory tasks within each session make it difficult to tease out their unique impact on ongoing task performance. Moreover,
because performance was nearly at ceiling on the high-interest tasks (only three children forgot to get a sticker), it is difficult to assess individual differences in performance. However, one might expect a positive relationship between these measures and the number of cues a participant needed to complete the task; if participants who were preoccupied with the sticker task required fewer cues to carry it out, this fixation may have interfered with performance on the ongoing tasks (executive functioning and belief/desire reasoning measures). However, a negative relationship was observed between prompts during the high-interest “remind me” task and the 1-back and digit-span measures of executive functioning (3-year-olds who required more cues to remember the sticker performed worse on these measures), and no other relationships were observed between ongoing tasks and high-interest prospective memory tasks.

From these findings, it is difficult to ascertain that people are engaging in prospective memory when they are highly motivated to complete the task, given the lack of evidence that a monitoring, representational, prospective, or retrospective component is invoked. When 3-year-olds were required to remember to retrieve a sticker, they were more likely to complete the task and needed fewer prompts to do so. Their performance was unrelated to executive functioning, suggesting they were unlikely to recruit limited cognitive resources involved in monitoring. Performance was also unrelated to representational ability, indicating they were unlikely to have relied on encoding and retrieval processes that required a mental model of the intention. Recall that Ellis and Kvavilashvili’s (2000) first requirement of prospective memory is a delay between the formation of an intention and the opportunity to carry it out. However, these finding underscore that prospective memory may be better distinguished from retrospective
memory processes if the following requirement is included in its definition: Ongoing tasks and/or mental activity must interfere with an intention to complete a prospective memory task during the delay interval, that is, prospective memory must involve interruption of an ongoing task and task switching (see Kliegel, Ropeter, and MacKinlay, 2006, for a similar suggestion).

The nature of low-interest tasks makes them more susceptible to forgetting, in contrast. Whereas high-interest tasks are likely to remain excited in memory (anecdotally, 3-year-olds were visibly excited to receive a sticker), less interesting tasks are more susceptible to interference from ongoing tasks and mental activity, especially if the alternatives are more “attention getting” (such as a sticker). Thus, remembering a task that is not particularly interesting may require greater effort on the part of the child, compared to relatively exciting tasks. The nature of this effort was examined here.

According to PAM theory, children require more cues to pass low-interest tasks relative to high-interest tasks because the former rely more heavily on preparatory attention processes and monitoring. Although the number of cues a 3-year-old received before remembering to change the sign on the door was related to individual measures of executive functioning, the executive functioning factor did not predict individual differences in the difficulty with which 3-year-olds passed low-interest, relative to high-interest, tasks, after controlling for age and gender effects. Because the number of cues participants received reflects the degree to which participants were engaging prospective and retrospective components, these findings suggest 3-year-olds were not using preparatory attentional processes to monitor the environment for the target event.

Alternatively, in accordance with the multi-process framework, I predicted the
extent to which 3-year-olds relied on the encoding and retrieval of an intentional representation would be reflected in the relationship between their representational ability and the number of cues needed to retrieve the intention and carry out the low-interest task. This hypothesis was supported by a marginal relationship between representational ability and the number of prompts 3-year-olds received when asked change the sign on the door, after controlling for differences in age, gender, and executive functioning. Three-year-olds with better representational abilities required fewer cues to pass tasks of low-interest, suggesting that representation underlies the development of prospective memory across the third year.

In contrast to the more stringent measure of prospective memory performance (remember vs. forget) described above, the analysis of cues revealed that 3-year-olds required fewer cues to remember tasks on their own than they did to remind the experimenter to carry out the task. Regression analyses revealed that, after controlling for effects of gender, executive functioning, and representational ability, age differences still negatively predicted the number of cues a participant required to pass “remind” tasks relative to “remember” tasks; older 3-year-olds required relatively fewer cues to carry out “remember” tasks compared to younger 3-year-olds. An interaction with age revealed that younger 3-year-olds required more prompting in the low-interest “remember” task compared to older participants. These findings suggest that low-interest “remember” tasks were susceptible to improvement with age across the third year, whereas low-interest “remind” tasks were not.

Low-interest “remind” tasks require an individual to represent an intention and to model what another person represents about this intention (how likely is it that he or she
has forgotten?). Consistent with this interpretation, regression analyses revealed that representational ability uniquely accounted for individual differences in the difficulty with which 3-year-olds remembered to remind an adult to change a sign on the laboratory door, after controlling for executive functioning, gender, and age effects. The processes required by this task not only require a general understanding of representation, prepositional negation, and the ability to use higher-order rules, but may reflect children’s developing reflective competence in coordinating symbolic representation, that is, the ability to reflect on conscious thought (Dick, Overton, and Kovacs, 2005).

These findings provide support for the goal-based model of prospective memory (Penningroth & Scott, 2007). Continued study of the motivational processes underlying prospective memory will hopefully yield a more complete and ecologically valid understanding of prospective memory.

**Gender Differences**

Girls remembered more tasks than boys, and this effect seemed to be driven by girls’ performance on the high-interest remind task, although small sample sizes limited the interpretation of gender comparisons within individual tasks. Somerville et al. (1983), Guajardo and Best (2000), and Smith et al. (2010) did not examine sex differences in their studies. To my knowledge, only Ceci and Bronfennbrenner (1985) have reported sex differences in the time-based prospective memory of children. Kerns (2000) examined sex differences in the time-based prospective memory performance on 7- to 12-year-olds, but failed to find any effects. To my knowledge, the results presented here are the first to document sex differences in event-based prospective memory performance in children as young as 3-years-old.
Female superiority in episodic memory (both recall and recognition; Maitland, Herlitz, Nyberg, Backmand, & Nilsson, 2004; Yonker, Eriksson, Nilsson, & Herlitz, 2003) and semantic memory (Maitland et al., 2004) has been documented in adults, suggesting that the advantage for girls during prospective memory tasks observed here may be related to processes supporting declarative memory, including verbal and representational abilities. Girls also outperform boys on tasks involving theory of mind (e.g., Baron-Cohen, 1995; Baron-Cohen, Knickmeyer, & Belmont, 2005) and other aspects of social cognition (see Baron-Cohen, 2002, for a review), however executive functioning and representational ability only partially mediated the relationship between gender and prospective memory performance here, suggesting that factors still remain which mediate sex differences in prospective memory.

**Limitations**

This study was limited by a small sample size, often resulting in trends approaching statistical significance. For these reasons, caution should be exercised when interpreting findings.

Furthermore, scripts might already exist for aspects of the naturalistic tasks that mirror everyday activities. For example, 3-year-olds at the preschool typically receive a sticker as compensation for participating in a study. It is unknown how often the participants in this study were involved in other studies and the extent to which they had already developed a script for receiving a sticker at the end of the session. This may be one explanation for their success on the high-interest task and may also explain why performance was not mediated by factors affecting prospective memory (i.e., executive functioning and representation). Future studies should present tasks that are of high-
interest to 3-year-olds but less likely to mirror everyday activities, to avoid the possibility that participants rely on scripts to complete the task and to reduce ceiling effects.

In this study, it was assumed that 3-year-olds were more interested in obtaining a sticker than in changing the sign on the door. However, the relative interest or importance of these tasks was not directly measured. Moreover, the interest in each prospective memory task relative to the ongoing tasks was not assessed either. It is possible that individual differences in the relative interest of ongoing and prospective memory tasks predicts how participants allocate attention to these tasks, and subsequently, their performance on each. This possibility is examined in Study 2. Performance on the low-interest task may have also been hindered because it was always paired with the high-interest task. It is unknown if performance on the low-interest task might have been better had it not been in competition with the high-interest task.

It is also a limitation that some executive functioning measures were eliminated from the analyses. Adaptations of the \(n\)-back and backward trail span tasks were explored in this study, and performance on these tasks was generally low, suggesting that these tasks may not be appropriate for preschool children. Moreover, continuous measures are best for capturing individual differences, and the binomial nature of prospective memory tasks and belief/desire reasoning tasks (i.e., pass/fail), as well as the bimodal distribution of some executive functioning tasks, posed a challenge for finding the appropriate analytical method to capture small effect sizes in a small sample. Factor analyses were conducted and proved beneficial in confirming the executive functioning and representational ability constructs, as well as highlighting those executive functioning measures that had not performed well in this study of 3-year-olds. Ordinal regression
analyses were considered but are problematic when predictor variables are continuous (as they were here), rather than categorical. As such, I proceeded with factor and subsequent linear regression analyses on the number of prospective memory cues participants received. However, caution should be used when interpreting these linear relationships, as the dependent variables (number of cues) are indeed discrete ordinal.

Finally, only first-order theory of mind measures were used in the present study. However, like the low-interest “remind” prospective memory task, second-order theory of mind is the ability to coordinate multiple symbolic representations of others’ mental states (“Justin thinks that I’ll remember that he wants...”) and is not expected to develop until age 5- to 7-years (Perner, Heinz, & Wimmer, 1985). Future studies should examine how individual differences in second-order theory of mind processes predict the development of prospective memory performance in tasks that require reminding.

**Conclusions and Implications**

Quantitative changes, such as central executive capacity, may be important to the development of prospective memory, but qualitative changes associated with the development of representational abilities should also be considered to contribute to prospective memory development. Although this study does not allow the conclusion that representational abilities precede prospective memory development, it does indicate that prospective memory is another symptom of the qualitative change that emerges in many domains across the third and fourth year of life.

This is the first study to directly demonstrate the role of individual differences in representational abilities in prospective memory development. These findings also replicate the work of others, demonstrating the effect of motivational factors and
individual differences in executive functioning on predicting prospective memory (e.g., Ceci & Bronfenbrenner, 1985; Kerns, 2000; Somerville, 1983). Continued work in this area is important for understanding age-appropriate prospective memory tasks, especially as children enter preschool and elementary school.
CHAPTER THREE

INTRODUCTION TO TIME-BASED PROSPECTIVE MEMORY

Examining the development of processes underlying prospective memory is complicated by the subtle, but important, distinction between time-based and event-based tasks. While the processes supporting event-based and time-based prospective memory are overlapping in many respects, there is widespread agreement among researchers that the retrieval of time-based tasks is fundamentally different from event-based tasks (Einstein & McDaniel, 1996a, 1996b; Sellen, Louie, Harris, & Wilkins, 1997).

The development of time-based prospective memory abilities is delayed relative to event-based tasks, possibly because time-based prospective memory tasks typically require an explicit monitoring component (Kvavilashvili & Fisher, 2007) and are less amenable to spontaneous retrieval processes than event-based tasks that benefit from relatively more explicit target events. Smith et al. (2007) suggested that preparatory attention processes involved in remembering that something must be done (i.e., the prospective component) undergo little change in the early elementary school years.

While some research supports this hypothesis, suggesting that individual differences in executive functioning provide a better account for time-based prospective memory than event-based prospective memory performance (Ellis, 1996; Kvavilashvili, Messer, & Ebdo, 2001; Kliegel et al., 2002; Wang et al., 2007), others suggest that these relationships are phase-specific; studies with adults suggest that event-based prospective memory recruits executive functions involved in task-switching, or set shifting, during
the final execution phase of event-based prospective memory tasks (e.g., Einstein, Smith, McDaniel, & Shaw, 1997; Smith et al., 2007, but see Martin, Kliegel, & McDaniel, 2003), while supervision effects, but not shifting effects, may reflect task-specific demands of time-based prospective memory. Below, I describe research that investigates these hypotheses of time-based prospective memory.

Lab-based research on time-based prospective memory tasks has suggested that certain patterns of behavior tend to occur prior to successful performance of a task. After receiving instructions, adult participants reportedly demonstrate what has been called the test–wait–test–exit (TWTE) cycle (Harris & Wilkins, 1978). Participants first test the time (typically by looking at a clock), then wait for a period of time until another test appears appropriate, and continue looping through test–wait cycles until the critical period to respond arrives, at which point they perform the action and then exit the loop.

Ceci and Bronfenbrenner (1985) examined time monitoring in a time-based prospective memory task in 10- to 14-year-old children who were asked to either remember to take cookies out of the oven or recharge a motorcycle battery after a specified amount of time had passed. Children were asked to carry out these tasks in either a familiar environment (the child’s home) or an unfamiliar environment (the lab). Children’s clock-checking behaviors were measured as the number of times each child turned around to look at a clock while playing a Pacman video game.

Ceci and Bronfenbrenner described older children as demonstrating a strategic, U-shaped pattern of checking behavior. Building on the TWTE pattern described by Harris and Wilkins (1978), Ceci and Bronfenbrenner reported that strategic participants initially calibrated themselves by checking the clock frequently, followed by a reduction in clock
checking in order to attend to other activities, and then finally engaged in a scalloping phase during which frequency of clock checking increases as the deadline approaches. The authors argued that this style of clock checking was indeed strategic. Children who used this strategic monitoring pattern the same level of accuracy as children who did not use the strategy. Anxious time monitoring, in contrast, was best illustrated by a linear function, as frequency of clock checking increased steadily across the time period (Ceci & Bronfenbrenner, 1985).

Factors other than age also seemed important in predicting strategic monitoring. For example, girls in Ceci and Bronfenbrenner’s study used strategic time monitoring less for baking, boys for battery charging. Ceci and Bronfenbrenner interpreted these findings to reflect anxiety provoked by sex-specific expectations. That is, boys may feel more pressure to perform well in the battery-charging context, whereas females feel more pressure to perform well in the baking context. In addition, 10-year-olds performed more strategically in the familiar home environment relative to the lab environment. The familiarity of the environment was especially important for 14-year-old females; strategic time monitoring began to disappear and look like anxious time monitoring when older females were baking in the lab setting as opposed to at home. This was the pattern for both sexes in the battery charge condition. These findings indicate that, in addition to age, there are important interactions between sex, the nature of the task, and the task setting that must be considered when predicting prospective memory performance.

While research indicates that 2-year-olds are capable of carrying out event-based prospective memory tasks (e.g., Somerville et al., 1983, but see Kliegel & Jager, 2007), time-based prospective memory has only been demonstrated in children as young as 7-
years-old (Kerns, 2000). However, younger children are often unable to tell time, precluding the use of time-based tasks involving analog clocks or the identification of numbers. Therefore, it is difficult to assess whether the relative delayed development of time-based prospective memory is the result of the more demanding nature of time-based prospective memory on executive functioning or to task-specific requirements (telling time versus recognizing a target event).

To conquer this issue, Kerns devised a clever game called the CyberCruiser that engages children in a car race but requires them to monitor their “gas tank” and refill it before it empties. Because the fuel tank features a gauge that consistently depletes over time, it serves as an analogous measure of adults’ time-monitoring and allows researchers to investigate children’s checking behaviors and success on time-based prospective memory tasks.

Using the CyberCruiser, Kerns (2000) examined the role of executive functioning in the time-based prospective memory performance of 7- to 12-year-old children. Participants’ primary goal was to earn points by using a joystick to maneuver their car around other vehicles and road hazards (i.e., the ongoing task). Their prospective memory task was to monitor their car’s fuel level, which was not visible but could be viewed as often as desired by hitting a button on the joystick. Participants were to refuel their car by hitting another button when the “tank” was less than 1/4 full and not before (i.e., the prospective memory task). In addition, participants completed a battery of executive function measures including two tasks of visuospatial working memory (delayed alternation - nonalternation and a self-ordered pointing task) and two measures of inhibitory capacity (go-no-go and a Stroop task).
Kerns (2000) found that age was negatively related to the number of times children ran out of gas ($r = -.29$), suggesting that time-based prospective memory performance improves with age. Kerns also demonstrated that children as young as 7-year-olds were demonstrating a strategic checking pattern, similar to adults (the TWTE pattern described by Harris & Wilkins, 1978). In fact, neither the frequency of gas-checking nor patterns of strategic monitoring were related to age. After partialling out age, failures on the prospective memory task (times out of gas) positively correlated with selective measures of executive functioning: namely, visuospatial working memory (delayed nonalternation errors), self-ordered pointing errors, and inhibition (Stroop interference), but frequency of gas-checking or checking pattern was not related to any measures of executive functioning. These findings are curious; even though children as young as 7-year-olds were demonstrating strategic monitoring on par with older children, they were more likely than older children to forget to refill the tank in time (cf. Ceci & Bronfenbrenner, 1985). Kerns suggested younger children may fail to benefit from strategic monitoring for two reasons: First, they may forget their intention to fill the car (a retrospective memory failure), or second, younger children may not effectively calibrate their inner clock with the fuel gauge, such that they overestimate how long they can go before they need to fill up.

To further examine these issues, Mäntylä, Carelli, and Forman (2007) compared the performance of 8- to 12-year-old children with that of adults on a time-based prospective memory task and three components of executive functioning: inhibition, updating, and mental shifting. In the time monitoring task, participants were asked to watch a movie and indicate the passing of every five minutes by pressing a red button.
Participants could check the clock at any time by pressing a green button. Mäntylä et al. (2007) differed from Kerns (2000) in the tasks used to measure executive functioning. The authors proposed that time monitoring involves the revision of task-relevant working memory representations and predicted that tasks assumed to tap the updating component of executive functioning would mediate the relationship between participants’ time monitoring performance and prospective memory success (a relationship Kerns [2000] failed to find). They used a matrix monitoring and n-back task to measure updating/working memory. Mäntylä and colleagues also assumed that updating and inhibition would be correlated, rather than distinct, components of executive functioning and that time monitoring performance would be related to individual differences in both components. The Stroop and stop-signal tasks were used to measure inhibition. Finally, a connections task, based on the trail-making task, and a category fluency task, were administered to assess task switching.

