UNIVERSITY OF ALBERTA

ADULT AGE DIFFERENCES IN THE COMPONENTS UNDERLYING PROSPECTIVE MEMORY

BY

M. BARBARA REEVES

A Thesis

Submitted to the Faculty of Graduate Studies and Research

in Partial Fulfillment

of the Requirements for the Degree of

MASTER OF SCIENCE.

DEPARTMENT OF PSYCHOLOGY

Edmonton, Alberta

FALL 1992

ISBN 0-315-77074-0

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: M. Barbara Reeves

TITLE OF THESIS: Adult Age Differences in the Components Underlying Prospective Memory

DEGREE: Master of Science

YEAR THIS DEGREE GRANTED: Fall, 1992

M. Barbara Reeves

(student's signature)

3493 Henderson Rd.
Victoria, B.C.
V8P 5A8

Date: Oct. 5, 1992

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled ADULT AGE DIFFERENCES IN THE COMPONENTS UNDERLYING PROSPECTIVE MEMORY submitted by M. BARBARA REEVES in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE.

Dr. Allen R. Dobbs

Dr. Edward Cornell

Dr. Robert B. Wilberg

Date: Aug. 21, 1992

Dedication

To Mom and Becca,

My Greatest Supporters

and

Best of Friends.

Abstract

Rather than treating prospective memory as a unitary process, it is suggested that prospective memory performance could be more usefully assessed by looking at performance on six components believed to underlie prospective memory. In this study, 24 young adults and 24 old adults performed a task which was designed to assess age differences on five of the PM components (metaknowledge, planning, monitoring, content recall, compliance and output monitoring). The participants' task was to monitor the progress of pebbles falling in three vials simulated on a computer screen. Within each vial, a pebble always fell at a constant rate but the rates of fall varied greatly between the three vials. Each vial could only be seen individually and only after its "show" button was depressed. Whenever a pebble entered a certain region on its vial, the participants had to press a button to reset the pebble's fall. Depending on which vial was reset and how many times it had been reset previously, the participants had to chose between 16 different reset buttons. The vial monitoring task allowed prospective memory data to be collected in terms of monitoring, content recall and output monitoring. Post-task questions supplied data on metaknowledge and planning. Whenever age differences were found, they favoured the younger adults. However, age differences were not found on all the measures either within or between components.

Acknowledgements

Thanks are due to many people who played a role in the completion of this thesis. To the volunteers who gave their time and effort for the sake of science, a sincere thank you. Thanks also to Brian Lake who is truly a programmer extraordinaire. I would also like to express my gratitude to the members of my research group, and in particular Marjorie Anderson, without whose support this thesis would never have been completed on time. A sincere thank you also is due to Edward Cornell and Robert Wilberg who were very tolerant of my deadlines and helpful with their comments. Finally, I would like to thank Allen Dobbs for being both a mentor and a friend.

Table of Contents

Chapter Page

I. Introduction 1

A. Metaknowledge 3

B. Planning 5

C. Monitoring 6

Monitoring accuracy 6

Monitoring behavior 7

D. Content Recall 10

E. Compliance 12

F. Output Monitoring 13

G. Paradigm for Studying the Components of

Prospective Memory 15

II. Method 20

A. Participants 20

B. Procedure 20

Participant information 20

Vial prospective memory task 21

Recognition of the reset words 23

Recollection of the monitoring pattern 23

Metaknowledge for the vial monitoring task ... 24

Planning for the vial monitoring task 24

III. Results 26

A. Monitoring 26

Monitoring errors 26

Monitoring pattern 27

Selective monitoring 33

Relation between monitoring behavior and errors 33
Concurrent task performance 34
Interference between the vial task and the lexical decision task 35
B. Content Recall 36
C. Output Monitoring 37
On-line output monitoring 37
Recognition of the reset words 38
Recollection of the monitoring pattern 40
D. Metaknowledge 43
Temporal measures:
Task length 43
Time to fall 44
Spatial measure:
Final position of the pebbles 44
Rate measures:
Identifying the fast and slow vial 45
First 10 resets 45
Total resets 46
E. Planning 48
Correlations between the strategies 51
IV. Discussion 52
A. Monitoring 52
Monitoring errors 52
Monitoring pattern 53

Selective monitoring 55
B. Content Recall 58
C. Output Monitoring 60
D. Metaknowledge 62
E. Planning 64
F. General Age Findings 66
F. Conclusion 69
V. References 71
VI. Appendix A. Planning Questions 80

List of Tables

Table	Description	Page
1	Percentage of Young and Old Participants Using each Strategy	49

List of Figures

Figure	Description	Page
1	Checks per Zone as a Function of Drop Rate	29
2	Slow Vial: Checks per Zone as a Function of Study Half	31
3	Medium Vial: Checks per Zone as a Function of Study Half	32
4	Estimated vs Actual Monitoring Patterns as a Function of Drop Rate and Zone	42
5	Estimated vs Actual Total Resets as a Function of Age and Drop Rate	47

Introduction

The purpose of this research is to investigate differences between younger and older adults abilities' to remember to carry out future actions. Past research on this topic has been equivocal. Four studies have found younger adults to be superior to older adults (Dobbs & Rule, 1987; Einstein, McDaniel, Cunfer, & Guynn, 1991, Study 1; Schonfield & Shooter, cited in Harris, 1984; West, 1988, Study 2), four studies have found older adults to be superior to younger adults (Devolder, Brigham, & Pressley, 1990; Martin, 1986, Study 2; Moscovitch & Minde, cited in Moscovitch, 1982; Poon & Schaffer, 1982, cited in West, 1988), and six studies have been indeterminate (Einstein & McDaniel, 1990; Einstein et al., 1991, Study 2; Kerr, 1992; Miller, 1990; West, 1984, Study 2, cited in Miller, 1990; West, 1988, Study 1). The reason for these disparate results lies in the nature of the prospective memory tasks used in the research. Experiments that found that older adults were superior to younger adults in remembering to carry out future actions were all done in real world settings, over long intervals, and with the use of memory aids uncontrolled. On the other hand, the experiments in which the younger adults were found to be superior to the older adults were all conducted in research labs, over short intervals, with the use of memory aids controlled by the experimenter. Control over extraneous factors, such as memory aids, therefore seems to be a major factor in determining the direction of age differences in prospective remembering. On the basis of past research, it seems possible to conclude that younger adults are superior to older adults when they have to rely on their memories alone, but older adults can surpass younger adults through the use of compensatory memory aids. However, before endorsing any general conclusions about age differences in prospective memory, it is necessary to consider the status of prospective memory as a theoretical construct.

In terms of theory, the field of prospective memory research is still relatively primitive. In the past, researchers frequently have been more interested in demonstrating that one group is impaired in prospective memory ability in comparison to another group, than in investigating the reasons underlying this difference. Consequently, many studies have looked at the differences between young and old adults, between different aged children, and between pathological groups and their controls, without making a priori predictions of why these groups should differ in this memory ability. Another limitation of the past research that seems to be related to the lack of theory-driven investigations is the field's overreliance on global measures of prospective memory performance (e.g., proportion of postcards returned on time). Taking global measures of prospective memory is tantamount to saying that prospective remembering is a unitary process or that all of the processing tasks are similarly affected by aging. More recently, however, prospective memory has been conceptualized in terms of components. A synthesis of the research indicates that there are as many as six components underlying the prospective remembering process. These components are metaknowledge (knowledge about the specific task demands), planning (the formulation of a plan that should allow the action to be carried out at the appropriate time), monitoring (how and if the person remembers to carry out some action, not necessarily the correct one, at the correct time), content recall (carrying out the correct action at any time), compliance (the person's willingness to carry out the action at the appropriate time) and output monitoring (remembering if the action has been carried out). It seems that rather than making predictions about how older adults will compare to younger adults on a global measure of prospective memory ability, it would be more profitable to consider how different aged adults are likely to differ on each of the components underlying prospective memory ability. Each of these components will now be considered in turn.

Metaknowledge

Within the context of prospective memory tasks, Dobbs and Reeves (1991) have defined metaknowledge as general knowledge about the prospective memory task that can be used to optimize performance. Before people can construct a plan that will help them remember to carry out a future action, they must understand the desirability of constructing such a plan. Young children frequently do not see this need: They just assume that they will be able to remember at the appropriate time (Istomina, 1986; Kreutzer, Leonard, & Flavell, 1975). Knowledge of prospective memory task demands also allows people to construct appropriate plans. A good prospective memory retrieval aid has been defined as one having a clear, specific and detailed associative link with the task to be performed and which will be seen, and recognized as a prospective memory cue, at the right time (Beal, 1985). Beal (1985) has shown that there is a developmental trend for children to recognize the components of a good prospective memory plan.

Several studies have asked older adults to rate their ability to remember to carry out different kinds of future actions. Personal perceptions of the difficulty of different prospective memory tasks can be taken as a measure of prospective memory metaknowledge. The results of these self-rating studies, although clearly biased by the prospective memory tasks chosen for rating, seem to indicate that older people give themselves better ratings on prospective memory ability than younger people give themselves, and that older people make more accurate predictions. For example, Martin (1986) asked people to rate their ability to attend appointments, pay bills, and take medication. In all cases her older participants gave themselves higher ability ratings than her younger participants gave themselves. Similarly, both McMillan (1984) and Bennett-Levy and Powell (1980, cited in McMillan, 1984) found that their participants believed that performance on returning a borrowed item and giving messages (both prospective memory items) increased with age. In addition, Devolder et al. (1990) found that not only did older adults rate their own ability on an appointment keeping task more highly than younger adults rated their own ability but that the personal performance predictions and postdictions of the older adults were more accurate.

Planning

The planning component of prospective memory reflects not people's knowledge of the efficacy of different prospective memory plans, but their actual use of prospective memory plans. It is important to realize that the use of plans and knowledge about the efficient use of plans are not the same thing. A person might know full well if probed that the best way to remember to check some cookies in 15 minutes would be to set a timer; but the same person may never actually utilize this method.

Many studies have asked people how they remember to carry out future actions. In general, it seems that external aids are used more frequently than internal aids (Harris, 1978; 1980; Intons-Peterson & Fournier, 1986; Kreutzer et al., 1975). There is also some evidence that older adults are more likely than younger adults to use external aids (e.g., Jackson, Bogers, & Kerstholt, 1988; Moscovitch & Minde, cited in Moscovitch, 1982). This might explain why older adults often outperform younger adults in real world prospective memory tasks.

Monitoring

Prospective memory monitoring is defined as remembering at the appropriate time that something has to be done, even if the person cannot remember what precisely has to be done. Conceptually, monitoring involves both the person's monitoring behavior while awaiting the correct circumstances to perform the prospective memory task, and the person's accuracy in identifying the circumstances appropriate for the action. Although accuracy is just the end point of the monitoring process, investigators have not always measured these two together. Consequently, they will be separated for the purposes of this review.

Monitoring accuracy. Investigators who treat prospective remembering as a unitary process have conceptually equated that process with monitoring accuracy. In these investigations, prospective memory performance is measured in terms of whether the person completes the appropriate action at the appropriate time. The components of prospective remembering are not differentiated. Other investigators have sought to differentiate between remembering to perform the future action (at the appropriate time) and remembering the content of the action. Both Dobbs and Rule (1987) and West (1988) looked at age differences in this way. Unfortunately, in both of these studies content recall is confounded with monitoring accuracy because a score on content recall was only possible after the person scored correctly on monitoring accuracy. It need not be the case that monitoring accuracy and content recall be confounded. In a recent study, Dobbs and Reeves (1991) measured content recall independently from monitoring accuracy. Dobbs and Reeves (1991) found that all of their participants were better at content recall than monitoring accuracy, that there was a decrease in prospective memory ability with age, but that there was no significant interaction between the two prospective memory components and age. The conclusions that can be drawn from these studies in terms of age differences are not yet clear. The number of studies that have separated monitoring accuracy from a global measure of prospective memory performance is very small. More studies need to be done looking at monitoring accuracy as a function of age.

Monitoring behavior. A prospective memory task is usually considered to have been performed incorrectly if it is performed at the wrong time. Consequently, when a person has to perform a prospective memory task in the future, it is important to monitor the environment so that when the right circumstances occur the prospective memory task can be performed promptly. Harris and Wilkins (1982) have proposed a model of how people monitor for the correct time or event after which to perform a future action. Important to this model is the idea that there is a critical period during which the action should be performed. Critical periods vary between prospective memory tasks in terms of their length and the nature of their boundaries. If the action is performed anytime during the critical period it is considered to have been performed on time. Harris and Wilkins refer to their model as the test-wait-test-exit model. According to this model people continually test for the circumstances under which to perform the future action. If the test is positive, the person exits the monitoring period. But if the test is negative, the person continues to wait and test. An important assumption of this model is that although testing is costly in terms of concurrent behavior, it is not as costly as missing the circumstances under which the prospective memory task was meant to be performed.

Investigations of the test-wait-test-exit model have been done by asking people to perform specific actions at certain future times. During the waiting interval the participants perform a background task such as playing a video game or watching a movie. Their monitoring behavior is measured by observing when, over the waiting interval, the participants check the time on a clock. Ceci and Bronfenbrenner (1985) found that their participants' clock checking patterns were indicative of three distinct monitoring strategies. Two of these strategies led to the prospective memory task being performed on time; the third did not. Of the two successful monitoring strategies, the most efficient (i.e., the one requiring the least number of clock checks and hence the one least disruptive to the concurrent task) was dubbed strategic time monitoring by the investigators. Strategic time monitoring produced a U-shaped clock checking distribution that Ceci and Bronfenbrenner (1985) interpreted as follows: During the beginning of the task the participants looked frequently at the clock in order to get their internal clocks in sync with the external clock. During the middle of the period they looked at the clock very infrequently because they could rely largely on their internal clocks to keep track of the time. Finally, as the response time approached, they relied mainly on the external clock to ensure that they were not late in performing the prospective memory task.

Ceci and Bronfenbrenner (1985) found a U-shaped distribution of clock checks when children monitored over one long interval. In contrast, Harris and Wilkins (1982) found a J-shaped distribution of clock checks when adults monitored over repeated short intervals. The reason for this discrepency may be due to the experience afforded Harris and Wilkins' (1982) participants over the repeated intervals or it may reflect a developmental difference such that it either takes longer for children to calibrate their internal clocks or it takes longer for them to become confident in the accuracy of their internal clocks (Ceci, Baker, & Bronfenbrenner, 1988). If there is a developmental difference it may be that older adults will be even better at setting their internal clocks, and/or will be more confident in their calibration than are younger adults (assuming that the younger adults have not achieved asymptotic performance).

Temporal calibration would not, of course, be the strategy of choice in all prospective memory tasks. When you need to give someone a message but you have no idea when this person will arrive, a test-wait-test-exit sampling strategy calibrated according to time would not be an efficient way to monitor. However, when you are monitoring for a certain time or for a temporally predictable event (e.g., when the gas gauge will be empty, when a person will finish walking across a room), a temporal calibration strategy is likely to ensure accurate performance.

<u>Content Recall</u>

As a prospective memory component, content recall refers to whether the correct action is carried out, regardless of when it is carried out. Very few studies have explicitly measured content recall in the context of prospective remembering. These studies have already been discussed in the section on monitoring accuracy. Although in most previous investigations it would probably have been possible to separate content recall from monitoring accuracy, it seems that many researchers have not realized the importance of making this conceptual distinction. An example should indicate why monitoring accuracy and content recall should be separated for study. If I have been asked to give a message to Karen the next time I see her, three different outcomes are possible: I might see Karen, remember I have to give her a message and remember what the message is (correct monitoring and correct content recall); or I might see Karen, remember that I have to give her a message but not remember what the message is (correct monitoring but incorrect content recall); or, finally, I might see Karen and not remember to give her the message but, if probed, I might have perfect recall of the content of the message (incorrect monitoring but correct content recall). Clearly, in order to investigate prospective memory most fruitfully the content of a prospective memory task can, and should, be separated out as a distinct unit of analysis.

Whether age differences will appear for the content of a prospective memory task depends on the nature on the content to be remembered. Content recall can be considered as the retrospective memory component of prospective memory. Consequently, it is likely that any manipulations that should produce age differences in the recall of a retrospective memory task, would also produce age differences in the recall of the content of a prospective memory task.

Compliance

Compliance is important to prospective remembering because people sometimes know perfectly well what action they are meant to perform, and they may know when to perform it, and yet they may simply decide not to perform it. To the lay-person, compliance may be the most important aspect of prospective remembering. Munsat (1966) points out that when a person forgets something he learned in the past, other people attribute this failure to his memory; but when a person forgets to do something in the future, other people attribute this failure to his character. Libraries and utility companies also seem to consider compliance an important aspect of remembering to carry out future actions because they punish people who fail to do so.

Compliance is probably the hardest component of prospective memory to interpret from the literature. This is because investigators have not measured compliance directly (i.e., by asking people if they did not perform an action because they did not want to). Instead, compliance has been studied by manipulating variables that "should" affect compliance and then seeing how these manipulations relate to performing the appropriate action. In this manner investigators have demonstrated that compliance can be affected by both personality variables (e.g., Lay, 1988; Searleman & Gaydusek, 1989; Wichman & Oyasato, 1983) and task demands (e.g., Meacham & Kushner, 1980; Meacham & Singer, 1977; Orne, 1970; Poon & Schaffer, 1982, cited in West, 1988; Somerville, Wellman & Cultice, 1983). However, very little research has looked at how variables that effect compliance interact with age. The only prospective memory study that seems to have investigated the relation between compliance and age was done by Poon and Schaffer (1982, cited in West, 1988) They found that the size of monetary payments affected older adults' performance but not younger adults' performance. In addition, although medical practioners have studied aging in relation to compliance for medical instructions (e.g., pill taking), it seems that what medical practioners refer to as compliance is what psychologists refer to as prospective memory rather than the prospective memory component of compliance (c.f., Feinstein, 1990). The medical literature will need to be carefully reviewed to see what predictions it can make in terms of age differences in prospective memory compliance.