In contrast to Kerns (2000), Mäntylä and colleagues did not find a relationship between timing error (failures on prospective memory task) and age. Older and younger children and adults provided more than 80% of the target responses within 10 s. Children and adults used the same general strategy for monitoring deadlines (i.e., a quadratic, J-shaped, TWTE function of interval reduction), replicating the findings of Kerns (2000). However, children required more clock-checks (i.e., a greater reliance on external time keeping) to maintain this high level of performance. Children checked the clock at a mean rate of 1.67 times/min while adults’ checked at a frequency of 0.70 times/minute.

Factor analyses revealed that the inhibition and updating tasks constituted a common “supervision” component, as predicted, and shifting tasks constituted a separate
“shifting” component for both adults and children. As predicted, the supervision component, but not the shifting component, of executive functioning related to timing error (i.e., prospective memory performance) for both children and adults, a finding consistent with Kerns (2000). In contrast, Mäntylä and colleagues found that selective executive functions (i.e., supervision) negatively correlated with frequency of clock checking for children (recall Kerns [2000] did not find a relationship between executive functioning and clock checking). That is, children demonstrating poorer performance in supervision components of executive functioning seemed to rely more on external time keeping as a compensatory measure during the prospective memory task compared to those children with better updating and inhibition task performance. However, those with more frequent clock checking (i.e., poor supervision performance) were still less accurate in reporting the passage of five minutes suggesting that these compensatory strategies were inefficient.

This pattern is consistent with the proposal that younger children suffer a utilization deficiency (Miller, 1990; Miller & Seier, 1994). These findings, coupled with Kerns’ (2000), indicate that although younger children may be implementing strategic monitoring to the same extent as older children, they fail to benefit from doing so. As Smith et al. (2010) suggested, this may be due to younger children’s poorer performance on retrospective components of the task.

Mäntylä et al. (2007) argued that possessing an intuitive sense of time functions to reduce the costs of monitoring and is most likely mediated by processes that allow for the temporal supervision (i.e., maintenance and updating) of working memory contents. This argument is supported by the selective effects of executive functioning (supervision
effects, but not shifting effects) and may reflect task-specific demands of time-based prospective memory relative to event-based tasks. That is, both time-based and event-based prospective memory may rely on monitoring the environment for target cues. In time-based tasks, variability in success is predicted by individual differences in one’s “sense of time” and executive functions that support temporal supervision, insofar as these processes allow one to predict the occurrence of the target cue. People with more efficient temporal supervision functioning are more likely to meet deadlines. In contrast, those with less efficient updating and inhibition functions would be more likely to experience discontinuities in their sense of time and require earlier and more frequent dependence on external time keeping (e.g., clock checking).

In the study of Mäntylä et al. (2007), the relationship between time monitoring and supervision was not observed in adults, suggesting that inefficient monitoring performance in children.

**Beyond Executive Functioning**

Prospective memory researchers have advocated for the assessment of developmental and individual differences in factors and processes influencing prospective memory performance (Einstein & McDaniel, 2010; Smith, 2010). Although these efforts have primarily focused on executive functioning accounts, findings are generating novel hypotheses about how higher-level processes support time-based prospective memory (e.g., Mäntylä et al., 2007). Ceci and Bronfenbrenner (1985) are among those who have indicated that factors beyond executive functioning are important for predicting time-based prospective memory performance, such as gender (Ceci & Bronfenbrenner, 1985) and motivation (Kvavilashvili & Fisher, 2007), as well as external
factors like setting and task type. In the next half of this chapter, I review some other factors, beginning with time estimation abilities.

**Time estimation**

Besides executive functioning, it has also been proposed that time-based prospective memory may rely partly on time estimation abilities (Goldstein, 2005; Mäntylä & Carelli, 2005; see also Carelli, Forman, & Mäntylä, 2008). Many time-based prospective memory tasks require the representation of a discrete temporal unit (e.g., press a key every 5 min); hence, the ability to estimate such durations may itself be a predictor of time monitoring and/or time-based prospective memory success (MacKinlay, Kliegel, & Mäntylä, 2009). A meta-analysis (Block, Zakay, & Hancock, 1999) indicated that children overestimate experienced durations, with the precision of these estimations increasing from 7 to 10 years of age, and underestimate time when asked to make time reproductions compared with adolescents and young adults. In other words, time seems to stand still for children. Children were generally able to make comparable productions of time durations, however.

The empirical evidence for time estimation as a potential factor underlying time-based prospective memory is both scarce and contradictory. Mäntylä and Carelli (2005) investigated time estimation as a possible predictor of performance on the time-based prospective memory movie-watching task described earlier (Mäntylä, Carelli, & Forman, 2007). Age differences were not observed between school-aged children and young adults when asked to reproduce durations of between 4- and 36-seconds, and time reproduction did not correlate with time monitoring performance on the time-based prospective memory task.
In contrast, and contrary to expectations, Goldstein (2005) reported a significant negative relationship between time estimation and time-based prospective memory accuracy as well as clock monitoring. Young adults who accurately produced 1-minute intervals made significantly more timing errors on a time-based prospective memory task (pressing a button every 7 min) and checked the clock more frequently than did inaccurate estimators.

More recently, MacKinlay, Kliegel, and Mäntyla (2010) examined individual differences in 7- to 12-year-olds’ time-based prospective memory (press a yellow button every 2 minutes while completing the other tasks) and time estimation, in addition to ongoing task performance, time monitoring, working memory (as measured by a digit span backward [HAWIK] task), task switching (using a task similar to the alternation tasks described above), and planning (researchers administered the Zoo Map Test, asking children to plan a route through the zoo, following certain rules). Time monitoring was measured using four tasks. In the prospective verbal estimation task, the child was asked to provide a verbal estimation of the time between two tones presented two minutes apart. In the prospective production unfilled task, the experimenter started a stopwatch and the child was asked to say “stop” when he or she thought 2 minutes had passed. In the retrospective estimation task, the child was asked to estimate the amount of time spent performing a task that had taken 2 minutes. The prospective production filled task was like the prospective production unfilled task, except the child completed a puzzle during the 2 minute interval.

Hierarchical regression analyses revealed that the majority of age-related variance in time-based prospective memory performance could be explained by executive
functioning (working memory, task switching, and planning), whereas no further independent contribution of time estimation was observed.

**Strategic monitoring**

Harris and Wilkins (1982) first described the TWTE pattern in an experiment in which adult participants were requested to lift a card from their laps either every 6 or every 9 min while watching a 2-hr video. As described elsewhere (Ceci & Bronfenbrenner, 1985; Kerns, 2000), Harris and Wilkins noted that participants checking behavior yielded a J-shaped function; participants checked the clock frequently at the outset of the task, after which clock checking decreased before increasing again just prior to the target time.

Subsequently, Ceci and Bronfenbrenner (1985, reviewed above) described a U-shaped pattern of checking behavior in children’s performance on time-based prospective memory tasks at home. In a follow-up study, Ceci, Baker, and Bronfenbrenner (1988) varied the speed of the clocks (by 10%, 33%, and 50%) when asking participants to perform prospective memory tasks at home. Their 10-year-old participants adjusted their TWTE strategy to the new clock speed in the 10% and 33% adjustment conditions, maintaining the U-shaped pattern of checking behavior observed in the 1985 study. However, in the 50% acceleration/deceleration condition, participants resorted to a linearly increasing, anxious clock-checking pattern, as if realizing they could no longer trust their internal clock because it did not match the external clock (Kvavilashvili et al., 2008).

Kerns (2000) and Mäntylä et al. (2007) also demonstrated the J-shaped pattern described by Harris and Wilkins (1982) in all children, 7 to 12 years old, irrespective of
age\textsuperscript{1}. Consistent replication of this pattern across multiple age groups (see also Einstein et al., 1995; Kvavilishvili & Fisher, 2007; Park et al., 1997) has led researchers to suggest strategic monitoring is automatic (Kerns, 2000; Smith, 2010). Children and adults seem to be fairly unaware that the strategy is being employed (e.g., Ceci & Bronfenbrenner, 1985). However, 10-year-olds abandonment of the strategy when the clock is accelerated or decelerated substantially suggests that on some level their strategic monitoring depends on their awareness of the fallible nature of their own internal clock under certain conditions (Ceci et al., 1988). This raises a question worthy of empirical investigation; is strategic monitoring influenced by individual’s confidence in their time estimation abilities?

Research has demonstrated that better time estimation does not predict better time-based prospective memory (e.g., MacKinlay et al., 2009), but there is yet to be an examination of the relationship between time estimation and strategic monitoring during a time-based prospective memory task. I examine this relationship in Study 2.

It may not be time estimation, per se, that predicts time-based prospective memory, but an individual’s understanding of his or her time estimation abilities that predicts checking patterns and subsequent performance. “Metacognition” is the term used

\textsuperscript{1} The discrepancy in the J-shaped pattern of performance observed by Harris and Wilkins (1982), Kerns (2000), Mäntylä et al. (2007), and the U-shaped pattern observed by Ceci & Bronfenbrenner (1985) has been attributed to the 1-trial measures of checking behavior in Ceci and Bronfenbrenner (1985) as opposed to the multiple trial measures adopted in other studies. Thus, once children calibrate themselves with the “clock” (or gas gauge needle) during the early phase of the first trial, they are likely to reduce this checking behavior during early phases of subsequent trials because they need not recalibrate their internal clock. This decrease in the checking frequency during early phases on subsequent trials drives down the mean frequency during this phase across trials, yielding a smaller “hook” or J-shaped function, instead of a U-shaped function, during the early calibration phase (Kerns, 2000).
to refer to “knowing about knowing,” and includes knowledge about when and how to use particular strategies for learning or for problem solving (Metcalfe & Shimamura, 1994). Anxiously monitoring time requires a participant to understand that his or her own time estimation is poor under present conditions. Young children’s tendency to overestimate their abilities on a number of skills (Bjorklund, Coyle, & Gaultney, 1992; Boulton & Smith, 1990; Flavell et al., 1970; Humphreys & Smith, 1987; Mills & Keil, 2004; Pickert & Wall, 1981; Plumert, 1995: Plumbert & Schwebel, 1997; Shin & Choi, 2002; Schneider, 1998; Yussen & Levy, 1975; Spinath & Spinath, 2005; Stipek, 1981) suggests poor metacognition, and may account for the utilization deficiencies described in the studies above. Young children may perform poorly on the time-based task because they think they are very good at estimating time, when in fact they are not, resulting in their failure to adjust their clock monitoring accordingly. In their monograph, Flavell, Green, and Flavell (1995) demonstrated that 3- to 5-year-olds know that thinking is an internal mental activity, but they are poor at recognizing that they themselves have just been thinking, and are mostly unable to describe what they were thinking about. However, by 7 to 8 years of age, children seem to reason about mental activity as less mysterious and develop the understanding that mental activity can be the result of preceding causes and may have subsequent effects.

If this is the case, time estimation abilities should improve with age, alongside participants’ metacognitive understanding of their own abilities. Thus, I predicted that young children who have poor metacognitive representations of their time estimation abilities would demonstrate lower-frequency, or even “strategic” (TWTE) monitoring and
poor time-based prospective memory performance (i.e., fail to remember to carry out the
task at the appropriate time). In contrast, older children and adults should be more likely
to appropriately understand their time estimation abilities. Those who accurately perceive
that they are good at estimating time should demonstrate strategic monitoring and
perform well on the prospective memory task. Those who accurately perceive that they
have a poor “sense of time” will demonstrate a linear, anxious pattern of checking, and
may still perform well on the prospective memory task. However, these participants may
suffer costs in an ongoing task due to nonstrategic (i.e., anxious) monitoring. This pattern
of findings would demonstrate that children’s poor metacognitive representations of their
time estimation abilities functions to promote strategic monitoring, allowing them to
continue performing well on ongoing tasks (see Shin, Bjorklund, & Beck, 2007), albeit at
the sake of the prospective intention.

Use of external strategies

In a similar vein, children’s ability to make use of external strategies should
influence their performance on time-based and event-based prospective memory tasks.
Adults use both internal and external cues to facilitate their remembering (Harris, 1978),
but external cues are usually more beneficial than internal cues (Einstein & McDaniel,
1990; West, 1988) and are more preferred (Intons-Peterson & Fournier, 1986). It has
been proposed that older adults are more likely to use memory aids because they are
aware of the fallible nature of their memories, compared to younger adults, and seek
external cues as a compensatory measure (for a review, see Phillips, Henry, & Martin,
2008).
Previous research has shown that children tend to use increasingly more effective memory strategies with age, yielding a corresponding improvement in memory performance (see Bjorklund, Dukes, & Douglas, 2009; Coyle & Bjorklund, 1997; Grammer, Purcell, Coffman, & Ornstein, 2010; Schneider & Pressley, 1997). With the exception of a few studies, little is known about how children understand and select prospective memory strategies.

When performing prospective memory tasks in Guajardo and Best’s (2003) study, 3-year-olds demonstrated a rudimentary understanding of effective strategies. For example, nearly one-third of the 3-year-olds chose a strategic means for remembering to press the key when the target picture appeared (see also Kreutzer, Leonard, & Flavell, 1975; Intons-Peterson & Fournier, 1986). However, the development of such skills across the next two years led 5-year-olds to perform better than younger children on prospective tasks across both short and long delays. Kreutzer et al. (1975) and Beal (1988) demonstrated that 8-year-olds were able to choose or describe effective external strategies in an event-based prospective memory task comparable to adults. However, the use of external strategies has yet to be examined in time-based prospective memory. As such, it is unclear whether children recognize the appropriate prospective memory contexts for adopting external strategies or whether they can carry through with the retrospective component of the task (remembering what is to be done) once the external strategy is adopted.

When the costs of remembering to carry out a particular activity at a specified time become great, one can reduce the interference of monitoring by creating an external reminder that effectively reduces the time-based task to an event-based task. For
example, if I need to remember to attend a meeting at 2:00 p.m., I can rely on my own internal sense of time, monitor an external time-keeper (checking the clock every time I think about the impending meeting, and thus cope with the accompanying distraction), and/or I can externalize the reminder by setting an alarm on my desktop calendar. In this way, strategies allow one to move prospective memory along the continuum toward prospective memory proper and away from vigilant time monitoring, described by Graf and Uttl (2001). Choosing the external reminder (the timer) requires that I (1) accurately assess my sense of time as poor (i.e., I’m likely to miss the 2:00 target time), (2) will forget what it is I have to do at 2:00 and/or (3) am aware of the costs (i.e., distraction) to ongoing tasks caused by monitoring the clock and consider this cost to be greater than the costs of setting an external reminder.

Thus, one might ask, do people’s awareness of costs associated with time-based prospective memory affect their willingness/ability to make use of appropriate external strategies? If an individual’s awareness of the costs of prospective memory predicts whether he or she will adopt strategies to reduce these costs, then participants who accurately assess a prospective memory task as costly (i.e., preoccupying or interfering) should adopt appropriate external reminders as a result. Moreover, I predict that, as children’s metacognition improves with age, they should become increasingly aware of the costs of the prospective memory task, and, at some point, this metacognitive awareness should account for their strategy use. Similarly, participants who are aware of the fallible nature of their time estimation abilities should adopt strategies that will be more reliable. Study 2 investigates these questions.

Motivation
Following up Study 1, Study 2 also examines how the relative importance and interest of the ongoing and prospective memory tasks interact with participants’ metacognitive understanding to influence attention allocation and strategy use in a time-based prospective memory task. Although Penningroth and Scott (2005b, 2006) reported that participants used strategies more frequently to carry out tasks rated relatively important, they did not find any differences in the frequency of internal (rehearsal or monitoring) or external strategies. These observations were reported for event-based, not time-based, prospective memory tasks, however. Recall that Kliegel et al. (2004) found that task importance did influence event-based prospective memory performance if the task was relatively monitoring-demanding.

Therefore, I explore how the goal-based model of prospective memory applies to time-based tasks in Study 2, examining how participants’ ratings of relative importance of the prospective memory task influence their fuel-checking and external strategy use. Based on the goal-based model (Penningroth & Scott, 2007) and assumptions of Kliegel et al. (2004), I predict that participants who view the prospective memory task as more important will engage in more frequent monitoring and use of external strategies, relative to those who rate the ongoing task as less important. To some degree, I do expect this relationship to strengthen with age, given that adults should be better able to strategically allocate attention in accordance with motivation. However, given that effects of motivation were apparent when 3-year-olds performed event-based tasks, I would not be surprised to find that motivation was influential in impacting the performance of 5-year-olds.
CHAPTER FOUR

STUDY 2. TIME-BASED PROSPECTIVE MEMORY, METACOGNITION, AND STRATEGY-USE IN ELEMENTARY-AGED CHILDREN AND ADULTS

At the first level, my second study examines time-based prospective memory in young elementary-aged children and adults using an Xbox 360®-style racing game, adapted from the Cyber Cruiser task developed by Kerns (2000).

The ongoing task was to drive the NASCAR-type car fast and avoid crashing so as to earn a “top-five” lap time. Participants’ lap time and number of crashes were recorded as dependent measures of ongoing task performance.

The prospective memory task was to refill the gas tank when needed. Participants’ frequency of gas checking was recorded as the measure of monitoring.