<u>Output Monitoring</u>

The final component of this conceptual framework is output monitoring. Output monitoring refers to the person's knowledge of the completion status of the prospective memory task. Incorrect output monitoring can either result in a task being repeated or in it not being carried out at all. The importance of output monitoring to prospective memory tasks has been pointed out by Koriat and Ben-Zur (Koriat & Ben-Zur, 1988; Koriat, Ben-Zur, & Nussbaum, 1990; Koriat, Ben-Zur, & Sheffer, 1988); however, although Koriat et al. (1988) have studied age differences in output monitoring (young people were superior to old people), the tasks they have used have not been prospective memory ones. Other studies that have investigated prospective memory have produced results that can be interpreted in terms of output monitoring. Wilkins and Baddeley (1978) found that if their participants performed a prospective memory task, they remembered that they had done so, but if they failed to perform it, they were often unaware of this failure. In contrast, Maylor (1990) reported that her middle and older aged participants were aware both of when they had performed a prospective memory task and when they had failed to do so.

Output monitoring age differences in prospective memory tasks definitely need more study. It seems likely that older adults would perform worse than younger adults at output monitoring because they have been previously shown to be inferior when it comes to reality monitoring and to making repetitions on free recall and verbal fluency tests.

Paradigm for Studying the Components of Prospective Memory

A paradigm has been developed that will allow the different components of prospective memory to be tested within the same study and on the same participants. No study has ever before measured more than two of these components at the same time. The paradigm to be used in this study is an extension of one used by Maule to investigate visual sampling behavior (Maule, 1985; Maule & Sanford, 1980). Maule's task was a vigilance task and not a prospective memory task. Conceptually, the distinction between prospective memory and vigilance tasks seems to be that in a prospective memory task the person is always engaged in another task while awaiting the correct circumstances to perform the future action. For example, if I take a phone message for my officemate I do not sit around waiting for her to return before I engage in any other tasks. In reality, the interval between when I take the message and when I deliver the message is filled with one or more other tasks. Maule's basic paradigm has been converted to a prospective memory paradigm through the addition of a concurrent task.

The concurrent task to be used in this study is a lexical decision task. The advantage of using a lexical decision task is that it is a simple, continuous task with an accepted indicator of normal performance (i.e., faster responses to high than low frequency words). For the lexical decision task in this study, pronounceable nonwords will be used so that the participants have to make lexically-based decisions and, hence, so that the concurrent task will be cognitively demanding.

Similar to Maule's methodology, the participants in this study will have to monitor the status of three objects (vials). Each vial contains a falling pebble. When each pebble reaches a special area on its vial it should be reset back to the top of the vial. Each pebble will be falling at a different rate (and so each vial will need to be reset a different number of times). Relative to the total number of resets, 60% of the resets will occur one vial, 30% of the resets will occur on another, and 10% of the resets will occur on the third. The vials will be displayed by a computer equipped with a touch screen. In order to see each vial, the participants must press a button and in order to reset it they must press another button. In Maule's task each object had its own reset button. However, in this study, in order to obtain measures of content recall and output monitoring, 16 different reset buttons will be located around the outside perimeter of the screen. Any of the buttons can be used to reset whichever vial is currently visible. All of the reset buttons will have a word written on them. The words will come from four different categories (three target categories and one distracter category). The participants will be told that they should use words from category 1 to reset the first vial, from category 2 to reset the second vial, and from category 3 to reset the third vial. They should not reuse any of the reset words for a vial until all members from the corresponding category have been used.

This paradigm will allow age differences to be investigated for most of the components of prospective memory. Monitoring (i.e., remembering to respond) can be measured both in terms of accuracy in resetting each vial and by examining the vial checking behavior displayed by the young and old participants on each vial. Content recall (i.e., remembering how to respond) can be measured by seeing how often the participants respond with a word from the right category. Output monitoring (i.e., remembering if you have responded) can be assessed on-line in terms of whether the participants repeat any of the target words on each vial. Off-line measures of output monitoring can be taken by giving the participants a recognition test for the words they used and by having the participants recall when they checked the vials. Metaknowledge of the task demands will be assessed at the end of the study (after the participants have gained experience) by having people answer questions meant to tap their temporal, spatial and rate-based knowledge for the task. Planning will be measured by asking people to indicate which of a series of strategies they used to keep track of the pebbles' positions. Compliance will not be measured but it will be controlled by keeping the instructions constant between participants and by using similar participants (i.e., unpaid community-dwelling volunteers) in both age groups. On every component, except compliance, performance will be assessed in terms of age differences.

Within this paradigm and on the basis of past research it is possible to make several predictions regarding prospective memory performance. At high drop rates, a linearly increasing monitoring pattern is expected (Maule, 1985). This monitoring pattern might result either as a strategy to keep track of a quickly dropping pebble or because the monitoring function will represent the average of many individual drops. At slow drop rates (and also when the monitoring pattern is collapsed over only a few drops), people's monitoring patterns will be consistent with a J-shaped pattern (Ceci et al., 1988) or a step function (i.e., low undifferentiated monitoring rates throughout most of a drop followed by a high monitoring rate at the end of the drop) (Einstein et al., 1991; Maule, 1985; Reeves, Dobbs & Li, 1992). Age differences in monitoring pattern are expected such that older adults will monitor less than the younger adults in the area before the reset region (Einstein et al., 1991; Reeves et al., 1992). In addition, although both age groups are anticipated to monitor selectively (by devoting most of their checks to the fast vial and least to the slow), the younger adults are likely to be more selective (Maule & Sanford, 1980; Sanford & Maule, 1971; 1973). Younger adults are also expected to perform better on the lexical decision task when not overtly monitoring (Howard, Shaw & Heisey, 1986; Madden, 1988). Monitoring (total checks and location of checks) should be related to resetting accuracy. And older adults should make more late resetting errors than younger adults (Einstein et al., 1991; Maule & Sanford, 1980; Reeves et al., 1992). No hypotheses are made concerning early resets, however, because there has not been enough research to form a conclusion. In addition, no hypotheses are made concerning which age group is most likely to reset a vial with a word from the right category. When a word from the wrong category is used, it may be the result of an action slip such that a high frequency response is substituted for a low frequency response (cf., Reason, 1979) and therefore related to the frequency with which the categories are used correctly. Output monitoring errors will be more commonly produced by older adults than younger adults both on-line (while performing the task) and off-line (when asked to recall what they did) (Koriat et al., 1988). Regarding metaknowledge and planning, no a priori predictions are made. It seems reasonable, however, that people will either use spatial or temporal strategies to keep track of the pebbles' positions.

Method

Participants

Forty-eight community-dwelling adults volunteered to participate in this study after answering a newspaper ad. Half of the participants were young (M age = 29.5, range = 21-35) and half were old (M age = 65.38, range = 60-77). All of the participants rated their health as average or above average and none of the participants were taking drugs for a depression or psychosis. None of the participants were color blind, all of them were right-handed and they all had adequate vision to read the experimental stimuli. On average, the younger adults had a achieved a higher level of education (M = 14.62, range = 10-19) than the older adults (M = 12.92, range = 8-19), $t(46) = 2.25$, $p < .03$, but the two age groups did not differ reliably in vocabulary level (M = 11.25/18). More younger (M = 91.67%) than older adults (M = 58.33%) spoke English as their native language, $t(46) =$ 2.83, $p < .007$.

Procedure

Participant information. Upon arrival, the participants were given some general information about the study and asked to sign a consent form. Then they were asked to provide some demographic information such as their sex, age, and years of education. Their vocabulary level was assessed via the short version of Ekstrom, French, Harman, and Derman's (1976) advanced vocabulary test. Their eyesight was assessed by asking them to read lists of words presented on the computers in the same fonts as the experimental stimuli.

Vial prospective memory task. In the vial prospective memory task, the participants were presented with a computer representation of three vials, each of which contained a pebble falling at a different rate. Each of the vials could only be seen individually and not for more than 5 seconds at a time. In order to see each vial the participants had to press the box next to it marked "show". The participants' task was to monitor the three vials so that whenever a pebble entered the reset region (marked in blue) near the bottom of its vial, the pebble could be sent back to the top of the vial. The pebble was reset by pressing one of 16 computer-drawn boxes placed around the inside perimeter of the screen.

The 16 reset boxes each contained a different word. The words came from four different categories: 4 kinds of birds (targets for vial 1), 4 articles of clothing (targets for vial 2), 4 kinds of fruit (targets for vial 3), and 4 parts of a building (distracters). (The reset words were chosen from the most frequent members of each category in Battig & Montague, 1969). In order to reset the visible vial, the participants were asked to press a box containing a word from the appropriate target category. They were asked not repeat any of the reset words for a vial until all of the words from the target category had been used. Each time a vial was reset, the position of all the words was scrambled.

While the participants were monitoring the vials they were completing a continuous (i.e., no ISI) lexical decision task on another computer. The participants were asked to perform as well as they could on both the monitoring and the lexical decision task. There were 1,640 letter strings used in the lexical decision task. Half of the letter strings were words and half were nonwords. Of the words, half were of high frequency and half were of low frequency as listed in Kucera and Francis (1967). None of the words were the same or similar to those used on the reset boxes in the monitoring task. Pronounceable nonwords were created by changing one letter in a real word. (The nonwords looked and sounded like real words to ensure lexical processing of the letter strings.) The letter strings were presented randomly with the restrictions that within each block of 20 items (a) half of the items were words, (b) half of the words were of high frequency, and (c) no more than three successive items required the same (word/nonword) response.

The participants sat facing the lexical decision computer and made responses on a button pad. The monitoring task was presented on a Zenith VGA monitor placed to the left of the lexical decision computer. Responses to this computer were made using a Carrol Smart-Frame touch-sensitive screen. The trial continued for at least 20 minutes and until each vial had been reset at least 12, 6, and 2 times, respectively. The number of resets required on each vial was counter-balanced across participants.

While performing the task, the participants were asked to imagine that they were nurses and that the vials were intravenous bottles (IVs) hooked up to patients in three different rooms. The pebble indicated the height of the medicine within each IV. Each patient's medicine had to be changed (i.e, reset) whenever the pebble was within the reset region marked near the bottom of the IV. Changing an IV either before or after the pebble was in the reset region would seriously endanger the patient's health.

Recognition of the reset words. At the end of the vial monitoring task, the participants were read a list of the 16 words that appeared on the reset boxes and 16 semantically related foils. As with the reset words, the foils were chosen from the most frequent members of each category in Battig and Montague (1969). For each word the participants were asked to indicate how many times they had used that word to reset a vial and how many times they had used it on the correct vial.

Recollection of the monitoring pattern. The participants were also shown a life-sized drawing of the three vials. The area above each vial's reset band was divided up into four equal sized zones. The participants were asked to indicate how many times they checked a vial when the pebble was in each of the four zones above the reset band, how many times they checked when the pebble was in the reset band, and how many times they checked when the pebble was below the reset band.

Metaknowledge for the vial monitoring task. The participants were asked several questions to reveal what they had learned about the monitoring task. To assess the participants' temporal knowledge for the task they were asked to estimate both the time for each pebble to fall and the task length. To assess the participants' spatial knowledge they were given a life-sized drawing of the vials and asked to draw in the position occupied by each of the pebbles when the end-of-task alarm sounded. The participants' knowledge of the rates associated with each of the vials was measured by having them identify the fastest & slowest vials, by having them estimate how many of the first 10 resets occurred on each of the vials, and by having them estimate the total number of resets that occurred on each of the vials.

Planning for the vial monitoring task. At the end of the study, people were questioned about different strategies they might have used to keep track of the pebbles' positions. The strategies addressed were temporally-based, spatially-based, rate-based, non-differentiated, nonconscious, external reminder and consistency. The questions for each category are presented in Appendix A.

Results

Note: Analyses done to compare behavior between the three vials frequently violated analysis of variance's (ANOVA's) assumption of sphericity. When this happened Greenhouse-Geisser corrections were done and the adjusted degrees of freedom are reported.

Monitoring

Monitoring errors. There were two types of monitoring errors. Late errors occurred when the participant reset the pebble after it exited the reset region. Early errors occurred when the participant reset the pebble before it entered the reset region. Because of the sensitivity of the equipment (early errors could result if the participants allowed their fingers to hover too close to the reset boxes), early reset errors were only counted if they occurred more than half the pebble height above the reset region.

The mean number of monitoring errors per drop was assessed in a 2 (age) x 3 (drop rate) x 2 (error type) repeated measures ANOVA. The analysis revealed that older adults made significantly more resetting errors (of any kind) per drop (M = .210) than the younger adults did (M = .129), $F(1, 46) = 5.89$, $p = .019$. However, age did not interact with any of the other variables. There also were significant main effects of drop rate, $F(2, 71) = 4.09$, $p < .05$, and error type, $F(1, 46) = 4.65$, $p = .036$, but these were qualified by a drop rate x error type interaction, $F(2, 76) = 46.78$, $p < .01$. Newman-Keuls tests indicated that the number of early errors decreased as drop rate increased (slow $M = 0.344$, medium $M = 0.191$, fast $M = 0.094$, all means significantly different) but the number of late errors increased as drop rate increased (slow $M = 0.021$, medium $M = 0.069$, fast $M = 0.295$, fast mean significantly greater than medium and slow means). However, there was no difference between the largest means for each error type (early slow and late fast) nor between the smallest means for each error type (late slow and early fast).

Monitoring pattern. To evaluate monitoring patterns, the area above each reset region was divided into 4 equal zones. This technique is similar to the one used by Ceci and Bronfenbrenner (1985) to plot monitoring behavior across a time interval. In the present study, four zones were used for the pattern analysis so that both the hypothesized linear and quadratic trends could be investigated and also to allow for the possibility of a cubic monitoring pattern. Each zone was equal to about 21% of the vial height. (The reset region was equal to 14% of the vial height.)

The mean number of checks in each of the four zones was subjected to a 2 (age) x 3 (drop rate) x 4 (zone) repeated measures ANOVA. There were no reliable effects involving age; however, the analysis produced significant main effects of drop rate, $F(1, 49) = 169.21$, $p < .01$, and zone, $F(2, 90)$

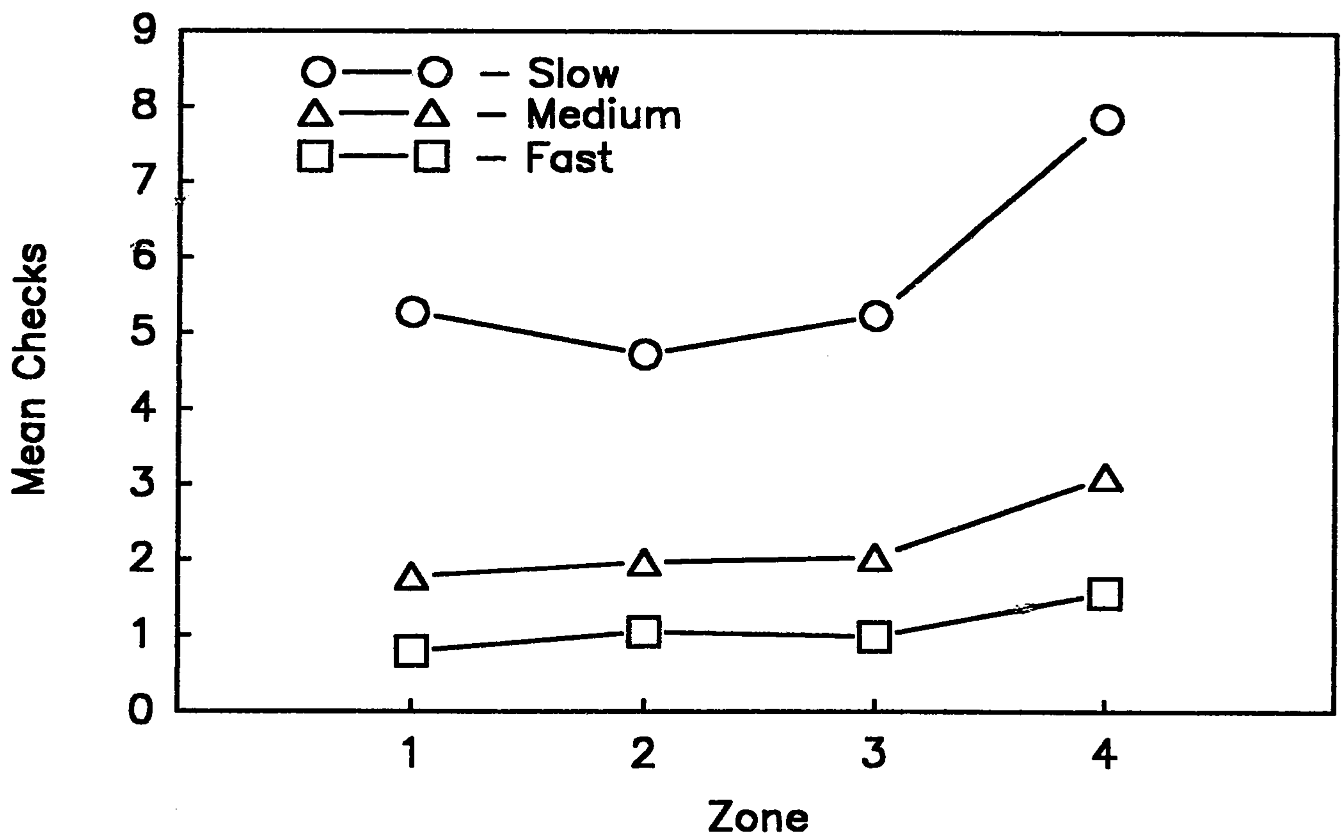
= 42.82, $p < .01$, and a drop rate by zone interaction, $F(2, 114) = 11.42$, $p < .01$. This interaction is displayed in Figure 1. Newman-Keuls tests on the interaction found four homogeneous groupings of the means: group 1 = zones 1 to 3 on the slow vial; group 2 = zones 1 to 3 on the medium vial and zone 4 on the fast vial; group 3 = zones 2 to 4 on the fast vial; group 4 = zones 1 to 3 on the fast vial. Whereas the means tend to be homogeneous within a vial, they are rarely homogeneous between vials. On the slow and medium vials the mean number of checks in zone 4 is significantly greater than the mean number of checks in zones 1 to 3. This monitoring pattern is best described as a step function that remains constant in zones 1 to 3 and then increases significantly in zone 4. In contrast, the pattern on the fast vial, where the means are harder to distinguish, is more consistent with a linearly increasing monitoring pattern with a small slope.