Participants completed the game three times, once needing to refill the gas tank (prospective memory block), once without needing to refill the gas tank (control block), and once with the option to adopt one of two external strategies to remember to refill the gas tank (external strategies block). By including a control block where participants only engaged in the ongoing task (i.e., driving the car), the costs of engaging in the prospective task could be computed by comparing participants’ lap time in the control block to their lap time in the prospective memory block.
Four time estimation tasks (adapted from MacKinlay et al., 2009) were also administered, allowing examination of the relationship between participants’ time estimation abilities and prospective memory performance. Participants were asked to estimate various aspects of their performance, either before or after completing the task, and their accuracy in doing so was used to predict their performance and strategy use. Therefore, at a second level, this study examined participants’ use of external strategies, gas-checking strategies, and prospective memory performance as a function of metacognitive estimations of (1) prospective memory performance, (2) costs incurred as a result of this task, and (3) confidence in time estimation abilities. Multilevel linear modeling techniques assessed how these relationships varied with age.

Participants were also asked to explicitly report the relative importance of the ongoing and prospective memory tasks. In this way, this study builds on Study 1 by further investigating the relationship between motivation and prospective memory performance.

Methods

Participants

Pilot research revealed that children as young as 5 years old were capable of understanding and participating in the new version of the CyberCruiser task (hereafter, CyberCruiser 2.0). A total of 115 participants completed the experiment: Twenty-seven kindergarten children (M age = 5y, 7mo; males = 13), 27 first-grade children (M age = 6y, 8mo; males = 18), 29 second-grade children (M age = 7y, 6mo; males = 17), and 32 adults (M age = 18y, 4mo; males = 16).
Children were recruited from Henderson Elementary school, affiliated with Florida Atlantic University, and were offered a sticker after the completion of the study. Adults were undergraduate students recruited from Florida Atlantic University’s subject pool who received partial course fulfillment for their participation. Ethnic distributions did not vary by age group; in total, 73 (64.0%) of participants were White/Caucasian, 17 (14.9%) were Black, 17 (14.9%) Hispanic, 6 (5.3%) Asian, and 1 (.9%) was classified as “other.”

Materials

The original CyberCruiser task, developed by Kerns (2000) was written in Visual Basic 4.0 and was no longer compatible with current software versions and libraries. The Cyber Cruiser 2.0 was designed, using a Microsoft XNA framework, to operate similar to a Xbox 360® 3-D racecar game and was operated using a Xbox 360® controller. The game was operated on an Apple MacBook 17-inch laptop.

The CyberCruiser 2.0 was designed to engage young children in ongoing and prospective memory tasks by asking them to race their virtual car around a racetrack and monitor their fuel level in the process, refueling when necessary. The CyberCruiser 2.0 offers three distinct racetracks: a beginner, intermediate, and advanced track. All participants raced on the beginner track because this was shortest, and practiced on the intermediate or advanced track. The game also allows the selection of various types of “cars,” each with unique aesthetic, as well as speed, acceleration, mass, braking, friction, and engine features. All participants raced the same car with the same settings but were allowed to choose their cars’ color (from white, yellow, cobalt, fuschia, red, green, teal, silver, tangerine, banana, or mint; See Appendix B for a screen shot).
The game was designed so that participants could monitor their fuel level by pressing a button on the controller to make a fuel gauge appear for 1-s on the lower left side of the screen. Upon the start of the trial (i.e., race), the fuel gauge depleted steadily for 30-s until running empty. The fuel was available for refill when the lever reached the “the red zone,” at 25-s post-trial initiation, by pressing a second button on the controller that would trigger the gauge to return to “full.” The car could not be refueled before this time, even if the participant pressed the refuel button. If the gas was not refueled when the lever was in the red zone, the car stalled out, and a message appeared on the screen directing the participant to push the appropriate button to refuel the tank.

Unlike the original CyberCruiser designed by Kerns, the 2.0 version used here allowed for the option of refueling the tank; in this way, the experimenter manipulated the settings such that participants could race without running out gas in one block, and therefore, without needing to monitor the gauge or refuel (henceforth, control block), and also race a second block requiring them to monitor and refuel the tank (henceforth, prospective memory block).

Participants were given five opportunities to refill the gas tank during the 2.5-minute prospective memory and external reminder blocks. The top five “best lap times” were displayed at the top right of the screen and the participant’s speed and current lap time were displayed in the bottom corners (see Appendix C for a screen shot).

A Timer application (Bonnin, 2009) on an Apple iPhone was used as one of the external reminders in the third and final external reminder block. The timer was pre-set for 27 seconds and started by the participant pushing the touch-screen button when beginning the race, such that the alarm “beeped” to indicate that the fuel gauge was ready
for refill. The timer continued to beep until the participant pressed “OK” or “cancel.”

Handily, the timer automatically reset to begin the next countdown so that even if the participant was delayed in silencing the indicator, the 27-s cycle was not delayed in restarting and reliably alerted the participant that the fuel gauge needed refueling. However, the participant also had the option to press “cancel” if he or she did not wish to continue using the timer.

After pilot testing, the controller sensitivity was adjusted for adults to make the driving task slightly more difficult. Kvavilashvili et al. (2008) noted the importance of making the ongoing task comparably difficult for older participants when comparing prospective memory performance across age groups.

A stopwatch, bell, and tangled lines task were used to assess time estimation. The Kaufman Brief Intelligent Test (KBIT) vocabulary and matrices tasks (Kaufman & Kaufman, 1990) were also administered to control for intelligence.

**Procedures**

Participants were tested individually in a private room. Figure 11 demonstrates the counter-balanced block presentations of the Cyber Cruiser 2.0 task. A 2 x 2 between-subjects factorial was used to counterbalance block order (prospective memory versus control blocks) and metacognitive condition (pre-trial questions or post-trial questions only), and participants were randomly assigned to one of four conditions. All participants received the external strategies condition last. This design was intended to control for practice effects associated with participating in two blocks. Likewise, the order of presentation of metacognition questions was counterbalanced between participants, such that some participants were asked to estimate their performance before racing the
prospective memory block, while others were asked about their performance retrospectively, after racing the prospective memory block. This was done to investigate any differences in performance that may be caused by probing children in advance about their performance.

First, the experimenter explained that the goal of the game was to drive the car fast and get the best lap time. The experimenter then demonstrated the task of driving the car before turning the controller over to the participant and asking him or her to practice. The practice block lasted for approximately 1-min for all participants. For children, the experimenter pointed out the passing minutes and seconds as indicated by the lap timer on the screen, describing that seconds are faster than minutes. Practice blocks took place on the “Expert” track, which was longer than the track used for the other blocks, so that participants would not complete an entire lap during practice, but simply practice handling the joystick and using it to control their racecar.

After practicing the game, participants assigned to receive the prospective memory block first were given verbal instructions and a demonstration on how to refill the gas tank while driving (prospective memory block). Half of these participants were asked to anticipate various aspects of their performance before they were allowed to race (pre-trial metacognition condition). Participants’ understanding of the prospective memory task was then confirmed by their indication of the correct buttons to push when checking the fuel gauge and refueling the tank. Each individual engaged in racing the prospective memory block for 2.5 minutes. Post-trial condition participants then answered questions about their performance. During the control block, everyone was instructed they no longer had to refuel the gas tank, but that they could simply
concentrate on driving fast and achieving a fast lap time. Pre-trial participants then answered metacognition questions. All participants raced the control block for one lap. Post-trial participants then answered metacognition questions.

For those assigned to receive the control block first, the game was reset after the practice block and participants were verbally instructed to play the racing game “for real” (control block). These individuals completed one lap of the control block before receiving verbal instructions and a demonstration for the prospective memory block. Half of these participants were asked to estimate various aspects of their performance prior to racing. Everyone raced the prospective memory block for 2.5 minutes, or until they had been given five opportunities to refill the gas. Those assigned to the post-trial condition were asked to estimate various aspects of their performance following the race.

The experimenter always pointed out participants’ lap times for the first block, regardless of condition, to offer a baseline for performance. For instance, “Great job! Your best lap time was 1 minute and 30 seconds.”

**Metacognition questions.** *Prospective memory metacognition:* Prior to racing the prospective memory block, only participants in the pre-trial condition were asked, “How many times do you think you will remember to fill the tank? If you have to refill the gas tank five times during the game, for example, how many times do you think you’ll fill it before it goes empty?” After completing the prospective memory block, participants in the post-trial condition were asked, “How many times did you refill the tank?” The question was phrased in this way because it was thought that running out of gas would be a relatively distinct event, given that the car stalled out and visual cues to refill the tank were presented, and that participants in the post-trial condition might have more accurate
recall of these salient negative events. In contrast, successful refills were not as marked as running out of gas, and participants’ estimates of these events post-trial would be a better gauge of their confidence in their prospective memory abilities, rather than a measure of their recall. Responses were reverse-coded (subtracted from five) to reflect estimates of the number of unsuccessful refills.

*Understanding the costs of prospective memory.* Prior to racing the second block, participants in the pre-trial condition were asked, “Now that you [don’t] have to refill the gas tank, do you think you will have a faster or a slower lap time than before? How many seconds [faster] slower do you think you will be, now that you [don’t] have to refill the gas tank?” Participants in the pre-trial condition who completed the prospective memory condition were asked the bracketed version of this question prior to completing the control condition. Participants in the pre-trial condition who completed the control condition first were asked the non-bracketed version prior to completing the prospective memory block second.

Participants in the post-trial condition were asked, “Now, that time you had to [didn’t have to] check the gas tank, so do you think that made you get a faster or a slower time than before? How many seconds [faster] slower do you think you went, now that you had to [didn’t have to] refill the gas tank?” Participants who completed the prospective memory condition first were asked the bracketed version of this question prior to completing the control condition. Everyone who completed the control condition first was asked the non-bracketed version prior to completing the prospective memory block.
**Motivation assessment.** Each individual was shown two tokens: a car, representing the ongoing task, and a fuel gauge, representing the prospective memory task. The experimenter pointed to each respective token while asking, “Can you point to the one that is more important to you? Which is more important to you -- to make sure the gas doesn’t run out or to get a fast lap time?” Participants then pointed to a token to indicate the prioritized task. Next, the experimenter showed participants a scale, ranging from 1 to 7, and asked, “Can you show me how important it is to you to get a fast lap time? Can you show me how important it is to you to be sure you don’t run out of gas?” By placing the tokens on the scale, participants rated the importance of the prospective memory and ongoing tasks. Pre-trial condition participants were asked this question prior to the prospective memory block, regardless of block order. Participants in the post-trial condition were asked this question after racing the prospective memory block.

**External reminder condition.** During the third and final block, all participants were instructed that they were to race one more time, and that they must again be sure to refill the gas tank and avoid letting the gas run out. The experimenter presented and demonstrated two external reminders by saying,

“This time I’ll give you some tips that might make it easier to remember to refill the tank. Sometimes I start this timer at the beginning of the race like this (experimenter demonstrates) so that it will beep when it’s time to refill the tank. That way, I don’t have to worry about checking the tank; I can just concentrate on driving fast and wait for the beep. Once it beeps, I just push the ‘OK’ button to restart it.
Another thing you can do is watch for the green billboards on the side of the track (experimenter points out). Sometimes I use those as a visual cue to check the gas tank. So every time I pass a green billboard, I check my gas level.”

The billboards were designed as a non-strategic reminder, given that billboards were interspersed at random around the track and passing them was in no way synchronized with the fuel depletion. In contrast, the timer was designed as a strategic reminder because it was synchronized with fuel depletion. Participants’ use of either or neither strategy was recorded.

**Intelligence task.** All participants were administered the vocabulary and matrices sub-tests of the KBIT last. The vocabulary sub-test measures crystallized intelligence while the matrices sub-test measures fluid intelligence. Raw scores were computed into standard scores in accordance with the participant’s age (mean of 100 and standard deviation of 15 in the standardization sample). The test-retest stability of the K-BIT is very good and evidences criterion-related validity with full battery measures of intelligence (Kaufman & Kaufman, 1990).

**Confidence in time estimation abilities.** *Prospective confidence in time estimation:* Prior to completing the time estimation tasks, participants were asked, “How good are you at guessing time? Like do you think you know exactly how long 1 minute is? Or do you think you might not be so good at guessing how long 1 minute is? Show me on here.” Participants were then shown a Likert-type five-point scale and asked to point to the smiley/frowny face that they are most similar to, “This guy (score of 4) knows exactly how long 1 minute is and gets it right every time. This guy (middle score
= 2) is just OK- sometimes he is off by 10 seconds or so. This guy (score of 0) has no idea how long one minute is - he never gets it right. Which one are you like?”

Retrospective confidence in time estimation: After completing the time estimation tasks (see below), participants were asked; “Now I just asked you to do four tasks about time. How many do you think you got exactly right? None, only 1 task, 2 tasks, 3 tasks, or all 4 tasks?”

Time estimation tasks. All time estimation tasks were identical to those used by Mackinlay et al. (2009).

Retrospective Estimation: Referring to the KBIT vocabulary sub-task, the experimenter asked, “Remember when I asked you to tell me the name of these pictures [experimenter pointed to KBIT vocabulary sub-task]? How long do you think you did that for? Can you guess?” Participants’ responses were recorded and compared to the actual duration of their vocabulary sub-task to obtain an accuracy score.

Prospective Verbal Estimation Task: The experimenter instructed, “Listen carefully. I’m going to ring this bell once, like this (ring bell), then I’ll pause for a while, and then ring it again (ring bell). I’d like you to guess how long I pause for. Are you ready?” The experimenter then performed the task with a 20-s delay between the first and second ring, before asking, “How many seconds do you think that was?” Participants’ responses were recorded and compared to the actual 20-s duration of the task to obtain an accuracy score.

Prospective Production Unfilled Task: Participants were shown a stop watch and told, “When I say ‘start’ I’d like you to say ‘stop’ when you think it has been 30
seconds.” Participants’ responses were timed and recorded and compared to the actual 30-s duration of the task to obtain an accuracy score.

**Prospective Production Filled Task:** Participants were introduced to the tangled lines task, a task similar to a maze, that requires a participant to find the beginning and end of as many overlapping lines as he or she can by tracing them with his or her finger. After explaining the task, the experimenter stated, “Now I want you to trace these tangled lines, but only for 45 seconds. When I say ‘start’ I’d like to you tell me to ‘stop’ after you think 45 seconds has passed.” Participants’ responses were timed and recorded and compared to the actual 45-s duration of the task to obtain an accuracy score.

**Results**

The rejection level for all analyses was set at .05 unless otherwise indicated, and effect sizes are estimated using partial eta squared. Root mean square errors were calculated and presented as RMSE. For each variable of interest, outliers above three standard deviations of the mean were brought back to the fence to conserve sample size. The percentage of participants for whom these adjustments were made is described within each section, where relevant, but generally was less than 2%.

Inter-experimenter reliability analyses were conducted on all experimenter-coded (as opposed to computer-recorded) variables for a random sampling of participants ($n = 51, 44.34\%$). Two experimenters observed these participants for an entire session and recorded responses independently. Inter-experimenter agreement was very high, between $K = .97$ and 1.0. Percent agreement was between 94% and 100%. Discrepancies were resolved through discussion.
Participants’ performance is described and analyzed for each measure first, followed by their metacognitive estimates of this performance measure, and third, by the analyses of the accuracy of these estimates. Finally, hypotheses tests regarding the relationships between these variables are presented.

**Prospective Memory Performance**

Consistent with findings reported elsewhere (e.g., Kerns, 2000), KBIT standardized scores were not related to prospective memory performance ($r = 0.03$) or frequency of gas-checking ($r = .00$).

**Running out of gas.** On average, participants ($n = 114$) missed 1.58 of the five opportunities to refill the gas tank. As a result of software failure, one kindergarten participant’s prospective memory performance was not recorded (<.01%), and she was excluded from analyses of this variable. No outliers were corrected.

Figure 7 presents the mean number of prospective memory trials in which participants ran out of gas for each age group. A 4 (age: kindergarten, 1st grade, 2nd grade, adult) x 2 (gender) x 2 (metacognitive condition: pre-trial, post-trial) x 2 (block order: PM first, PM last) analysis of variance (ANOVA) with all variables between subjects, produced a significant main effect of age group, $F(3, 82) = 9.98$, $RMSE = 1.45$, $\eta_p^2 = .27$. Least significant difference (LSD) post-hoc comparisons confirmed that kindergarten participants were more likely to run out of gas than 1st grade participants and adults, but did not differ from 2nd grade participants. First-grade children were also more likely to run out gas than adults, but did not differ from 2nd grade children. Second-grade children only differed significantly from adults.
Also significant were a gender x block order interaction, $F(1, 82) = 7.18, \eta_p^2 = .08$, and a three-way gender x age x block order interaction, presented in Table 11, indicating that the differences between males and females in each block order varied as a function of their age, $F(3, 82) = 3.96, \eta_p^2 = .13$. Performance was not affected by whether participants were asked to estimate their abilities before or after racing.

Multiple contrasts examined differences between participants who raced the prospective memory block first or second, as a function of age x gender group. Kindergarten males were less likely to run out of gas when racing the prospective memory block first, compared to kindergarten males who raced this block second, $t(9) = -4.28$. However, this trend was reversed for females in kindergarten, $t(3.32) = 4.16$. First grade males and females did not differ significantly in the number of times they ran out of gas as a function of block order: males, $t(16) < 1, ns$; females, $t(7) = 1.05, ns$. This was also the case for first and second grade participants: males, $t(15) < 1, ns$; females, $t(10) < 1, ns$. Like kindergarten males, adult males appeared slightly less likely to run out of gas when the prospective memory version was raced first, but this difference only approached statistical significance, $t(7) = -2.03, p = .08$. Adult females did not differ in the number of times they ran out of gas, dependent on block order, $t(8.54) = -1.27, ns$.

These findings suggest that racing the prospective memory block first particularly hindered female kindergarten children, relative to males. This trend was also maintained for first grade females, but did not reach statistical significance. Second grade females performed worse than males, regardless of block order.

**Accuracy of prospective memory estimates.** Recall that, prior to racing the prospective memory block, participants in the pre-trial condition were asked to estimate
how many times, out of a hypothetical five opportunities, they would remember to refill the gas tank on time. Participants in the post-trial condition were asked, after racing the prospective memory block, to estimate how many times they had successfully refilled the gas tank. Participants’ estimates were reverse coded (subtracted from five) to be comparable with the actual number of missed refills.

KBIT scores did not predict participants’ estimates in how often they would run out of gas \( (r = -0.06) \), and were excluded from further analyses. On average, participants’ estimates were low \( (M = .67 \text{ out of 5.0, } SD = 1.02) \).

Accuracy scores were calculated by subtracting the actual number of times a participant ran out of gas from his or her estimate. Negative scores therefore indicate a participant underestimated how many times he or she ran out of gas. That is, a low negative score indicates a participant severely overestimated his or her performance on the prospective memory task. Likewise, a high positive score indicates a participant overestimated how many times he or she ran, or would run, out of gas. In other words, the higher the positive score, the more a participant underestimated his or her prospective memory abilities. A score of zero means participants were accurate in their estimates.

As depicted in Figure 8, younger participants were more likely to overestimate their performance (i.e., to underestimate the number of times they would run out of gas) than older participants. One-sample t-tests were conducted for each age group and confirmed that kindergarten, \( t(25) = -4.14 \), 1st grade \( t(26) = -4.75 \), and 2nd grade, \( t(28) = -2.3 \), children all significantly overestimated their performance (compared to a score of 0 = accurate). In contrast, adults’ estimates were not significantly different from zero, \( t(31) = 1.04 \), \( ns \).
A 4-way ANOVA examined whether accuracy scores differed as a function of age, as well as metacognitive condition, gender, and block order. The analysis produced a significant main effect of age, $F(3, 82) = 4.99$, $RMSE = 1.59$, $\eta^2_p = .15$. Kindergarten participants overestimated their prospective memory abilities more than 2nd grade and adult participants, but not 1st grade children. First graders did not differ from 2nd graders, but significantly overestimated their abilities relative to adults. Second graders also overestimated their abilities relative to adults.

There was also a significant 2-way interaction between block order and gender, $F(1, 82) = 9.41$, $\eta^2_p = .10$, and, as depicted in Table 12, a significant 3-way interaction between age, block order, and gender, $F(3, 82) = 2.88$, $\eta^2_p = .10$.

Post-hoc contrasts examined gender differences within each age x block order group. Only kindergarten males and females differed when racing the prospective memory block first; kindergarten females significantly overestimated their abilities when compared to kindergarten males, $t(12) = 3.32$. No other contrasts reached significance. Post-hoc contrasts also compared block order effects within each gender x age group. Kindergarten males were more likely to overestimate their prospective memory abilities when racing the prospective memory block second rather than first, $t(11) = 2.59$. In contrast, kindergarten females did not significantly differ across block order,

$t(11) = -1.58$, ns. No other contrasts reached significance. These findings indicate that kindergarten males’ estimates were affected by racing the prospective memory block first, moreso than were estimates of other age and gender groups; racing the prospective memory block first caused these males to underestimate their prospective memory abilities.
**Frequency of gas checks during first minute of play.** Participants checked the gas a mean of 6.38 times within the first minute of racing (SD = 4.33; no outliers were observed or corrected). Participants who successfully refilled the tank on the first trial checked the gas an average of 8.22 times (n = 77, SD = 3.83) during the waiting period. Those who did not refill the tank on the first trial checked the gas only 2.66 times, on average (n = 38, SD = 2.52), t(103.79) = 9.31.

As depicted in Figure 9, and contrary to the findings of Mäntylä and colleagues (2007), kindergarten participants checked the gas less frequently during the first minute than older children, and especially adults. Frequency of checking was submitted to a 4 (age) x 2 (gender) x 2 (metacognitive condition) x 2 (block order) ANOVA, and confirmed the main effect of age group, $F(3, 83) = 3.13$, $RSME = 3.81$, $\eta_p^2 = .10$. There was also a main effect of gender; males checked the gas more frequently ($M = 7.28$, $SD = 4.60$) than females ($M = 5.25$, $SD = 4.70$), $F(1, 83) = 4.20$, $\eta_p^2 = .05$. No other main effects or interactions were significant.

Post hoc LSD multiple comparisons conducted to follow-up the main effect of age confirmed that kindergarten and 2nd grade children checked the gas significantly less often than adults. First grade children were not significantly different from adults. No differences between kindergarten, 1st, and 2nd grade children were observed.

**Relationship between monitoring and accuracy of prospective memory estimates.** Multilevel linear models assessed the relationship between participants’ accuracy in estimating their prospective memory performance and their monitoring (i.e., frequency of fuel checking). As a first step, accuracy of prospective memory estimation scores were entered as a predictor in an ANOVA with frequency of gas checks as the
dependent variable, ignoring the structure of the data. Participants’ accuracy in estimating their own abilities significantly predicted their frequency of gas checking, $b = 1.25$, $SE = .19$, CI[.87-1.62], $F(1, 114) = 44.20$.

Age, block order, gender, and metacognitive condition were examined as second-level subject variables, both independently, and in conjunction. Hurvich and Tsai’s criterion (AICC) was used to assess whether the fit of the model improved at each step in this and all subsequent multilevel models. This criterion corrects for model complexity much like Akaike’s information criteria (AIC), but is appropriate for small sample sizes.

Allowing the intercept and slope of the relationship to vary by age, block order, or metacognitive condition did not improve the fit of the model, nor did including KBIT scores as a covariate. Accuracy was not significant in predicting monitoring (i.e., gas-checking frequency) as a quadratic term, nor did it improve the fit of the model, indicating that the direction of inaccuracy (whether performance was under- or overestimated) influenced monitoring.

As depicted in Figure 10, these findings confirmed that participants who overestimated their prospective memory abilities engaged in less frequent gas checking compared to participants who were accurate. Participants who underestimated their prospective memory performance engaged in more frequent gas checking than those who were accurate.

**Pattern of checking.** An examination of the pattern of gas checks across the length of the first-trial waiting period was conducted by tallying all the gas checks within each 5-second interval. Previous work demonstrated that strategic checking patterns during the first trial of a time-based prospective memory task follow a U-shaped pattern,
whereas when subsequent trials are taken into account, the mean frequency across all trials yields a J-shaped pattern. This is likely due to the decreased need for participants to calibrate themselves after the initial trial (cf. Ceci & Bronfenbrenner, 1983; Kerns, 2000). As such, first-trial checking behaviors were analyzed here. In addition, time intervals became less accurate after the first trial because some participants refilled early in the 5-s refill period, others late, and others not at all, causing variation in the time at which the second trial was started (26-s to 38-s).

Figure 11 displays the slightly positive, linear distribution of clock checking during the waiting period. Overall, participants tended to monitor more frequently during time periods that were closer to the prospective memory target times. These data were analyzed by means of a 5-way mixed design (MANOVA: age x gender x block order x metacognition presentation x time interval), with the first four factors examined between subjects and the final factor, each 5-s interval, analyzed as a repeated measure within subjects. A significant main effect of 5-s interval was revealed, $F(4, 80) = 8.26$, $\eta^2_p = .29$. No other significant main effects or interactions were observed. Within-subjects polynomial contrasts revealed the linear trend was significant, $F(1, 93) = 22.80$, $RMSE = .77$, $\eta^2_p = .22$, but not the quadratic term, $F(1, 93) < 1$, $ns$.

This trend was followed up using multilevel modeling techniques. The linear relationship between time and checking frequency showed significant variance in intercepts across participants, $\text{var}(u_{0j}) = .21$, $\chi^2(1) = 95.69$, but the slope of the linear trend remained the same. This model revealed a significant, positive linear trend between checking frequency and time, $b = .11$, $SE = .02$, $F(1, 460) = 37.87$. This pattern is inconsistent with the strategic patterns described elsewhere (Ceci & Bronfenbrenner,
1983; Harris & Wilkins, 1978; Kerns, 2000), and suggests that participants were not engaging in a TWTE strategy during the CyberCruiser 2.0.

**Ongoing Task Performance**

**Lap time.** To increase computational precision of the costs of prospective memory, for each individual I took into account the presentation procedure (metacognitive condition and block order), following a method used by Einstein, McDaniel, Thomas, Mayfield, Shank, Morissette, and Breneiser (2005). Due to software failure, one kindergarten participant’s lap time was not recorded in the control or prospective memory conditions, and another first grader and kindergartener did not have decipherable lap times in the prospective memory condition only (3.5%). No outliers were corrected.

I identified participants who showed an increased lap time for the prospective memory versus control block and those who did not show an increased lap time. Consistent with Einstein et al. (2005), and the multiprocess framework, this analysis clearly revealed that there were substantial numbers of participants displaying both patterns. Eighty-nine participants (77.4%) had increased latencies for the prospective memory block, whereas 26 (22.6%) did not.

A possible issue in interpreting these individual differences is that the tendency to display increased versus decreased latencies on the prospective memory relative to

---

2 I computed the average speed-up across blocks one and two when each task version (ongoing task block and prospective memory block) appeared first for each metacognitive condition. For a particular participant, the average value (speed-up) obtained for her or his metacognitive condition, block order, and age group was added to his or her latency for the second block of trials. Therefore, for every participant the difference between responses for prospective memory and control blocks took into account the artifact of speed-up across blocks and the fact that this speed-up varied across counterbalancing conditions.
control block might be linked to the order in which the two blocks were presented. There was, however, no association between whether participants displayed a cost in response latency for the prospective memory block and which block was presented first, \( \chi^2(1, N = 112) = .01, \ ns. \)

Having ruled out the effects of block order on cost- versus no-cost participants, adjusted lap times were entered as the dependent variable in a block (control v. prospective memory) x age x gender x metacognitive condition mixed model ANOVA. Block served as the within subjects variable. Age, gender, and metacognition presentation were analyzed between subjects.

On average, participants were faster in the control block \( (M = 91.62s, SD = 3.17) \) than in the prospective memory block \( (M = 106.07s, SD = 3.94) \), \( F(1, 96) = 52.99, \ RMSE = 1.48, \eta_p^2 = .36 \). The main effect of age was also significant, \( F(3, 96) = 37.22, \ RMSE = 3.28, \eta_p^2 = .22 \), as was the main effect of gender, \( F(1, 96) = 22.35, \eta_p^2 = .19 \). The 2-way interaction between block and age was significant, \( F(3, 96) = 3.55, \eta_p^2 = .10 \), as was the interaction between block, age, and gender, \( F(3, 96) = 7.84, \eta_p^2 = .20 \), presented in Figure 12.

Post hoc comparisons were used to follow-up this interaction by examining differences in lap times for the prospective memory and control blocks within participants within each age x gender group. As depicted in Figure 12, on average, males in kindergarten suffered a prospective memory cost; their lap times were significantly faster in the control block than they were in the prospective memory block, \( t(12) = 4.07 \). The difference between blocks for males in 1st grade was not significant, \( t(16) = 1.98 \). Second-grade males also suffered a prospective memory cost; they were significantly
faster in the control block relative to the prospective memory block, $t(16) = 2.84$. In contrast, adult males’ did not incur a cost; lap times did not differ across block, $t(15) = 1.52$, ns. Unlike males in kindergarten, female kindergarteners did not appear to incur a prospective memory cost; they were no faster in the control than prospective memory block, $t(11) = .12$, ns. First-grade females did incur a cost, however, $t(8) = 3.07$, as did 2nd grade females, $t(11) = 8.41$. Like male adults, female adults did not incur a prospective memory cost, $t(15) = -2.16$, ns.

**Crashes.** Crash rates were adjusted in the same manner as lap times to control for block order effects. Participants with crash rates above three standard deviations of the mean were brought back to the fence (prospective memory block = 2.61%, control block = 3.48%). An error in the software resulted in a failure to record crash rates for one participant in the control block and one participant in the prospective memory block. These data were eliminated pairwise from analyses. As such, for within-subjects comparisons, sample size is reduced by two. Fifty-six (49.6%) participants experienced more crashes in the prospective memory block than the control block, whereas 57 (50.4%) did not. This effect was not related to block order, $\chi^2(1, N = 113) = .22$, ns.

Adjusted crash rates were examined using the same 4 x 2 x 2 x 2 mixed design ANOVA used to assess lap time. Only the main effect of age was significant, $F(3, 97) = 24.46$, $RMSE = 6.10$, $\eta_p^2 = .44$. LSD pairwise comparisons revealed that kindergarten participants ($M = 12.25$, $SE = .86$, CI[10.54-13.95]) crashed more than 1st graders ($M = 8.35$, $SE = .92$, CI[6.53-10.17]), 2nd graders ($M = 6.28$, $SE = .84$, CI[4.62-7.94]), and adults ($M = 2.41$, $SE = .76$, CI[.90-3.93]). Adults also crashed significantly less than 1st and 2nd grade children, but the difference between 1st and 2nd grade children was not
significant. While no other significant main effects were produced, an interaction between metacognition presentation and block was revealed. These means are presented in Table 13, $F(1, 97) = 9.10$, $RMSE = 5.92$, $\eta_p^2 = .09$.

Post-hoc tests confirmed that the difference between crash frequency did not differ from the prospective memory to the control block among participants who were not asked to estimate their performance prior to racing (post-trial condition), $t(58) = 2.01$, ns. However, when participants were asked to estimate their performance prior to racing (pre-trial condition), they crashed significantly more often in the prospective memory block relative to the control block, $t(53) = -3.00$. These findings suggest that asking participants to estimate their performance prior to racing may have caused them to become more anxious in the prospective memory block, increasing crash rates relative to the control block.

Crash scores served as a proxy for accuracy and were used to examine the possibility of a speed-accuracy tradeoff, adopting a method used by Einstein et al. (2005). A 2 x 2 mixed ANOVA with lap time cost group ($n = 89$ with increased lap times, $26$ with decreased or equal lap times) as the between-subjects variable and the presence of the prospective memory task (block) as the within-subjects variable confirmed that the groups did not differ in the number of crashes, $F(1,111) < 1$, $RMSE = 7.82$, ns, and there was no interaction or effect of block, $F(1,111) < 1$, $RMSE = 6.19$.

**Estimated cost of prospective memory block.** Recall that participants were asked to estimate whether the task of monitoring and refueling the gas would cause them to race a faster or slower lap time. Participants were also asked to estimate by how many seconds their lap times would differ.
When asked to estimate whether refilling the gas tank would cause a faster or slower lap time, 57% \((n = 65)\) of participants stated “slower,” 41.2% \((n = 47)\) stated “faster,” and 1.8% \((n = 2)\) stated “the same.” Table 14 indicates that children were more likely than adults to report they would achieve a “faster” lap time when required to monitor and refuel the gas tank, \(\chi^2(6, N = 114) = 23.27\).

However, the direction of a participants’ estimated cost was also influenced by whether he or she raced the prospective memory block first or second, \(\chi^2(2, N = 114) = 29.08\). When the relationship between block order and estimated costs was examined in each age group, it was revealed that all three age groups of children were influenced by block order when making these estimates. The majority of those who estimated the prospective memory task would cause them to have a faster lap time did so when racing the prospective memory block second; Kindergarten, \(\chi^2(1, N = 26) = 18.96\); 1\textsuperscript{st} grade, \(\chi^2(1, N = 27) = 12.49\); 2\textsuperscript{nd} grade, \(\chi^2(1, N = 29) = 12.46\). Block order was not related to estimated costs for adults, \(\chi^2(1, N = 32) = 2.04, ns\). These responses indicate that children in this condition may have been confused by the question and made estimations based on perceived practice effects. For this reason, it is also difficult to conclude that children who received the prospective memory block first attributed the cost to the prospective memory task rather than practice. As such, children’s responses are disregarded from further analysis of cost estimates. For these reasons, only adults’ estimates of the prospective memory costs can be analyzed further.

On average, adults estimated that the prospective memory task would add 8.71 seconds to their lap time \((SD = 19.52)\). Adults’ estimates were not significantly related to
gender, block order, or metacognitive condition, \( F(1, 24) < 1, RMSE = 19.94, \text{ns} \). Adults’ estimates were not related to KBIT scores, \( r = .08, n = 32, \text{ns} \).

**Accuracy of cost estimates.** Accuracy scores were calculated in the following steps: First, a participant’s adjusted lap time on the ongoing task was subtracted from his or her adjusted lap time on the prospective memory task, yielding a measure of the actual cost of prospective memory (positive scores indicated the time by which lap time increased, whereas negative scores indicated the time by which lap time decreased); next, the actual cost of prospective memory was subtracted from the amount of time participants’ estimated (as a continuous measure). Positive scores indicate the amount participants’ overestimated, whereas negative scores indicate the amount participants’ underestimated the costs of the prospective memory task on lap times.