It was anticipated that the monitoring patterns might change between the first and second half of the study as a result of learning. Because the number of drops was so different between the three vials, separate 2 (age) x 4 (zone) x 2 (half) repeated measures ANOVAs were done on each vial to investigate learning. In these analyses only the effects involving half are of relevance.

On the fast vial there were no significant effects involving half. However, there was a half x zone

Figure 1

Checks per Zone as a Function of Drop Rate



interaction on the medium vial, $F(3, 138) = 7.34$, $p < .001$, and a half x zone trend, $F(3, 138) = 2.58$, $p = .056$, on the slow vial. The interactions for the slow and medium vials appear in Figures 2 and 3 respectively. Newman-Keuls tests on the slow vial indicated that the zone 4 means for each half did not differ; nor did the zone 1 to 3 means for each half. However, the means for zone 4 were significantly greater than the means for zones 1 to 3. This indicates that the same pattern was present for half 1 and half 2 and that the pattern is that of a step function which remains constant in zones 1 to 3 and then increases significantly in zone 4. Newman-Keuls tests on the medium vial found four groupings of means: group 1 = zone 4 in both halves; group 2 = zones 1 and 3 in half 1 and zones 2 and 3 in half 2; group 3 = zones 1 and 2 in half 1 and zone 3 in half 2; group 4 = zone 2 in half 1 and zones 1 and 3 in half 2. Although there are several different ways to combine the zones 1 to 3 means for each half (the half x cubic trend was significant, $p < .001$), the pattern is essentially the same as that on the slow vial. The means for zones 1 to 3 go together and are significantly smaller than the means for zone 4. Again this is indicative of a step function which increases in zone 4.

A caveat must be mentioned regarding the scoring of these data. It would have been better to have calculated the mean checks per zone based on the number of times the

Figure 2
Slow Vial: Checks per Zone as a Function of Study Half

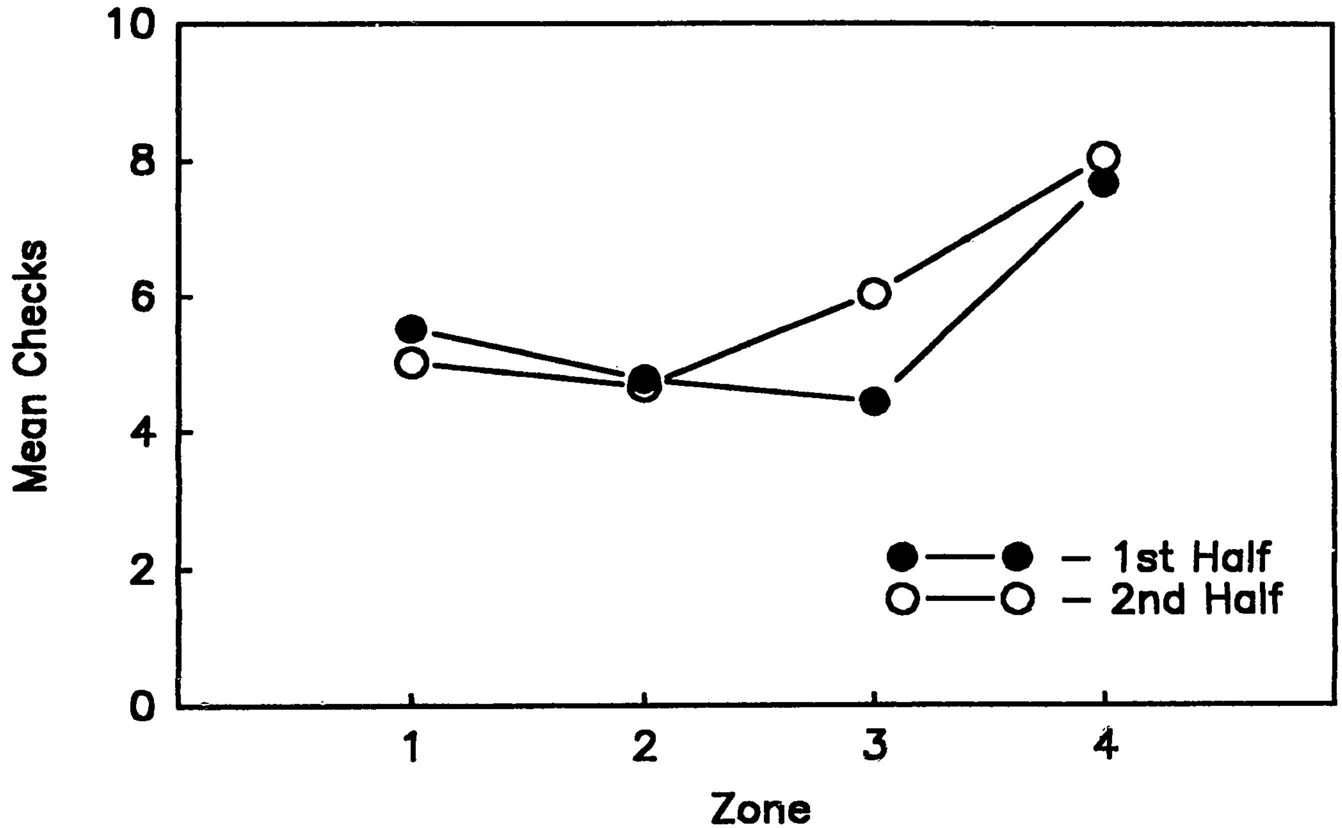
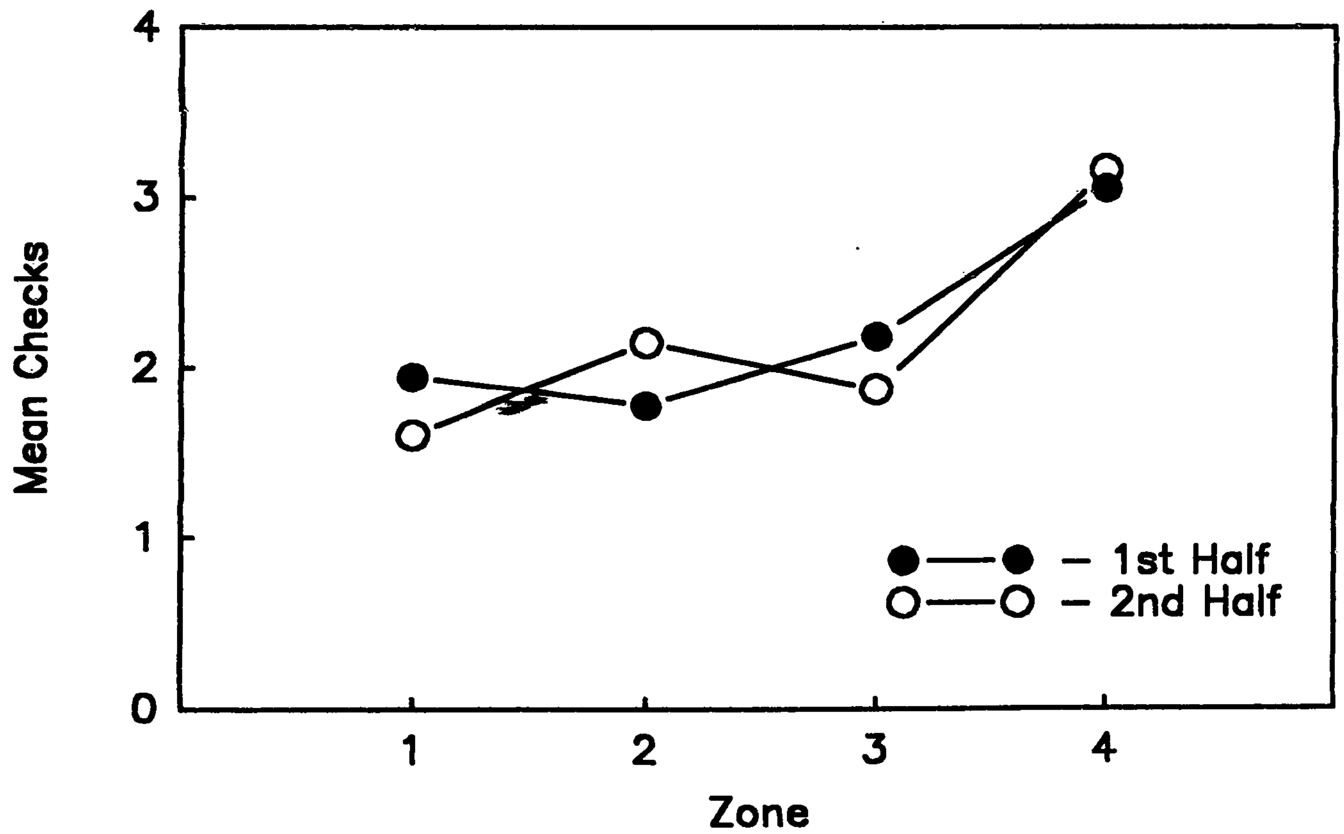


Figure 3

Medium Vial: Checks per Zone
as a Function of Study Half



participant was in each zone, rather than the number of drops the participant completed. This would have meant that when a participant reset a pebble in, for example, zone 1, the means for zones 2, 3 and 4 would not have been deflated by including zeroes in their calculations. However, this was not done as (because the slow pebble only dropped twice) it would have necessitated either scoring the slow vial differently from the medium and fast vials or throwing out 11 participants who reset before zone 4 on the slow vial.