The mean of adults’ accuracy scores was 4.26 s (\( \text{SD} = 18.5 \) s), indicating that adults overestimated the costs of prospective memory, on average, although this estimate was not significantly different from zero, \( t(31) = 1.30, \text{ns} \). Because children’s estimations were confounded by block order, only adults’ accuracy scores were submitted to a 3-way ANOVA, examining the effects of block order, metacognitive condition, and gender as between-subjects variables. No significant main effects or interactions were produced.

**Relationship between prospective memory performance and adults’ accuracy estimating prospective memory costs.** Adults’ accuracy in estimating the costs incurred by the prospective memory task was first used to predict prospective memory performance. The intercepts of the relationship between accuracy in estimating costs and prospective memory performance varied significantly by block order and gender, \( \text{var} (u_{0j}) = .06 \), and revealed a small but significant linear relationship between accuracy and
performance, after controlling for KBIT scores, $b = -0.01, SE = 0.00, F(1, 30.96) = 8.37$. Including cost estimate accuracy as a quadratic term did not improve the model. This final model was significant, $\chi^2(2) = 8.28$, and Figure 13 presents the collapsed partial correlation. As adults moved from underestimating the costs of the prospective memory task to accurately estimating the costs, through overestimating these costs, they became less likely to run out of gas in the prospective memory task.

**Relationship between monitoring and adults’ accuracy estimating**

**Prospective memory costs.** Adults’ accuracy in estimating the costs incurred by the prospective memory task was also used to predict their frequency of gas checking. An ANOVA failed to reveal a significant relationship between the two variables, $F(1, 32) < 1, ns$. Allowing the intercept to vary by gender, metacognitive condition, and/or block order did not improve the model, nor did covarying KBIT scores. To assess whether the relationship between checking frequency and accuracy might be quadratic, the squared accuracy term was entered into the model. Doing so failed to improve the model or reveal a relationship between accuracy and gas checking, $F(1, 32) < 1, ns$.

**Use of external reminders**

Recall that during the third and final round of racing, participants were presented with the option of adopting either one of two external strategies; participants were shown how to use a timer that could be set to go off when the gas was ready for refill, eliminating the need to check the gas level while racing (a strategic reminder), and how to use the billboards as a visual cue to check the gas (a non-strategic reminder, given that billboards were interspersed at random around the track and passing them was in no way synchronized with the fuel depletion). Table 15 presents the percentage of participants
within each age group who chose either strategy or who did not use a strategy in the final round. Overall, participants were more likely to use the timer than the billboard strategy or no external strategy at all. A greater proportion of kindergarten children did not use an external strategy compared to other age groups. Adults were, surprisingly, evenly disbursed in their use of the billboard, timer, or neither strategy, whereas the majority of 1st and 2nd grade children used the timer strategy.

Males and females differed in their use of external strategies during the final round. As depicted in Table 16, males were more likely to use the timer compared to females, whereas females were more likely to use the billboard than males. There were no interactions between gender and age group; males did not differ in their use of strategies by age, $\chi^2(6, N = 64) = 7.77, ns$, nor did females, $\chi^2(6, N = 51) = 10.17, ns$.

To determine whether participants benefited from reducing the time-based task to an event-based task (using the timer or billboards vs. neither), prospective memory performance (number of successful refills) was compared across the prospective memory and external reminder blocks, as a function of strategy. Missed refill opportunities were entered as a dependent variable in a 5-way mixed MANOVA where block (prospective memory or external reminder) was a within-subjects variables, and strategy, gender, age, metacognitive condition, and block order were compared between subjects. As mentioned above, one kindergarten female’s prospective memory performance was not recorded due to software error, and was excluded from the analysis. It was noted that she did not use either strategy during the external reminder block, however. The effect of block was not significant, $F(1, 49) = 1.27, ns$, indicating participants did not differ in the number of times they ran out of gas by block, nor did block interact with strategy, $F(2, 49) = 1.20,$
There was a significant 3-way interaction between block, strategy, and age, $F(6, 49) = 5.08$, $RMSE = .86$, $\eta_p^2 = .38$.

The block x strategy x age interaction was examined using post-hoc contrasts. Prospective memory performance in each block was compared across each strategy x age group. These means are presented in Table 17. No contrasts reached significance, but trends are interpreted at $p < .10$. Kindergarten participants’ ability to refill the tank was unaffected by whether they adopted the timer, $t(12) < 1$, $ns$, or billboard, $t(2) < 1$, $ns$, but those who chose to adopt neither strategy tended to run out of gas more often in the external reminder block relative to the prospective memory block, $t(9) = 2.14$, $p = .06$. First graders who used the timer were slightly more likely to run out of gas in the external reminder condition, $t(16) = 1.78$, $p = .10$, but did not differ when using the billboard, $t(6) < 1$, $ns$, or neither strategy, $t(2) = 1.15$, $ns$. Second grade children did not differ in the number of times they ran out of gas as a function of adopting the timer strategy, $t(16) = 1.25$, $ns$, billboard strategy, $t(3) = -1.22$, $ns$, or neither strategy, $t(7) < 1$, $ns$. This was also the case for adults, $ts < 1$.

Next, I examined whether adopting the timer strategy would influence participants’ ongoing task performance, that is, their lap times in the external reminder, relative to prospective memory, block. Lap times were entered into a 4 (age) x 3 (strategy) x 2 (gender) x 2 (metacognitive condition) x 2 (block) mixed MANOVA, with block as the repeated measure. As with prospective memory performance, lap times did not differ by block, $F(1, 72) < 1$, $RMSE = 22.29s$, $ns$. No other main effects or interactions were produced.
The finding that participants did not seem to benefit from using the timer was surprising. If participants actually relied on the strategies, their frequency of gas checking should be reduced in the external reminder, relative to the prospective memory block. To examine this hypothesis, gas-checking frequency was compared across the prospective memory and external reminder blocks, as a function of strategy use. The analysis produced a significant interaction, $F(2, 87) = 13.91$, $RMSE = 3.72$, $\eta_p^2 = .24$. Post-hoc paired samples t-tests confirmed that participants who chose the billboard strategy seemed to rely on this strategy, checking the gas more frequently in the external reminder block ($M = 10.38$, $SD = 9.29$) relative to the prospective memory block ($M = 5.69$, $SD = 4.84$), $t(15) = -2.41$. In contrast, participants who opted to use the timer checked the gas significantly less often in the external reminder block ($M = 3.25$, $SD = 3.33$) than in the prospective memory block ($M = 6.51$, $SD = 4.39$), $t(50) = 50.03$. Participants who chose to use neither strategy did not differ in their frequency of gas-checking across blocks (external reminder, $M = 3.46$, $SD = 4.04$; prospective memory, $M = 4.48$, $SD = 3.31$), $t(22) < 1$, ns.

**Relationship between strategy-use and adults’ understanding of the costs of prospective memory.** I predicted that participants who understood the costs of prospective memory on ongoing task performance would adopt strategies to reduce these costs. Because only adults’ estimates of these costs were reliable, I examined the interaction between the costs of prospective memory on lap time (how much faster or slower) and their accuracy in estimating these costs as a predictor of their use of the timer in the external reminder condition. Binary logistic regression analysis did not produce a significant effect of the interaction, $b = 0.00$, SE = .00, Wald $\chi^2 (1) = .12$, ns. Only KBIT
scores predicted adults’ use of the timer, $b = -.06$, $SE = .03$, $R^2 = .13$ (Cox & Snell), .18 (Nagelkerke), Model $\chi^2 (1) = 4.52$.

**Motivation**

Participants’ reported the importance of driving and filling the gas tank on a sliding Likert-type scale, ranging from 1 = “not important at all” to 7 = “very important.” To confirm that children understood the scale, that is, that their ratings on the scale were consistent with whether driving was considered more or less important than filling the gas tank, children were also asked to point to the token that was most important. All children’s responses were consistent with their ratings on the Likert scale; those who rated the driving task as more important on the scale also selected the car token, representing the driving task, as most important, whereas those who rated the prospective memory task as more important on the Likert scale unanimously selected the gas token as most important.

Categorically speaking, 17 participants (14.8%) rated the ongoing task (a fast lap time) as more important than the prospective memory task (refueling). Thirty participants (26.1%) rated the ongoing and prospective memory tasks as equally important. The majority ($n = 68, 59.1\%$) rated the prospective memory task as more important.

**Relationship between motivation and prospective memory performance.** A single score was calculated to represent each participant’s “relative prospective memory motivation” by subtracting his or her Likert rating of the ongoing task from his or her Likert rating of the prospective memory task. In this way, negative scores indicate the ongoing task was more important to the participant; conversely, positive scores indicate the prospective task was more important.
These scores were used as a predictor of prospective memory performance in multilevel linear models. The best-fitting model, $\chi^2(2) = 18.85$, allowed the intercepts of the relationship to vary by age, and covaried K-BIT scores, but still failed to produce a significant effect of motivation, $F(1, 111.16) < 1$, ns. No relationship between motivation and frequency of checking or external strategy use was observed either.

**Time Estimation Tasks**

Time estimation scores were calculated for each participant for each of the four time estimation tasks by dividing the participants’ estimated time by the actual time, resulting in a score representing the percent over/under estimated. Across the four tasks, 3.47% of participants were outliers brought back to the fence. Percent deviation was then averaged across the four time estimation tasks for each participant. The session was ended for one 1st grade participant before the time estimation tasks could be completed. On average, participants overestimated time durations by 134% ($SD = .74$). One-sample t-tests confirmed that mean time estimation was significantly different from 0 = accurate for all age groups. Time estimation abilities did not significantly differ by age, $F(3, 106) < 1$, ns, gender, $F(1, 106) < 1$, ns, or the interaction of age and gender, $F(3, 106) < 1$, ns, $RMSE = .74$.

**Time estimation abilities and prospective memory performance.** To replicate Mackinlay et al.’s (2009) test of the hypothesis that individual differences in time estimation abilities predict prospective memory performance, a participant’s time estimation scores were used to predict the number of times he or she ran out of gas and their frequency of gas checking. In contrast to Mackinlay’s findings, time estimation predicted prospective memory performance, after covarying KBIT scores, $b = .40$, $SE = .108$.
.19, $F(1, 110.87) = 5.01$. The intercept of this relationship varied significantly by age, $\text{var}(\mu_{ij}) = .86$, $\chi^2(2) = 16.65$. These findings indicate that as participants move from underestimating the amount of time that has passed, through accurately estimating time, to overestimating the amount of time that has passed, they become more likely to remember to refill the gas tank.

**Time estimation abilities and gas-checking frequency.** Time estimation predicted gas checking frequency, $b = -1.43, SE = .53, F(1, 113) = 7.31$. This negative linear trend indicated that as participants moved from underestimating time intervals through accurate estimation, to overestimating time intervals, they check the gas less frequently. That is, those who thought time passed faster than reality tended to check the gas more often, whereas those who felt time moved more slowly checked the gas less frequently.

**Metacognitive understanding of time estimation.** Recall that prior to the time estimation tasks, all participants were asked to indicate how good they thought they were at “guessing time” on a visual scale ranging from 0 = “not good at all” to 4 = “very good.” Following the trials, participants were asked to indicate how many tasks they thought they had gotten exactly right (from 0 to 4). Confidence scores were reverse-coded and then divided by five to yield a proportion ranging from 0 to 1, where low scores represented higher confidence ratings. Participants’ actual mean time estimation scores were centered around zero, such that 0 = accurate time estimation abilities, and the further a score moves away from zero, the less accurate he or she was in estimating time. Participants’ metacognitive estimates of his or her own time estimation abilities were then calculated by subtracting a participants’ metacognitive estimate (reversed
confidence rating, proportion out of five) from his or her actual mean time estimation score (centered around zero).

As such, negative scores indicate the participant was overconfident in his or her time estimation abilities (i.e., performed worse than he or she thought) and positive scores indicated the participant was under-confident in his or her time estimation abilities (performed better than he or she thought).

These accuracy scores were entered as a dependent variable in a mixed ANOVA where age (children versus adults) and gender were between-subjects factors and “time of estimate” (pre-trial or post-trial) was the repeated measure. The main effect of “time of estimate” was significant; participants underestimated their time estimation abilities more at post-test ($M = .07, SD = .76$) than at pre-test ($M = -.04, SD = .72$), $F(1, 106) = 8.38$, $RMSE = .21, \eta^2_p = .07$. There was also a main effect of age. Kindergarten ($M = -.33, SE = .13, 95\% CI [-.59, -.07])$, 1st grade ($M = -.13, SE = .14, 95\% CI [-.40, .15])$, and 2nd grade ($M = -.01, SE = .13, 95\% CI [-.26, .25])$ children were significantly more accurate than adults ($M = .46, SE = .12, 95\% CI [.22, .70])$, $F(3, 106) = 7.32, \eta^2_p = .17$. The block x age interaction was significant, and is presented in Table 18, $F(3, 106) = 6.64, \eta^2_p = .16$.

One-sample t-tests indicated that kindergarten participants’ accuracy was not significantly different from $0 = \text{accurate at pre-trial}$, $t(26) = -1.93$, $ns$, but that kindergarten children grew overconfident by post-trial, $t(26) = -2.29$. First grade children were not significantly different from accuracy at pre- or post-trial, $t(25) < 1$, $ns$, nor were 2nd grade children, $t(28) < 0, ns$. In contrast, adults significantly underestimated their abilities at both pre- and post-trial, $t(31) = 8.82$ and 12.80, respectively.
Next, I evaluated my hypotheses that participants’ metacognitive representations of their own “sense of time” would predict their frequency of gas checking and their use of external strategies in the third block.

**Metacognitive understanding of time estimation and monitoring.** I predicted that participants who were not very good at estimating time, and who understood these personal limitations, would spend more time monitoring the fuel gauge. To test this hypothesis, I entered participants’ checking frequency as a dependent measure in a multilevel linear model with the interaction between metacognitive understanding of time estimation and actual time estimation abilities as a continuous predictor and trial as a repeated factor. The interaction significantly predicted gas-checking frequency, $b = .32$, $SE = .13$, $F(1, 227.98) = 9.41$, and the intercepts varied significantly by age, gender, block order, and metacognitive condition, var($u_{0j}$) = 5.92, $\chi^2(1) = 50.57$. When participants metacognitively assessed their time estimation abilities (pre- versus post-trial) did not interact with their metacognitive understanding or time estimation abilities to predict monitoring, $F(1, 198.48) < 1$, ns.

To break down this interaction, time estimation abilities were split at the median (1.14) to reflect two groups; one of participants with relatively good time estimation abilities and another of participants with relatively poor time estimation abilities. The relationship between metacognitive understanding of time estimation and monitoring was then examined within participants who were good at estimating time and those who were not. The models specified were the same as above. As predicted, participants who better understood they were poor at estimating time checked the fuel gauge more frequently, $b = 1.34$, $t(113.93) = 3.12$. This was not the case for participants who were relatively good
at estimating time; that is, their understanding of how good they were at estimating time did not predict their gas-checking frequency, $b = .28, t(110.35) < 1, ns.$

**Metacognitive understanding of time estimation and use of an external reminder.** I predicted that participants who were not very good at estimating time, and who understood these personal limitations, would be more likely to adopt the timer as an external strategy, as a compensatory measure for their poor sense of time. To test this hypothesis, I entered participants’ use of the timer as a dependent measure in a multilevel linear model with the interaction between metacognitive understanding of time estimation and actual time estimation abilities as a continuous predictor and trial as a repeated factor. The interaction was not significant in predicting use of the timer, however, $F(1, 225.52) < 1, ns. When$ participants metacognitively assessed their time estimation abilities (pre- versus post-trial) did not interact with their metacognitive understanding or time estimation abilities to predict timer use, $F(1, 218.13) < 1, ns.$ Allowing the slope and intercepts to vary by age did not improve the model.

**Discussion**

Research on prospective memory development has mainly focused on event-based prospective memory. Few studies have specifically examined the processes involved in remembering time-based tasks. This study fills this gap by examining the nature and development of thought processes that support success on time-based prospective memory tasks, in both young children and adults.

**Successful refills**

Children performed worse on the time-based prospective memory task, the CyberCruiser 2.0, failing to refill the gas tank more times than adults. This age effect is
consistent with others (e.g., Kerns, 2000; Rendell, Vella, Kliegel, & Terrett, 2009) providing further support that time-based prospective memory performance improves with age.

Couched within this effect was a gender x block interaction, indicating that younger female children performed worse than male children, particularly when asked to race the prospective memory task first (although 2\textsuperscript{nd} grade females ran out of gas more often than 2\textsuperscript{nd} grade males in both blocks). In contrast, Study 1 showed that female children outperformed male children on event-based prospective memory tasks. It is possible that males have more experience with CyberCruiser-type games than females in early elementary school. As a result, younger females may have felt more anxious or uncertain about performing the task, particularly when the ongoing task was coupled with the prospective memory task in the first block, as reflect by their poorer performance than male children in this condition. By adulthood, the male advantage may disappear as females have gained more experience with videogames.

These findings are somewhat consistent with Ceci and Bronfennbrenner’s (1985) who reported sex differences in the performance of adolescents on laboratory-based prospective memory tasks. However, in Ceci and Bronfennbrenner’s study, it was older boys (14-year-olds) who performed more anxiously on male sex-typed tasks in the laboratory than younger boys (10-year-olds). Ceci and Bronfenbrenner did not report age x gender x task interactions for participants’ actual success on prospective memory tasks, but only on their patterns of clock checking. However, taken together these patterns reveal that there are important interactions between sex, the nature of the task, and the task setting that impact children’s performance.
Monitoring

In the current study, kindergarten and second grade children checked the gas less often than adults. This pattern of results run counter to those reported by Kerns (2000), who failed to find a relationship between frequency of gas-checking and age among 7- to 12-year-olds, and Mäntylä and colleagues (2007), who found that children checked the clock more often than adults. However, the younger age group tested here (5- to 7-year-olds) may account for these differences. Moreover, females checked the gas less frequently than males. Therefore, although it is still possible that female children performed worse than male children because they were more anxious or uncertain about the task, they did not seem to adopt a checking-strategy that would reflect a pattern of anxious monitoring. Rather, it seems they performed more poorly on the prospective memory task because they spent less time monitoring the fuel gauge and were more likely to miss the refill opportunity as a result.

A participant’s accuracy when estimating how likely he or she was to run out of gas positively predicted his or her frequency of fuel checking. Participants who did not appropriately estimate the difficulty of the task checked the gas less frequently than participants who more accurately estimated their abilities. By comparison, participants who perceived their prospective memory abilities to be relatively poor engaged in more anxious monitoring, checking the gas more frequently. This relationship did not vary by age or gender, indicating that young females’ poorer performance on the task and infrequent monitoring was partially mediated by their failure to assess that the task would be difficult for them rather than to anxiety about their performance.

Strategic monitoring?
Analyses of participants’ checking patterns across the delay interval revealed a largely linear function. Frequency of checking increased as the opportunity to refill the gas tank approached. This pattern is inconsistent with that reported by Kerns (2000), who measured checking behaviors in 15-second intervals across a 1-minute waiting period and observed a significant J-shaped pattern across all age groups (see Kerns, 2000, p. 67; see also Ceci & Bronfenbrenner, 1983; Harris & Wilkins, 1978). However, Ceci and Bronfenbrenner reported that checking patterns were more likely to take on a positive linear trend, like that observed here, when time-based tasks were conducted in a lab versus home context (1983, p. 156-157). The lab-based context of the current study may be one reason the linear pattern is more prominent here.

A second reason may be the relatively short trial period in the current study; recall that Kerns’ trial periods were 1-minute, while Ceci and Bronfenbrenner (1983) examined checking patterns in successive 5-minute intervals over a 30-minute waiting period. A third reason may be the relative uncertainty of the trial period in the present study; participants were not advised how long the waiting period would be before the gas ran out, whereas in Ceci and Bronfenbrenner’s study, participants knew they had to take the cookies out of the oven, for example, after precisely 30-minutes. Moreover, Ceci et al. (1988) found that when the clock speed was manipulated by 50% (faster or slower), 10-year-olds resorted to a positive linear pattern of clock checking. Similarly, in this case, participants may have been inclined to adopt a more “anxious” linear pattern of checking because they did not know, initially, how long before the tank ran out of gas or the rate at which the gas gauge would deplete.
Therefore, these findings generate two complementary hypotheses regarding the patterns observed in the present study: (1) that the linear pattern of checking is *strategic* in time-based prospective memory tasks with relatively shorter waiting intervals, or (2) that the linear pattern is driven by anxious-monitoring when task conditions are uncertain and is not strategic.

To test these hypotheses, I computed the correlation between the amount of time spent by each participant in clock checking and the linear distribution of his or her gas-checking pattern, as measured by his or her linear regression coefficient (see Ceci & Bronfenbrenner, 1983, for a similar method). The resulting correlation indicated there was no significant relationship between linear checking and frequency of checking, \( r = - .15, \text{ns} \). Participants’ linear coefficients did not predict prospective memory performance either, \( r = .05, \text{ns} \). These findings suggest that participants were not engaging in a strategic pattern of checking during the prospective memory task, possibly because of the ambiguous, yet rapid delay interval.

Consistent replication of the TWTE pattern across multiple age groups (see also Einstein et al., 1995; Kvavilishvili & Fisher, 2007; Park et al., 1997) has led researchers to suggest strategic monitoring is automatic (Kerns, 2000; Smith, 2010), and that time-based tasks depend on periodic rehearsal of intention during a period preceding the critical time. However, others have argued that participants are engaging in deliberate and self-initiated monitoring (Kvavilashvili & Fisher, 2007) because there is no external signal that the time for task-execution has arrived. Also, laboratory studies have mostly short time delays, throughout which participants are likely to maintain conscious awareness of the prospective memory task. In this vein, the linear pattern observed here
might indicate participants’ thoughts were not allowed the opportunity to “wonder elsewhere” given the short delay interval. Rather, participants may have been actively engaged in monitoring their intention to refill the tank throughout the entire delay interval. As a result, there was no dip, or “hook” in their patterns. Therefore, the linear pattern observed in the current study provides strong support that the TWTE pattern is not automatic, is affected by experimental conditions, and is the result of deliberate monitoring.

If it is the case that the prospective memory task never leaves conscious awareness under laboratory circumstances where participants are required to carry out a task several times with short delay intervals (1 or 2 min), then it becomes difficult to determine whether prospective memory is being measured, or just a participant’s ability to measure the length of elapsed time (see also, Graf & Uttl, 2001; Kvavilashvili & Fisher, 2007; Park et al., 1997; Sellen et al., 1997). Therefore, one limitation of the present study is the short delay interval (30-sec). The delay interval was shortened from that used by Kerns, primarily to allow for an increased number of prospective memory trials (five per block) within a time period that would not cause fatigue in children as young as 5 years old (2.5 min per block). However by omitting the KBIT assessment in the future (which occupied a majority of the session time, yet was unrelated to prospective memory performance or metacognitive accuracy), more time can be allotted to the CyberCruiser 2.0 task without increasing overall session time.

**Ongoing task performance**

This study demonstrated that some participants suffer costs to ongoing performance as a result of completing a prospective memory task, whereas others do not.
Einstein et al. (2005) have interpreted this pattern as supporting the multiprocess framework in the context of event-based prospective memory, suggesting that some participants engage in monitoring an external source, whereas others rely on spontaneous retrieval. However, in the context of time-based prospective memory, explored here, it is difficult to understand how participants might be relying on spontaneous retrieval processes in the absence of an explicit, discrete target event. Thus, another explanation is required.

Adults incurred less of a cost than children, overall, suggesting that the costs associated with prospective memory were related to task difficulty more than monitoring or spontaneous retrieval processes. However, the age x gender x block interaction is difficult to interpret. Males in kindergarten and 2\textsuperscript{nd} grade suffered prospective memory costs, but not 1\textsuperscript{st} grade or adult males. In contrast, kindergarten and adult females did not incur a cost, whereas 1\textsuperscript{st} grade and 2\textsuperscript{nd} grade females did. The small sample sizes associated with these groups, as well as the noise inherent to their lap times, make meaningful interpretations of these interactions difficult. However, the interaction between block and age suggest the costs of prospective memory are reduced with age.

Unfortunately, children’s estimates of the costs of prospective memory were unreliable. Future iterations of this study will take measure to ensure that children understand they are being asked to estimate differences in ongoing task performance associated with the prospective memory task, not practice. However, adults’ estimates proved viable and revealed interesting findings regarding the relationship between metacognitive understandings of prospective memory costs and performance. As predicted, adults who underestimated the costs of the prospective memory task were
more likely to run out of gas relative to adults who accurately estimated the costs and those who overestimated these costs. However, their accuracy was not related to fuel-checking frequency or whether they adopted the timer as an external strategy, contrary to predictions. Therefore, it is unclear how adults were allocating attention strategically to compensate for what they perceived to be the costs of the prospective memory task.

**Strategy use**

Although it seems reasonable to propose that the timer was the most strategic reminder, relative to using the billboards as a visual cue or to monitoring the gas tank, this could not be confirmed in the present study. Although participants were more likely to use the timer in the final round, these individuals were not more likely to refill the gas than those who elected to rely on billboards or no external strategy. Using a strategy did not improve ongoing task performance either.

One possible explanation was that participants were not actually relying on the strategy, perhaps because they did not trust it or did not properly understand its use, but analyses revealed this was not the case; participants who used the timer engaged in less fuel-checking relative to participants who used the billboards or simply relied on gauge monitoring. An alternative explanation is that the timer device was not reliable in producing the cue at the right time to facilitate appropriately timed execution. However, my own observations and the observations of other experimenters confirmed this was not the case. Therefore, one might conclude that participants experienced a failure of retrospective memory when the timer sounded (i.e., they forgot the timer indicated they were to refill the tank). While I cannot entirely rule out this explanation, I find it unlikely given that the external reminder block was always raced last, by which point participants
had been presented with five opportunities to refuel the tank while racing the prospective memory block. Thus, participants seemed to suffer some other utilization deficiency, appropriately implementing the new and effective strategy, yet failing to benefit from it.

I predicted that that when an individual is aware of the costs of the prospective memory task and considers these costs to be greater than the costs of setting a timer, he or she should opt to do so. Perhaps the timer was costly upon initial use, due to its unfamiliarity or even cumbersome nature. Some supporting evidence for this conjecture is provided by the fact that 1st graders who used the timer were slightly less likely to refill the gas tank in the external reminder condition but not when using the billboard or neither external strategy.

Participants’ unfamiliarity with the external strategies may also account for the failure to find a relationship between adults’ accuracy when estimating prospective memory costs and their choice of external strategy. Recall I predicted that individuals who understand the costs suffered by engaging in monitoring during a prospective memory task should adopt strategies to reduce these costs. One effective strategy is to slide the time-based prospective memory task down the continuum toward event-based prospective memory by setting a timer to serve as an explicit reminder that the delay interval is ending. However, adults who accurately perceived they had raced a slower lap during the prospective memory block were no more likely to adopt the timer than adults who underestimated or overestimated these costs. It is possible that, without experience using the timer, adults were unable to reason about its utility. Consistent with this interpretation, recall that KBIT scores were related to adults’ use of the timer, suggesting that more intelligent adults had better insight into the potential effectiveness of the timer.
strategy. However, to date there are no studies examining the relationship between IQ and causal reasoning about objects and their affordances based on observation. Future studies will allow participants additional opportunities to practice using these strategies, to assess whether increased familiarity with their utility will (a) improve the relationship between metacognitive understandings of the costs of prospective memory and strategy-use and (b) eliminate the utilization deficiencies observed here.

**Motivation**

In contrast to Study 1, predictions set forth by the goal-based model of prospective memory, as well as those of Kliegel and colleagues (2001, 2004), motivation did not predict prospective memory performance or monitoring. Surprisingly, most participants found the prospective memory task more important, even though the superordinate goal was to achieve a fast lap time, a direct product of the ongoing activity (i.e., racing). It remains unclear how a goal-based model accounts for time-based prospective memory performance. One interpretation is that the immediacy of the prospective memory task led participants to prioritize it over the superordinate goal of earning a fast lap time but did not lead to increased fuel-checking because participants did not want to suffer costs to their lap times, the ultimate goal.

**Time estimation**

Consistent with the findings of Block et al.’s meta-analysis (1999), participants overestimated durations. Contrary to their findings, however, precision did not improve with age. It should be noted that Mäntylä and Carelli (2005) also failed to demonstrate age differences between school-aged children and young adults asked to reproduce durations of between 4- and 36-seconds.
Recall that Mackinlay et al. (2009) and Mäntylä and Carelli (2005) failed to demonstrate a relationship between time estimation and time-based prospective memory performance after controlling for individual differences in executive functioning. Mackinlay et al. (2009) did find that time estimation accuracy negatively predicted time-based prospective memory performance in bivariate correlations, however ($r = -.35$). Similarly, in the present study, time estimation was related to prospective memory performance, such that participants who overestimated time were more likely to refuel the gas tank.

The discrepancy in findings across studies may be due to the fact that others have contrasted accuracy with inaccuracy, whereas I examined the direction of this inaccuracy. For example, Goldstein (2005) reported that young adults who accurately produced 1-minute intervals made significantly more timing errors on a time-based prospective memory task (pressing a button every 7 min) and checked the clock more frequently than did inaccurate estimators. From a methodological standpoint, my study underscores the importance of considering the direction of participants’ inaccuracies; findings indicated that participants who perceive time as passing faster than reality are more likely to refuel the gas tank, relative to those who are accurate or underestimate time. Consistent with this interpretation, the relationship between time estimation abilities and monitoring revealed that participants who perceived time as passing faster than reality were more vigilant in monitoring their fuel level. These findings further confirm that prospective memory success in the CyberCruiser 2.0 relied less on prospective memory ability and more on estimations of elapsed time.
Age effects in participants’ accuracy when reasoning about their time estimation abilities were inconsistent with those indicating that young children tend to overestimate their abilities relative to adults (Bjorklund, Coyle, & Gaultney, 1992; Boulton & Smith, 1990; Flavell et al., 1970; Humphreys & Smith, 1987; Mills & Keil, 2004; Pickert & Wall, 1981; Plumert, 1995; Plumert & Schwebel, 1997; Shin & Choi, 2002; Schneider, 1998; Yussen & Levy, 1975; Spinath & Spinath, 2005; Stipek, 1981). Kindergarten children seemed to have an accurate representation of their time estimation abilities when prospectively predicting performance, but overestimated their abilities following the tasks, indicating they felt they had done better than they actually had (see Bjorklund, Gaultney, & Green, 1993, for a similar pattern in children’s metaimitation). First and 2nd grade children were accurate when making prospective and retrospective estimates. In contrast, adults significantly underestimated their abilities, both before and after the time estimation tasks.

Although these findings are unexpected in an absolute sense, the trend is consistent with that reported in the literature, namely that young children are more confident in their abilities than are older children and adults. With respect to metamemory, young children typically overestimate their performance relative to older children and adults, who are accurate. In the case of time estimation here, a shift in the intercept of the accuracy-age relationship put children at a point of accuracy when predicting their abilities, whereas adults shifted toward doubting their abilities. This shift may have been the result of two factors: Because children are taught about time (how to read clocks, for example) during the early elementary years, children this age may possess an explicit awareness that they are not good at telling or knowing about time
(e.g., when they have difficulty reading clocks). Adults’ inaccuracy, by comparison, may be an artifact of the methods used here to solicit estimations. When asked about their knowledge of time durations prior to the tasks, adults may have found it reasonable to believe they possessed a generally poor sense of time. Likewise, adults may have also felt it unreasonable to think they were “exactly right” on any of the time estimation tasks, further lowering their accuracy scores relative to their actual time deviations during the four time estimation tasks. In the case of time estimation, it may be beneficial to slightly underestimate one’s own sense of time, if this motivates one to vigilantly monitor an external timing source. This prediction was partially supported here by a relationship between accuracy and fuel monitoring frequency for participants with poor time estimation abilities.

Participants’ metacognitive understanding of time estimation abilities predicted monitoring frequency in the prospective memory task. Participants who accurately understood they were poor at estimating time checked the gas more frequently, as if understanding they could not trust their own judgments of time, whereas those who understood their competence when estimating time tended to rely on this internal sense of time and engaged in less frequent gauge monitoring. This relationship was consistent across all age groups, suggesting that even 5-year-olds who accurately assessed their sense of time as poor were capable of engaging in more frequent monitoring, improving their time-based prospective memory performance as a result. Developmental differences in metacognitive understanding of time estimation predict monitoring behaviors in time-based prospective memory.
Metacognitive understanding of time estimation abilities did not predict external strategy use. As described above, this is possibly because participants’ were yet to become familiar with using the timer, and they may have perceived its use as costly relative to the monitoring the fuel tank, a strategy they had already practiced.

This study also demonstrated modest effects associated with when participants were asked to estimate their performance. For example, when participants were asked to estimate their performance prior to racing (pre-trial condition), they crashed significantly more often in the prospective memory block relative to the control block. These findings have important implications for future studies involving metacognitive inquiry and suggest that asking participants to estimate their performance prior to a task may have caused them to become more anxious on the subsequent task.

**Limitations and Future Directions**

Issues that might have limited the tests of specific hypotheses are described above, including the short delay interval during the CyberCruiser 2.0 and that participants were only given one block to use the timer.

More generally, the sample size also limited statistical power to interpret some three- and four-way interactions. One of the major limitations of the current study is methodological. The CyberCruiser 2.0 does not currently allow for assessments of frequency or patterns of gas-checking past the first minute. This is because there could have been up to 5 sec of variability in when participants refilled the tank for the first time, given that the fuel tank was available for refill during the entire last 5-sec of the 30-sec trial interval. Because the “clock” tracking participants’ actions continues to run throughout the game, rather than restarting at the beginning of each trial, some variability
emerges in the time participants begin subsequent trials. Therefore, for participants who refuel the tank as soon as it becomes available, their second trial may begin as early as 26-sec into the game and could be expected to end prior to 60-sec into the game, for example. For participants who do not refuel the tank until it is almost empty, their second trial will begin a few seconds later than those who refill soon (e.g., around 30-sec) and may end closer to 60-sec from the beginning of the game. However, if a participant fails to run out of gas during the first trial, more time is “eaten up” by the car stalling out and the tank refueling before the second trial can begin. In these cases, a participant may not begin the second trial, and could not be expected to begin checking the fuel tank again, until well after 30-sec and may not end this trial until well into the second minute of play.