Selective monitoring. The participants' abilities to selectively monitor vials with different drop rates were tested in a 2 (age) x 3 (drop rate) x 2 (half) repeated measures ANOVA on the total checks made to each vial. None of the effects involving age or half were reliable; however the drop rate main effect was significant, $F(1, 67) = 23.40$, $p < .01$. Planned comparisons done in a MANOVA confirmed that the participants devoted more of their looks to the medium vial ($M = 28.20$) than the slow vial ($M = 23.24$) and more of their looks to the fast vial ($M = 31.15$), than the medium vial (Fs reliable at $p < .002$). However, the proportion of looks devoted to each vial (slow = .28, medium = .34, fast = .38) was far below the level of optimum selectivity (.10, .30, .60).

Relation between monitoring behavior and errors. Correlations were calculated on each vial to investigate the relationship between vial checking and monitoring errors.

Because the participants had been told that both kinds of errors were equally deleterious, monitoring errors were not differentiated into early and late errors for these correlations.

On all three vials the same pattern of correlations was found. The total number of checks made by the participants was not related to their number of monitoring errors. However, the number of checks in zone 4 was always inversely correlated with monitoring errors (slow r = -.51, medium r = -.41, fast r = -.45, all p's < .01) and there were no significant correlations between monitoring errors and the number of checks in any of the preceding three zones. This is consistent with the results of the monitoring pattern analyses which found that (on the slow and medium vials, at least) people monitored differently in zone 4 than in zones 1 to 3. It also suggests that it is not the total number of checks that matters but when those checks are made.

Concurrent task performance. While monitoring the vials the participants performed a lexical decision task which was meant to keep them cognitively engaged. One way to assess if the participants were responding in this task in an appropriate manner is to evaluate whether word frequency had is usual effect when the participants were not overtly monitoring the vials. Both the 2 (age) x 2 (word frequency) repeated measures ANOVAs on decision times and on errors produced significant main effects of word frequency.

The participants were slower on low frequency words (M = 1306 ms) than high frequency words (M = 1084 ms), $F(1, 46) = 91.98$, $p < .001$, and they made more errors on low frequency words (M = 8.6%) than high frequency words (M = 1.9%), $F(1, 46) = 50.92$, $p < .001$. This confirms that the participants were responding in the lexical decision task in the usual manner.

Although there was a tendency for the younger adults to make faster decisions (M = 1142 ms) than the older adults (M = 1248 ms) this was not significant, $p = .089$. However, even though there were no age differences in the decision times or errors when people were not overtly monitoring the vials, the younger adults were able to complete significantly more of these trials (M = 554) than the older adults (M = 394) within a 21 minute interval, $t(46) = 3.57$, $p = .001$.

Interference between the vial task and lexical decision task. Correlations for each age group were done to see if there were any indications of interference between the two tasks when the participants were not overtly monitoring. Such interference might be an indication of how often the participant thought about monitoring (but did not actually monitor). Two types of correlations were done: one on the raw performance scores for each of the tasks, and one on people classified as below-average, average, or above-average on each of the tasks. Neither of these analyses could find any significant relationships between the two tasks.

However, there is one indication that the two age groups differentially sacrificed performance on one task in order to do well on the other. A larger percentage of the young adults' lexical decision trials (M = 88.13%) than the old adults' trials (M = 79.35%) did not contain overt monitoring, $t(46) = 2.86$, $p = .006$. A higher percentage of trials containing overt monitoring indicates that the older participants decided more often to sacrifice a decision time in order to monitor the vials.

Content Recall

Content recall (i.e., remembering how to respond) was assessed in a 2 (age) x 3 (drop rate) repeated measures ANOVA on mean category accuracy per drop. Averaged across drop rate, younger adults used the correct category on a larger proportion of drops (M = 0.882) than did the older adults (M = 0.697), $F(1, 46) = 11.01$, $p = .002$. No other effects were reliable.

Resetting errors (i.e., using a word from the wrong category to reset a vial) were assessed in a 2 (age) x 3 (drop rate) x 3 (frequency of error) repeated measures ANOVA. The main effect of age, indicating that older adults made more errors than younger adults was significant, $F(1, 46) = 11.01$, $p = .002$; however, because this is essentially the same finding as category accuracy it will not be discussed further. There was also a main effect of frequency (high frequency $M = 0.082$, low frequency $M = 0.084$, distracter $M = 0.045$), $F(2, 92) = 4.38$, $p = .015$. Planned comparisons (after equating for the number of words) indicated that slips to words that should be used on the other vials (high and low frequency) were more common than slips to words that should never be used (distracters), $p = .009$.

Output Monitoring

On-line output monitoring. An on-line output monitoring error occurred whenever the participants reused a category member before they had used up all the other members of the same category. The mean number of times an item was repeated, given that the correct category had been used, was assessed in a 2 (age) x 3 (drop rate) ANOVA. The analysis produced a main effect of drop rate, $F(2, 82) = 25.28$, $p < .001$, for which Newman-Keuls tests indicated that errors were more common on the medium ($M = .278$) and fast vials ($M = .355$) than on the slow vial ($M = .093$). The lower number of errors on the slow vial is uninteresting, however, because it would have been very difficult to have made a same category error on the slow vial (it could only happen if the person used the same word on both drops). Finally, this analysis also produced a main effect of age, $F(1, 41) = 4.10$, $p = .049$. However, because there were no meaningful effects involving drop rate, and because four old adults and one young adult had to be excluded from the above analysis (because they never used the correct category on the slow vial), a new analysis of age differences was done. Because this new analysis collapsed over drop rate, data could be included for all the participants. This analysis confirmed that older adults made more on-line output monitoring errors (M = .340) than did younger adults (M = .261), $t(46) = 2.32$, $p = .01$.

Recognition of the reset words. Another measure of output monitoring was obtained by asking the participants, once the monitoring task was over, to recognize the reset words they had used. As with on-line category accuracy, this also provides a measure of what they remembered doing. However, because this measure was taken after the task was completed I have referred to it as off-line output monitoring.

Age differences were examined in several different measures of off-line output monitoring. In all cases the analyses were done summing over legitimate categories (not distracters).

The first analysis was done to investigate what the participants correctly remembered doing. It was found that the older adults forgot more of the words they used (M = 31.13%) than did the younger adults (M = 8.55%), $t(46)$ = 4.80, $p < .001$.

A 2 (age) x 2 (intrusion type) repeated measures ANOVA was then done to investigate what the participants correctly remembered not doing. Two types of intrusions were compared: Intrusions in which the participants failed to reject words they had seen while monitoring but had not used to reset a vial and intrusions in which the participants failed to reject foils (i.e., high frequency members of the reset categories that were not seen while monitoring). The analysis revealed that the participants failed to reject more of the words seen but not used (M = 53.01%) than the words not seen (M = 13.30%), $F(1, 45) = 46.78$, $p < .001$. In addition, the older adults failed to reject more unused words (M = 39.76%) than the younger adults did (M = 26.27%), $F(1, 45) = 4.26$, $p = .045$. However, intrusion type did not vary as a function of age.