Although adjustments were made to control for this variability when calculating checking frequency across the first “minute,” or first two trials, here, this example is presented to provide the reader with an idea of the variability in trial onset and offset that can accumulate after five trials, making the measure of checking frequency and pattern a tedious one, at best, when measured by the minute, as was done here. Patterns of checking were only assessed during the first trial here. While this was justified, given the different patterns observed by Kerns (2000) and Ceci and Bronfennbrenner (1983), described in Chapter 3, this may have limited detection of a TWTE strategic pattern. Participants may have adjusted their checking patterns during subsequent trials, after becoming calibrated with the fuel gauge’s rate of decline during the first trial. Future work is needed to examine whether checking patterns become more strategic following the initial trial. Revisions to the CyberCruiser 2.0 are needed that will allow for (1) an ongoing measure of time, to determine lap time and (2) a clock that recycles at the
beginning of every new trial, to allow for better accuracy when assessing participants’ refueling, monitoring frequency, and crashing behaviors.

Future iterations of this study might also benefit from more standardized measures of metacognition, such as the Wurzburg Metamemory Test (Schlagmuller et al., 2001), Metacognitive Awareness Inventory and Inaccurate/Accurate Metacognition Inventory (Schraw & Dennison, 1994), to assess how general metacognitive abilities predict the aspects of performance investigated here. Measures of temporal perspective (the TimeStyle Inventory, Fortunato & Furey, 2010) have also been recently developed to assess individual differences in future, present, and past orientation. Exploring how these personality expressions correlate with prospective memory performance and strategy use may be an interesting step toward further understanding individual differences in prospective memory.

Other studies of metacognition ask participants to make confidence judgments about their responses following each trial (e.g., Kelemen, Frost, & Weaver, 2000). In the future, I will use this method when asking participants to predict their time estimation abilities and apply statistical comparisons of bias and accuracy (gamma). Moreover, controls for executive functioning and experience with video games may have helped to flesh out some of the noise that was possibly responsible for gender and block interactions.

Conclusions

Children as young as 5 years old are capable of participating in the CyberCruiser 2.0 and seem to enjoy doing so. However, performance on this time-based prospective memory task improved with age, suggesting development in the abilities supporting time-
based prospective memory. Younger children tended to overestimate their prospective memory abilities and checked the gas less frequently, causing them to miss more opportunities to refuel compared to older children and adults.

Similarly, participants who underestimated the costs of prospective memory were more likely to run out of gas relative to those who more accurately assessed these costs. Although this relationship was limited to adults, it suggests that a poor metacognitive understanding of the costs of prospective memory may result in missed opportunities to carry out a delayed intention if individuals fail to allocate attentional resources appropriately.

Finally, time estimation was related to prospective memory performance, such that participants who overestimated time durations (thought more time had passed than in reality) were more likely to refuel the tank. Time estimation did not improve with age, nor did the relationship between time estimation and prospective memory performance change. Therefore, the time-based performance of children as young as 5 years old seems to be influenced to some degree by individual differences in their time estimation abilities. Moreover, participants who accurately understood they possessed a poor sense of time engaged in more frequent clock-checking, suggesting they engaged in compensatory measures to improve their chances of detecting the 5-sec interval during which the tank could be refilled. This relationship also remained constant across age groups, suggesting that individual differences in the extent to which 5-year-olds can metacognitively represent their internal clock predicts differences in their monitoring patterns.

Although time estimation and metacognitive representations thereof did not differ
across age group, the number of successful refills and frequency of fuel-checking did. These findings are puzzling and suggest that, although these abilities are sufficient to account for individual differences in performance, there remain developmental differences in the ability to remember to carry out a delayed intention at the appropriate time that are not captured by the higher level measures examined here.

The recent increased interest in metacognitive processes as they relate to our everyday activities (meditation, mindfulness) underscores the importance of considering how individual differences in metacognitive abilities and strategy-use support prospective memory performance.
CHAPTER FIVE

FINAL COMMENTS AND FUTURE DIRECTIONS

Prospective memory is suffering an identity crisis. With critics who argue that prospective memory is fundamentally no different than episodic memory, researchers who seek to understand prospective memory failures must first endure the task of defining prospective memory as a unique construct.

Graf and Uttl (2001) have made strides in this direction, as have Penningroth and Scott (2007, 2008). The emerging model of prospective memory appears to be one that is continuous, extending between tasks predominantly executed through vigilant monitoring and strategic practices and tasks for which “pop-up” spontaneous retrieval processes can be relied upon (Graf & Uttl, 2001). Building on motivational-cognitive models, Penningroth and Scott (2007, 2008) argue that processes supporting higher-order goal achievement are important for prospective memory. Based on the findings reported here, I argue that motivation (i.e., task-importance) is one factor that may influence the position of a prospective memory task on the continuum by influencing the strength of goal representations and the reliance on representational ability.

In Chapter 2, I reported that children as young as 3 years old were more likely to carry out prospective memory tasks of high-interest, and did so with little reliance on executive functioning capacity or representational ability. Interpreted through the models
of Graf and Uttl (2001) and Penningroth and Scott (2007, 2008), these findings suggest that 3-year-olds were more likely to retrieve a sticker than they were to remember to change a sign on the door because the former aligned with goal-relevant behaviors and achieved greater activation as a result. In contrast, participants were less likely to change a sign on the door and required significantly more cues to do so. Moreover, performance on this task was more reliant on executive function and representational ability. Accordingly, because this task was less goal-relevant (i.e., less interesting), it did not benefit from the “activation boost” granted to the high-interest sticker task and required advanced goal representation and monitoring to complete.

Quantitative changes, such as central executive capacity, may be important to the development of prospective memory, but qualitative changes associated with the development of representational abilities should also be considered to contribute to prospective memory development. Although the findings reported in Chapter 2 do not allow the conclusion that representational abilities precede prospective memory development, they do indicate that prospective memory is another symptom of the qualitative change that emerges in many domains across the third and fourth year of life.

Chapter 2 also described the first study to directly demonstrate the role of individual differences in representational abilities in prospective memory development. These findings replicated some findings of others, demonstrating, for example, how motivational factors and individual differences in executive functioning predict
prospective memory (e.g., Ceci & Bronfenbrenner, 1985; Kerns, 2000; Somerville et al., 1983). Continued work in this area is important for understanding age-appropriate prospective memory tasks, especially as children enter preschool and elementary school.

Building on the concept that representational ability is important for prospective memory development and influential in predicting individual differences in success, I described, in Chapter 3, how individual differences in metacognition might influence strategy use in time-based prospective memory performance. In Chapter 4, I described the development of the CyberCruiser 2.0 task and demonstrated that children as young as 5 years old are capable of completing a time-based prospective memory task. I also described that performance on this time-based prospective memory task improved with age and presented findings suggesting that the development of metacognitive skills may partially mediate the age differences in success. Younger children tended to overestimate their prospective memory abilities and checked the gas less frequently, causing them to miss more opportunities to refuel compared to older children and adults.

Similarly, participants who underestimated the costs of prospective memory were more likely to run out of gas relative to those who more accurately assessed these costs. Although this relationship was limited to adults, it suggested that an individual’s poor understanding of the costs of prospective memory might result in missed opportunities to carry out a delayed intention if that individual fails to strategically allocate attention.

Behaviorists argued that consciousness was an epiphenomenon and not a causal component of the relationship between external stimulus and behavioral response. However, the studies reported here, coupled with the research in other areas of psychology and education, underscore the importance of considering how individual
differences in metacognitive abilities and strategy-use support prospective memory performance.

Moreover, this capacity, and the capacity to mentally time travel, more broadly, may have been a driving force in the evolution of *Homo sapiens*. Reasoning about future goal states may allow an individual to predict, shape, and respond flexibly to the environment. Imagining future scenarios means conceiving of possibilities. This was likely advantageous for our ancestors when adapting to a variable environment; if individuals can maneuver their current environments to match the who, what, when, and where of their superordinate goals, they have effectively adapted the environment to meet their needs, instead of the contrary.

Using prospective memory to achieve a future goal may also provide individuals with motivation to delay gratification, encouraging restrained choices, and a decrease in the extent to which the future is “discounted” (Boyer, 2008). Thus, it is facilitated by inhibitory capacity, and executive functioning more generally. Moreover, rehearsal allows one to think through a problem space, representing an initial state, a means, and an end. This capacity was likely advantageous to our ancestors, allowing them to stay “one step ahead” of the competition.

In this way, prospective memory is distinct from episodic memory in that it recruits autonoesis in the service of future scenarios and goals. The development of this autonoetic ability in humans permits the representation of intentions, and eventually the ability to adopt strategies to successfully carry out these intentions. Representational abilities may subserve event-based prospective memory at a relatively basic level, by allowing an individual to hold a future intention in mind, even when the intended action
is not particularly enticing. Moreover, by imagining the future episode during which the activity must be completed, one can build a strong association between the target event and intended action, leading to successful retrieval and implementation. Regarding time-based prospective memory, the development of advanced representational abilities allows one to reflect on one’s own performance and capacities and to adopt appropriate strategies for completing the task on time. These abilities develop across early childhood and account for developmental differences in success on these tasks.

**Future Directions**

The studies presented here describe how motivational and metacognitive factors influence where a task falls on the prospective memory continuum and demonstrated how the development of these abilities predicts children’s performance on prospective memory tasks at across early childhood. Future research is charged with continuing the task of identifying these factors and their development.

Because so many prospective memory tasks in the real world involve reminding another individual to carry out a task, a clear future direction is toward a better understanding of the processes that facilitate this ability. Preliminary evidence, described in Chapter 2, suggests that this ability develops across the third year and may rely on second-order theory of mind.

Another direction for future research is to examine how a representational theory of prospective memory accounts for prospective memory development and decline across the life span. For many years, it was thought that representational abilities do not decline with age (e.g., Hedden & Gabrieli, 2004). However there is some suggestion that older adults perform more poorly on future-oriented tasks because of declines in relational
abilities, that is, in their capacity to recombine mental representations to construct future episodes (see Schacter, Addis, & Buckner, 2007). Research has also demonstrated a decline in performance on theory of mind tasks in later adulthood (Maylor, Moulson, Muncer, & Taylor, 2002), but others have shown this decline is accounted for by diminishing executive functioning capacities (German & Hehman, 2006).

Studies have also revealed that older adults are more likely to compensate for their poorer time-based prospective memory performance by adopting external strategies. These findings suggest that metacognition, particularly self-monitoring, may play an even more important role in prospective memory later in life. Relatively few studies have examined changes in metacognition later in life, but those available find that metamemory does not decline with age (Lachman, 1979) or is a biproduct of declines in executive functioning (Souchay & Isingrini, 2004). Therefore, a clear future direction is to examine the role of representational abilities in accounting for changes in prospective memory across the lifespan.
Table 1

*Mean Performance on Reminding Tasks for 2-, 3-, and 4-year-olds, from Somerville, Wellman, and Cultice (1983, p. 91)*

<table>
<thead>
<tr>
<th>Age group</th>
<th>HI-SD</th>
<th>HI-LD</th>
<th>LI-SD</th>
<th>LI-LD</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spontaneous</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 years</td>
<td>.80(^a)(1.14)(^b)</td>
<td>.50(1.14)</td>
<td>.20(.29)</td>
<td>.00(.29)</td>
<td>.38(.71)</td>
</tr>
<tr>
<td>3 years</td>
<td>.60(1.67)</td>
<td>.50(1.17)</td>
<td>.30(1.00)</td>
<td>.20(.33)</td>
<td>.40(1.04)</td>
</tr>
<tr>
<td>4 years</td>
<td>.80(1.57)</td>
<td>.60(1.14)</td>
<td>.20(4.3)</td>
<td>.30(.43)</td>
<td>.48(.89)</td>
</tr>
<tr>
<td>Average</td>
<td>.73(1.45)</td>
<td>.53(1.15)</td>
<td>.23(.55)</td>
<td>.17(.35)</td>
<td>.42(.88)</td>
</tr>
<tr>
<td><strong>Prompted</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 years</td>
<td>.90(1.71)</td>
<td>.70(1.29)</td>
<td>.40(.43)</td>
<td>.10(.43)</td>
<td>.53(.96)</td>
</tr>
<tr>
<td>3 years</td>
<td>.90(2.00)</td>
<td>.70(1.50)</td>
<td>.40(1.17)</td>
<td>.30(.50)</td>
<td>.58(1.29)</td>
</tr>
<tr>
<td>4 years</td>
<td>.80(1.71)</td>
<td>.80(1.29)</td>
<td>.40(.57)</td>
<td>.30(1.00)</td>
<td>.58(1.14)</td>
</tr>
<tr>
<td>Average</td>
<td>.87(1.80)</td>
<td>.73(1.35)</td>
<td>.40(.70)</td>
<td>.23(.65)</td>
<td>.56(1.13)</td>
</tr>
</tbody>
</table>

*Note: HI-SD = high interest, short delay; HI-LD = high interest, long delay; LI-SD = low interest, short delay; LI-LD = low interest, long delay.*

*\(^a\) Scores represent one reminding task of each type for each child. Thus, the maximum score = 1.00; \(N = 10\) at all ages.*

*\(^b\) Scores in parentheses are based on two reminding tasks of each type for each child. Thus, the maximum score = 2.00; \(N = 7\) for 2-year-olds, \(N = 6\) for 3-year-olds, \(N = 7\) for 4-year-olds.*
Table 2

*Results from Guajardo and Best (2000, p. 86)*

Proportion of 3- and 5-year-old children who remembered the naturalistic, prospective memory tasks

<table>
<thead>
<tr>
<th>Age</th>
<th>Tasks</th>
<th>Sticker</th>
<th>Door</th>
<th>Pencil</th>
<th>Picture</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>no prompt</td>
<td>0.64</td>
<td>0.16</td>
<td>0.43</td>
<td>0.78</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>first prompt</td>
<td>0.24</td>
<td>0.42</td>
<td>0.43</td>
<td>0.17</td>
<td>0.32g</td>
</tr>
<tr>
<td></td>
<td>second prompt</td>
<td>0.12</td>
<td>0.42</td>
<td>0.14</td>
<td>0.05</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>0.52 b,c</td>
<td>0.25 b,d</td>
<td>0.29 c</td>
<td>0.37 f</td>
<td>0.38</td>
</tr>
<tr>
<td>5</td>
<td>no prompt</td>
<td>0.70</td>
<td>0.67</td>
<td>0.60</td>
<td>0.82</td>
<td>0.70</td>
</tr>
<tr>
<td></td>
<td>first prompt</td>
<td>0.25</td>
<td>0.28</td>
<td>0.24</td>
<td>0.09</td>
<td>0.22a</td>
</tr>
<tr>
<td></td>
<td>second prompt</td>
<td>0.05</td>
<td>0.05</td>
<td>0.16</td>
<td>0.09</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>0.83 c</td>
<td>0.75 g</td>
<td>0.52 c,g</td>
<td>0.73 f,g</td>
<td>0.71</td>
</tr>
</tbody>
</table>

a $p < 0.05$.

b, c $p < 0.01$.
d, f $p < 0.001$.
e, g $p < 0.02$. 