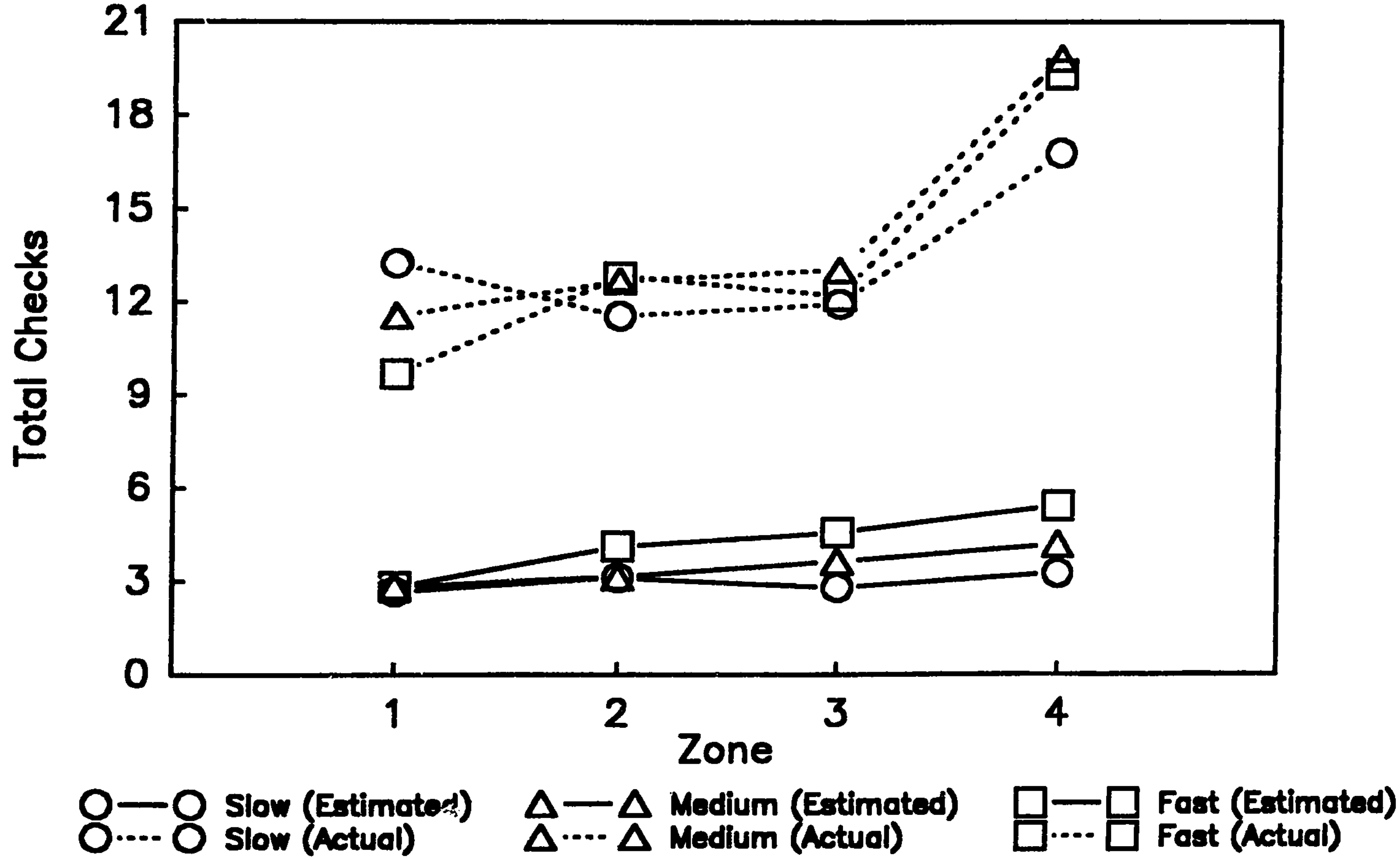
A final analysis based on this data set was done to see if the participants could remember how many times they had reset a vial with a category meant to be used on one of the other vials. A 2 (age) x 2 (response type) repeated measures ANOVA was done in which response type had the levels of estimated and actual errors. This analysis found main effects for age, $F(1, 46) = 15.00$, $p < .001$, and response type, $F(1, 46) = 70.98$, $p < .001$, and a reliable age x response type interaction, $F(1, 46) = 22.75$, $p < .001$. Newman-Keuls tests on the interaction indicated that there was no difference between the younger (M = .67) and older adults' estimates (M = .50). However, although both age groups thought they made a small and equivalent number of errors, the older adults actually made significantly more errors (M = 6.67) than the younger adults (M = 2.38) and both groups made many more errors than they thought they did.

Recollection of the monitoring pattern. A final measure of off-line output monitoring was obtained by asking the participants to recall how many times they checked a vial when the pebble was in each of the four zones above the reset region. The estimated versus actual number of checks in the whole study was analyzed in a 2 (age) x 2 (response type) x 3 (drop rate) x 4 (zone) repeated measures ANOVA. The variable of response type had two levels: estimated and actual checks. This ANOVA produced many significant effects, not all of which were meaningful after collapsing over response type. Other effects were qualified by reliable higher order interactions and will not be discussed separately (i.e., the main effect of response type, $F(1, 46) = 98.35$, $p < .001$, and the drop rate x response type, $F(1, 67) = 4.13$, $p < .05$, and response type x zone, $F(2, 80) = 26.52$, $p < .01$, interactions). Of the two remaining reliable effects, Newman-Keuls tests done on the age x response type interaction, $F(1, 46) = 5.85$, $p = .020$, were unable to differentiate between any of the means (young estimated M = 4.00, old estimated M = 3.11, young actual M = 11.69, old actual M = 15.75). Because the Newman-Keuls procedure looks for differences between pairs of means, the significant age x response type interaction must have resulted from a combination of means differing from another combination of means. However, because the exact nature of this interaction is unknown, the age x response type interaction will not be discussed further.

In this analysis, therefore, only the response type x drop rate x zone interaction, $F(4, 200) = 3.27$, $p < .05$, is of interest. This interaction is displayed in Figure 4. Newman-Keuls tests on the interaction indicated that the following groups of means are not significantly different: group 1 = zone 4 on the medium and fast vials for actual checks; group 2 = zones 1 to 3 on the slow and medium vials, and zones 2 and 3 on the fast vial, all for actual checks; group 3 = zone 4 on medium vial and zones 2 to 4 on the fast vial, all for estimated checks; group 4 = zones 2 to 4 on the slow vial, zones 1 to 4 on the medium vial, and zones 1 to 3 on the fast vial, all for estimated checks; group 5 = zones 1 to 4 on the slow and medium vials, and zones 1 and 2 on the fast vial, all for estimated checks. Taken as a whole, these groupings seem to indicate that the pattern of actual checks is very different from the pattern of estimated checks. For actual checks there is a clear distinction between the first three zones and the fourth. However, for estimated checks no clear distinctions between the means can be found. In conclusion it seems that both

Figure 4

Estimated vs. Actual Monitoring Patterns as a Function of Drop Rate and Zone



age groups greatly underestimated the number of checks they made and both age groups are unable to reproduce the observed differential monitoring pattern. This seems to indicate that the step function monitoring pattern is not the result of a conscious strategy.

Metaknowledge

The metaknowledge measures were given in questionnaire form and in many cases the participants had to give their responses separately for the slow, medium and fast vials. A problem with this is that not all of the participants were able to identify the slow or fast vials. For these participants, questions before the speed identification question (i.e., first 10 resets, estimated spatial positions, total checks) were scored according to the real slow, medium and fast vials. Questions after the speed identification question (i.e., time to fall, total resets) were scored in terms of which vials the participants identified as slow, medium and fast.

Metaknowledge for the monitoring task was assessed via temporal, spatial and rate-based measures. The results for each of these will be listed separately.

Temporal measures: Task length. In order to investigate the younger and older adults' abilities to estimate the length of the task, age differences in the ratio of estimated to actual task length were compared in a t-test. Although the older adults were significantly worse at approximating the length of the task (*M* = .45) than the younger adults (*M* = .60), $t(46) = 2.01$, $p = .05$, both age groups greatly underestimated the task length.

<u>Temporal measures: Time to fall</u>. The participants' estimates of the time it took each pebble to fall were compared by dividing the estimated drop time on each vial by the actual drop time. This provided a measure of how close each estimate was to the actual times and controlled for the different actual drop times on each vial. The ratio of estimated to actual drop times was assessed in a 2 (age) x 3 (drop rate) repeated measures ANOVA. Only the main effect of drop rate was significant, $F(1, 63) = 42.84$, $p < .01$. Participants became worse at estimating the drop time of the pebbles as the speed of the pebbles decreased (fast *M* = .648, medium *M* = .497, slow *M* = .272). Newman-Keuls tests confirmed that these three means were significantly different. However, on none of the vials were the participants very good at estimating the drop times and on all of the vials they underestimated the actual drop times.

<u>Spatial measure: Final position of the pebbles</u>. The final position of the pebbles was calculated as the percentage of the distance down the vial occupied by the center of each pebble at the end of the task. The difference between the actual and estimated final position of the pebbles was assessed in a 2 (age) x 3 (drop rate) repeated measures ANOVA. No significant effects were found.

However, although the means were quite small for each cell (range = 2.8-11.9%) the standard deviations were large (range = 19.8-42.2%).

Rate measures: Identifying the fast and slow vial. The number of participants who could identify the vials with the fastest and slowest drop rates was analyzed in a 2 (age) x 2 (rate) repeated measures ANOVA. The analysis produced a significant main effect of rate indicating that more people could identify the fast (M = 95.8%) than the slow vial (M = 75.0%), $F(1, 46) = 9.66$, $p = .003$. Neither the main effect of age nor the age x rate interaction was significant; however, these effects may have failed to reach significance due to a ceiling effect for the young adults knowing the fast vial. In an attempt to pull out any age differences, a t-test was done just on the slow vial data. This analysis also failed to reveal any age differences.

The fact that all participants were not at ceiling on these measures is somewhat surprising given the extreme differences in rates of fall between the three vials. It is also surprising that a correlation done on the older participants (one was not possible on the younger participants) failed to find any significant relation between people who were able to identify the fast vial and people who were able to identify the slow vial.