137
Table 3

**Counterbalancing of Executive Functioning Task Sets with Prospective Memory Tasks Across Session.**

<table>
<thead>
<tr>
<th></th>
<th>Session 1</th>
<th>Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Executive Functioning</td>
<td>Prospective Memory Tasks</td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Set A</td>
<td>High-interest, “remind me” / Low-interest, “remember”</td>
</tr>
<tr>
<td>8</td>
<td>Set B</td>
<td>High-interest, “remind me” / Low-interest, “remember”</td>
</tr>
<tr>
<td>8</td>
<td>Set A</td>
<td>Low-interest, “remind me” / High-interest, “remember”</td>
</tr>
<tr>
<td>8</td>
<td>Set B</td>
<td>Low-interest, “remind me” / High-interest, “remember”</td>
</tr>
</tbody>
</table>
Table 4

*Correlations and Descriptive Statistics for Performance on Measures of Executive Functioning.*

<table>
<thead>
<tr>
<th>Executive Function Tasks</th>
<th>Simon Says</th>
<th>1-back</th>
<th>Digit Span</th>
<th>FTS</th>
<th>BTS</th>
<th>Stroop</th>
<th>DCCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simon Says</td>
<td>3.31</td>
<td>.24</td>
<td>.24</td>
<td>.30*</td>
<td>-.06</td>
<td>.22</td>
<td>.19</td>
</tr>
<tr>
<td>SD = 2.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-Back</td>
<td>3.07</td>
<td>.31*</td>
<td>.25+</td>
<td>-.01</td>
<td>.01</td>
<td>.09</td>
<td></td>
</tr>
<tr>
<td>SD = 3.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Digit Span</td>
<td>4.43</td>
<td>.49**</td>
<td>-.18</td>
<td>.11</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 1.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward Trail Span</td>
<td>2.73</td>
<td>.13</td>
<td>.24</td>
<td>.118</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 1.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backward Trail Span</td>
<td>.59</td>
<td>.30+</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 1.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroop Error Difference</td>
<td>2.33</td>
<td>.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD = 3.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCCS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.90</td>
</tr>
<tr>
<td>SD = 1.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>-.05</td>
<td>.37**</td>
<td>.23</td>
<td>.40*</td>
<td>.29*</td>
<td>.12</td>
<td>.20</td>
</tr>
<tr>
<td>SD = .10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* +p < .10; *p < .05 level; **p <.01; “FTS” is an abbreviation for forward trail span; “BTS” is an abbreviation for backward trail span; “DCCS” is an abbreviation for dimensional change card sort. Correlations in bold are reported as Kendall’s tau.
Table 5

Correlations for Prospective Memory Performance and Measures of Executive Functioning

<table>
<thead>
<tr>
<th>EF</th>
<th>Prospective memory</th>
<th>Simon Says</th>
<th>1-back</th>
<th>Digit Span</th>
<th>FTS</th>
<th>BTS</th>
<th>Stroop</th>
<th>DCCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall prospective memory performance</td>
<td>.37*</td>
<td>.31*</td>
<td>.13</td>
<td>.43*</td>
<td>.14</td>
<td>-.06</td>
<td>.30+</td>
<td></td>
</tr>
<tr>
<td>Number of Cues Received:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-interest “Remember”</td>
<td>.03</td>
<td>-.22</td>
<td>-.05</td>
<td>-.15</td>
<td>-.19</td>
<td>.20</td>
<td>.05</td>
<td></td>
</tr>
<tr>
<td>High-interest “Remind”</td>
<td>-.13</td>
<td>-.32*</td>
<td>-.41*</td>
<td>-.14</td>
<td>.04</td>
<td>.09</td>
<td>-.21</td>
<td></td>
</tr>
<tr>
<td>Low-interest “Remember”</td>
<td>-.32*</td>
<td>-.40**</td>
<td>-.40*</td>
<td>-.52**</td>
<td>-.04</td>
<td>-.22</td>
<td>-.10</td>
<td></td>
</tr>
<tr>
<td>Low-interest “Remind”</td>
<td>-.32*</td>
<td>-.14</td>
<td>.01</td>
<td>-.09</td>
<td>-.10</td>
<td>.11</td>
<td>-.42*</td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .10; **p < .05 level; “FTS” is an abbreviation for forward trail span; “BTS” is an abbreviation for backward trail span; “DCCS” is an abbreviation for dimensional change card sort. Correlations in bold are reported as Kendall’s tau.
Table 6

Factor Loadings and Communalities based on a Principal Components Analysis with Varimax Rotation for Five Items (N = 31)

<table>
<thead>
<tr>
<th>Task</th>
<th>Executive Function (Eigen value = 2.23)</th>
<th>Representational Ability (Eigen value = 1.14)</th>
<th>Communality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simon Says</td>
<td>.81</td>
<td>.13</td>
<td>.67</td>
</tr>
<tr>
<td>Stroop Error Difference</td>
<td>-.79</td>
<td>.04</td>
<td>.62</td>
</tr>
<tr>
<td>Forward Trail Span</td>
<td>.74</td>
<td>.38</td>
<td>.69</td>
</tr>
<tr>
<td>DCCS</td>
<td>.24</td>
<td>.74</td>
<td>.61</td>
</tr>
<tr>
<td>Sum Belief/Desire Reasoning</td>
<td>-.02</td>
<td>.89</td>
<td>.79</td>
</tr>
</tbody>
</table>
Table 7

*Hierarchical Regression Analysis Predicting Individual Differences in Overall Prospective Memory Performance of Children*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
<td>B</td>
</tr>
<tr>
<td>Age</td>
<td>.05</td>
<td>.04</td>
<td>.23</td>
<td>.05</td>
</tr>
<tr>
<td>Gender</td>
<td>.92</td>
<td>.40</td>
<td>.40*</td>
<td>.85</td>
</tr>
<tr>
<td>Executive Functioning</td>
<td></td>
<td></td>
<td></td>
<td>.66</td>
</tr>
<tr>
<td>Representational Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.05</td>
<td>.21</td>
<td></td>
<td>.33</td>
</tr>
<tr>
<td>$F$ for change in $R^2$</td>
<td>1.67</td>
<td></td>
<td></td>
<td>5.62*</td>
</tr>
</tbody>
</table>

*Note:* **$p < .01$; *$p < .05$; +$p < .10.$*
Table 8

*Hierarchical Regression Analysis Predicting Individual Differences in the Number of Cues Children Received on Prospective Memory Tasks varying by Interest Level*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
<th>Model 3</th>
<th></th>
<th></th>
<th>Model 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
<td>B</td>
<td>SE(B)</td>
</tr>
<tr>
<td>Age</td>
<td>.04</td>
<td>.03</td>
<td>.21</td>
<td>.04</td>
<td>.03</td>
<td>.21</td>
<td>.04</td>
<td>.03</td>
<td>.21</td>
<td>.01</td>
<td>.03</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td>-.31</td>
<td>.34</td>
<td>-.17</td>
<td>-.32</td>
<td>.35</td>
<td>-.17</td>
<td>-.43</td>
<td>.34</td>
</tr>
<tr>
<td>Executive Functioning</td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
<td>.28</td>
<td>.03</td>
<td>-.08</td>
<td>.27</td>
<td>-.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Representational Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.50</td>
<td>.26</td>
<td>.38*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.05</td>
<td></td>
<td>.07</td>
<td>.07</td>
<td></td>
<td>.19*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$F$ for change in $R^2$:
- Model 1: 1.37
- Model 2: $<1$
- Model 3: $<1$
- Model 4: 3.79*

*Note:* **$p < .01$; *$p < .05$; +$p = .06$; Difference scores were calculated as mean cues for high-interest tasks minus mean cues for low-interest tasks; lower, negative difference scores indicate more cues were given for the low-interest relative to high-interest tasks.*
Table 9

Hierarchical Regression Analysis Predicting Individual Differences in the Number of Cues Children Received for Prospective Memory Tasks varying by Instruction.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th></th>
<th></th>
<th>Model 2</th>
<th></th>
<th></th>
<th>Model 3</th>
<th></th>
<th></th>
<th>Model 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
<td>B</td>
<td>SE(B)</td>
</tr>
<tr>
<td>Gender</td>
<td>.07</td>
<td>.31</td>
<td>.04</td>
<td>.08</td>
<td>.32</td>
<td>.05</td>
<td>.09</td>
<td>.33</td>
<td>.05</td>
<td>.04</td>
<td>.30</td>
</tr>
<tr>
<td>Executive Functioning</td>
<td></td>
<td></td>
<td></td>
<td>-.09</td>
<td>.25</td>
<td>-.06</td>
<td>-.08</td>
<td>.27</td>
<td>-.06</td>
<td>-.04</td>
<td>.25</td>
</tr>
<tr>
<td>Representational Ability</td>
<td></td>
<td></td>
<td></td>
<td>-.02</td>
<td>.24</td>
<td>-.02</td>
<td>.17</td>
<td>.23</td>
<td>.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.07</td>
<td>.03</td>
<td>-.46*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.00</td>
<td></td>
<td></td>
<td>.01</td>
<td></td>
<td></td>
<td>.01</td>
<td></td>
<td></td>
<td>.19</td>
<td></td>
</tr>
</tbody>
</table>

$F$ for change in $R^2$        | <1      | <1    | <1    | 5.96*   |       |       |         |       |       |

Note: **$p < .01$; *$p < .05$; +$p < .10$; Difference scores were calculated as mean cues for “remember” tasks minus mean cues for “remind” tasks; lower, negative difference scores indicate more cues were given on “remind” relative to “remember” tasks.
Table 10

Hierarchical Regression Analysis Predicting Individual Differences in the Relative Difficulty with which Children Passed each Prospective Memory Task.

<table>
<thead>
<tr>
<th>Variable</th>
<th>High-interest “Remind”</th>
<th>Low-interest “Remember”</th>
<th>Low-interest “Remind”</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE(B)</td>
<td>β</td>
</tr>
<tr>
<td>Age</td>
<td>-.26</td>
<td>.23</td>
<td>-.23</td>
</tr>
<tr>
<td>Gender</td>
<td>-.02</td>
<td>.22</td>
<td>-.02</td>
</tr>
<tr>
<td>Executive Functioning</td>
<td>.18</td>
<td>.36</td>
<td>.10</td>
</tr>
<tr>
<td>Representational Ability</td>
<td>-.07</td>
<td>.34</td>
<td>-.05</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$ for model</td>
<td>&lt;1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .05; Difficulty scores were calculated as the number of cues received for the high-interest “remember” tasks minus the number of cues received for each of the three tasks described in the top row of the table.
Table 11

*Mean Missed Refill Opportunities, Presented as a Three-way Interaction Between Gender, Age, and Block Order.*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Block Order:</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PM First</td>
<td>PM Second</td>
<td>PM First</td>
<td>PM Second</td>
<td>PM First</td>
<td>PM Second</td>
<td></td>
</tr>
<tr>
<td>Kindergarten ((n = 26))</td>
<td>0 (0)</td>
<td>2.8 (1.87)</td>
<td>3.36 (1.56)</td>
<td>.50 (.71)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st Grade ((n = 27))</td>
<td>1.70 (1.20)</td>
<td>1.65 (1.67)</td>
<td>2.33 (1.84)</td>
<td>1.25 (.95)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2nd Grade ((n = 29))</td>
<td>1.92 (1.86)</td>
<td>1.43 (1.72)</td>
<td>2.54 (1.83)</td>
<td>2.50 (2.26)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adult ((n = 32))</td>
<td>0 (0)</td>
<td>.41 (.57)</td>
<td>.13 (.35)</td>
<td>.63 (1.06)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 12

*Mean Accuracy of Prospective Memory Performance Estimates by Age, Block order, and Gender.*

<table>
<thead>
<tr>
<th>PM first, M (SD)</th>
<th>PM second, M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Kindergarten</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Kindergarten</td>
<td>-1.33 (1.53)</td>
</tr>
<tr>
<td>n = 3</td>
<td>n = 11</td>
</tr>
<tr>
<td>1st Grade</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>1st Grade</td>
<td>-1.50 (1.41)</td>
</tr>
<tr>
<td>n = 5</td>
<td>n = 5</td>
</tr>
<tr>
<td>2nd Grade</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>2nd Grade</td>
<td>-.82 (2.20)</td>
</tr>
<tr>
<td>n = 10</td>
<td>n = 6</td>
</tr>
<tr>
<td>Adults</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>Adults</td>
<td>0.0 (0.00)</td>
</tr>
<tr>
<td>n = 8</td>
<td>n = 8</td>
</tr>
</tbody>
</table>

*Note: * gender difference is significant at $p < .05; ^+p < .10.
Table 13

*Mean Crash Rates for Participants, presented by Block and Metacognitive Presentation Condition*

<table>
<thead>
<tr>
<th>Metacognitive Presentation:</th>
<th>Block</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prospective Memory</td>
<td>Control</td>
</tr>
<tr>
<td></td>
<td>$M$ ($SD$)</td>
<td>$M$ ($SD$)</td>
</tr>
<tr>
<td>Pre-trial ($n = 54$)</td>
<td>8.45 (7.61)</td>
<td>4.77 (7.93)</td>
</tr>
<tr>
<td>Post-trial ($n = 59$)</td>
<td>6.06 (6.89)</td>
<td>8.04 (4.98)</td>
</tr>
</tbody>
</table>
Table 14

*Percent (and number) Estimating Direction of Prospective Memory Cost, by Age Group*

<table>
<thead>
<tr>
<th>Response</th>
<th>Kindergarten % (n)</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Grade</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Grade</th>
<th>Adult</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>“slower”</td>
<td>61.5 (16)</td>
<td>25.9 (7)</td>
<td>51.7 (15)</td>
<td>84.4 (27)</td>
<td>57% (65)</td>
</tr>
<tr>
<td>“the same”</td>
<td>0 (0)</td>
<td>3.7 (1)</td>
<td>0 (0)</td>
<td>3.7 (1)</td>
<td>1.8 (2)</td>
</tr>
<tr>
<td>“faster”</td>
<td>38.5 (10)</td>
<td>70.4 (19)</td>
<td>48.3 (14)</td>
<td>12.5 (4)</td>
<td>41.2 (47)</td>
</tr>
</tbody>
</table>


Table 15

*Percentage of Participants Using each Strategy in the Final Round of Racing, broken down by Age Group*

<table>
<thead>
<tr>
<th>Age</th>
<th>Strategy</th>
<th>Kindergarten % (n)</th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Grade % (n)</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Grade % (n)</th>
<th>Adult % (n)</th>
<th>Total % (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neither</td>
<td>40.7 (11)</td>
<td>11.1 (3)</td>
<td>27.6 (8)</td>
<td>31.3 (10)</td>
<td>27.8 (32)</td>
</tr>
<tr>
<td></td>
<td>Billboard</td>
<td>11.1 (3)</td>
<td>25.9 (7)</td>
<td>13.8 (4)</td>
<td>34.4 (11)</td>
<td>21.7 (25)</td>
</tr>
<tr>
<td></td>
<td>Timer</td>
<td>48.1 (13)</td>
<td>63 (17)</td>
<td>58.6 (17)</td>
<td>34.4 (11)</td>
<td>50.4 (58)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100 (27)</td>
<td>100 (27)</td>
<td>100 (29)</td>
<td>100 (32)</td>
<td></td>
<td><strong>100 (115)</strong></td>
</tr>
</tbody>
</table>

$\chi^2(6, N = 115) = 12.17, p = .06$
Table 16

*Distribution of Strategy Use across Gender*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Strategy</th>
<th>Male % (n)</th>
<th>Female % (n)</th>
<th>Total % (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Neither</td>
<td>26.6 (17)</td>
<td>29.4 (15)</td>
<td>27.8 (32)</td>
</tr>
<tr>
<td></td>
<td>Billboard</td>
<td>14.1 (9)</td>
<td>31.4 (16)</td>
<td>21.7 (25)</td>
</tr>
<tr>
<td></td>
<td>Timer</td>
<td>59.4 (38)</td>
<td>39.2 (20)</td>
<td>50.4 (58)</td>
</tr>
<tr>
<td></td>
<td><strong>Total</strong></td>
<td><strong>100 (64)</strong></td>
<td><strong>100 (51)</strong></td>
<td><strong>100 (115)</strong></td>
</tr>
</tbody>
</table>

*Note: χ²(2, N = 115) = 6.28.*
Table 17

*Mean Missed refills for each Age Group, presented by Block and Strategy Use*

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Timer</th>
<th>Billboard</th>
<th>Neither</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prospective Memory</td>
<td>External Reminder</td>
<td>Prospective Memory</td>
</tr>
<tr>
<td>Age</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Kindergarten</td>
<td>2.46 (1.98)</td>
<td>2.52 (1.59)</td>
<td>2.67 (2.08)</td>
</tr>
<tr>
<td>1st Grade</td>
<td>1.46 (1.55)</td>
<td>1.95* (1.69)</td>
<td>2.18 (1.65)</td>
</tr>
<tr>
<td>2nd Grade</td>
<td>1.82 (2.00)</td>
<td>2.21 (1.64)</td>
<td>1.38 (1.60)</td>
</tr>
<tr>
<td>Adult</td>
<td>.57 (.94)</td>
<td>.56 (.56)</td>
<td>.18 (.40)</td>
</tr>
</tbody>
</table>

Note: * Contrast significant at $p < /= .10$. 

$n = 13, n = 17, n = 17, n = 17, n = 11, n = 11, n = 10, n = 7, n = 4, n = 8$
Table 18

*Mean Accuracy of Participants’ Time Estimation Metacognition, presented by Age and Time of Estimate.*

<table>
<thead>
<tr>
<th>Age</th>
<th>Pre-trial</th>
<th>Post-trial</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Kindergarten (n = 27)</td>
<td>-.31 (.83)</td>
<td>-.36 (.81)</td>
</tr>
<tr>
<td>1st Grade (n = 25)</td>
<td>-.13 (.95)</td>
<td>-.13 (.91)</td>
</tr>
<tr>
<td>2nd Grade (n = 29)</td>
<td>-.08 (.60)</td>
<td>.08 (.59)*</td>
</tr>
<tr>
<td>Adult (n = 32)</td>
<td>.33 (.21)</td>
<td>.58 (.26)*</td>
</tr>
</tbody>
</table>

Note: * paired samples t-test significant at p < .05. Negative scores reflect the magnitude by which participants over-estimated abilities.
Figure 1. Percentage of participants passing each prospective memory task.
Figure 2. Percent within each gender passing prospective memory task.
Figure 3. Mean number of cues (1 – 4) presented by prospective memory task type. Error bars represent 5% of the mean.
Figure 4. Mean number of cues received by younger versus older children as a function of prospective memory instruction. Error bars represent 95% confidence intervals.
Figure 5. Percentage of participants passing each measure of belief/desire reasoning ($n = 31$).
Figure 6. Design of Study 2
Figure 7. Mean number of prospective memory trials where participant ran out of gas, presented by age group. Error bars represent 95% confidence interval.
Figure 8. Accuracy of estimates of prospective memory performance, presented by age group. Error bars represent 95% confidence intervals.
Figure 9. Mean gas-checking frequency during first minute of prospective memory block, presented by age. Error bars represent 95% confidence intervals.
Figure 10. Relationship between participants’ accuracy in predicting prospective memory abilities and monitoring.
Figure 11. Mean checking frequency at each 5-second interval during the wait period.
Figure 12. Mean lap time for each block (in seconds) presented by gender and age group.
Figure 13. Partial relationship between adults’ accuracy in predicting the costs to their lap time as a result of the prospective memory task and their prospective memory performance, controlling for KBIT scores. *Note:* Intercept varies by block order and gender, \( \text{var}(u_{ij}) = .06 \).
REFERENCES


Bonnin, F. Apple iPhone timer applications, Copyright 2009.


Einstein G. O., McDaniel M. A. (2010). Prospective memory and what costs do not reveal about retrieval processes: A commentary on Smith, Hunt, McVay, and