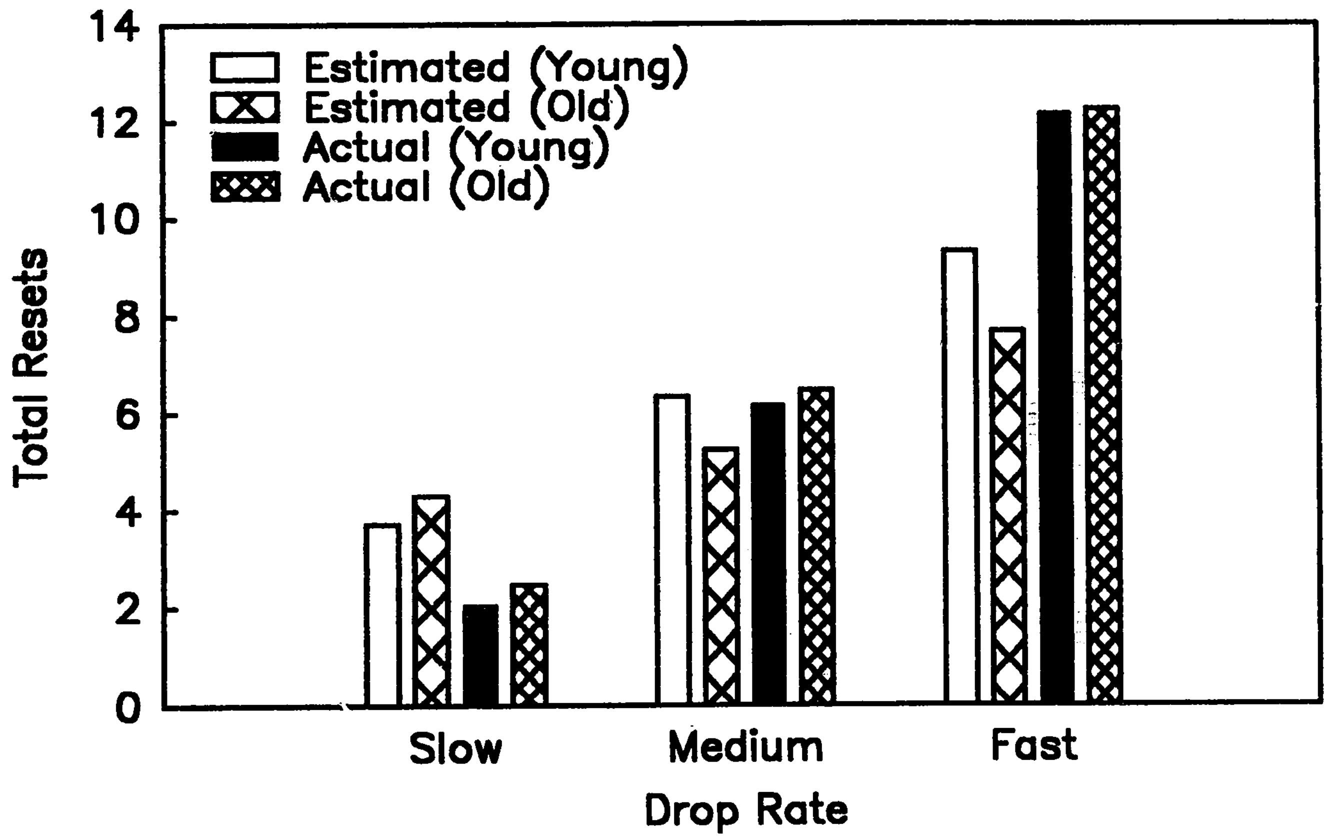
Rate measures: First 10 resets. The participants were asked to estimate how many of the first 10 resets occurred on each of the vials. It was hoped that this measure would provide a subjective approximation of the relative frequency with which each of the vials needed to be reset. Of interest was whether the participants could reproduce the actual distribution of 1-3-6.

The estimated number of the first 10 resets on each of the vials was analyzed in a 2 (age) x 3 (drop rate) repeated measures ANOVA. The main effect of drop rate was found to be significant, $F(1, 62) = 51.28$, $p < .01$. People thought that the most resets occurred on the fast vial ($M = 4.8$), followed by the medium vial ($M = 3.1$), followed by the slow vial ($M = 2.1$). Newman-Keuls tests confirmed that these means were significantly different. However, even though people knew that the most resets occurred on the fast vial and the least on the slow vial, the estimated distribution is quite different from the actual distribution of 1-3-6. People reduced the subjective ratios between the three vials by overestimating the number of resets on the slow vial and underestimating the number on the fast vial.

Finally, a t-test was done to compare the number of older and younger adults who could correctly state the distribution of the first 10 resets as 1-3-6. More younger ($M = 37.5\%$) than older adults ($M = 4.2\%$) could do this, $t(46) = 3.05$, $p = .004$.

Rate measures: Total resets. The participants' abilities to estimate the total number of resets in the

Figure 5
Estimated vs. Actual Total Resets
as a Function of Age and Drop Rate



whole study was assessed in a 2 (age) x 2 (response type) x 3 (drop rate) repeated measures ANOVA. The analyses produced a main effect of response type, $F(1, 46) = 4.25$, $p = .045$, and a response type x drop rate interaction, $F(2, 72) = 114.70$, $p < .01$; however, both of these are qualified by the reliable age x response type x drop rate interaction, $F(2, 72) = 3.79$, $p < .05$. This interaction is depicted in Figure 5. Newman-Keuls tests indicated that the actual number of resets did not vary as a function of age on any of the vials. Similarly, on the slow vial there was no age-related difference in the estimated number of resets. However, there was an age-related difference in estimated resets on both the medium and the fast vials such that the older adults estimate they made less resets than the younger adults. In addition, both age groups overestimate the number of resets they made on the slow vial and underestimate the number of resets they made on the fast vial. This suggests that people think the differences between the vials in terms of resets were much smaller than they actually were. Only on the medium vial does the estimated number of resets ever equal the actual number of resets and this is only true for the young adults.

Planning

The percentage of young and old participants using each of the strategies in the planning questionnaire is depicted in Table 1. Age differences were only apparent for the

Table 1
Percentage of young and old participants using each strategy

Strategy Type	Strategy	Young	Old	$t(46)$
Temporal	Estimate time	29.2 (46.4)	20.8 (41.5)	.66
Spatial	Use one pebble to estimate others	33.3 (48.2)	50.0 (51.1)	1.16
	Visually imagine	75.0 (44.2)	70.8 (46.4)	.32
Rate	Figure out fastest vial	83.3 (38.1)	70.8 (46.4)	1.02
	Figure out slowest vial	70.8 (46.4)	66.7 (48.2)	.31
Non-differentiated monitoring	Check all together	75.0 (44.2)	75.0 (44.2)	.00
Nonconscious monitoring	Forget & have need to check spring into awareness	8.33 (28.2)	41.7 (50.4)	2.83*
External reminder	Lexical decision reminder	12.5 (33.8)	8.33 (28.2)	.46
Consistency	Same strategy always	54.2 (50.9)	37.5 (49.5)	1.15

Note. Standard deviations are in brackets.

* $p = .007$

"strategy" of nonconscious monitoring. Nonconscious monitoring occurred when the need to monitor temporally left conscious awareness. The older adults were either more likely to seemingly forget about the monitoring task or more willing to admit that they had forgotten.

The most popular strategies (reportedly used by more than 68% of people) were figuring out the fastest and/or slowest vials (rate), visually imaging the position of the pebbles (spatial) and checking all the vials together (non-differentiated monitoring). The least popular strategies (reportedly used by less than 42% of people) were using one pebble to estimate the others' positions (spatial), estimating the time for one or more pebbles to fall (temporal), forgetting about monitoring but having the need to do so suddenly spring into awareness (nonconscious monitoring) and receiving an idiosyncratic prompt from the lexical decision task (external reminder). In addition, just over half of the participants changed their strategies during the task, presumably as a result of learning.

As one would expect, there was a significant correlation between the people who tried to identify the slow vial during the task and those who could successfully identify it at the end of the task, $r = .55$, $p < .001$. However, there was no comparable correlation for the fast vial.

<u>Correlations between the strategies</u>. Relations between the planning questions were investigated in 2-tailed correlations calculated separately for young and old adults. Only a few correlations were significant. Not surprising, the two rate questions were positively correlated for both young (r = .70, p < .001) and old adults (r = .52, p < .01): If people tried to figure out the fastest vial they also tended to try to figure out the slowest vial. In addition, there was a positive correlation between "nonconscious monitoring" and "external reminder" for the young adults (r = .80, p < .001). Both of these "strategies" were infrequently reported by younger adults and they are both quite anti-strategical. The items might correlate together because they identify individuals who are likely to receive nonconscious prompts. All other monitoring strategies were uncorrelated.

Discussion

If this prospective memory task had been scored in the usual manner, the conclusion would have been that the younger adults had outperformed the older adults. This conclusion would have been based entirely on the number of monitoring errors made by each age group. However, the data were analyzed in terms of many components which underlie prospective memory, not just one, and hence the age-related conclusions that can be drawn are not nearly so clear-cut. Indeed, the conclusions are much more interesting. The results on each component will now be discussed in turn.

Monitoring

Monitoring errors. The older adults in this study made more monitoring errors than the younger adults. Monitoring errors occurred either when the participants reset a pebble too early (before it entered the reset region) or too late (after it exited the reset region). The age differences in late resetting are consistent with many previous vigilance (Maule & Sanford, 1980) and prospective memory studies (Einstein et al., 1991; Kerr, 1992; Reeves et al., 1992). Age differences in early prospective memory performance extend the literature into a new area.

Most studies have not measured or reported the incidence of early performance errors. However, these errors should not be overlooked. In this study older adults were more likely than the younger adults to make both late and early resetting errors. In addition, an interaction was found between error type and speed indicating that early errors may be more common than late errors under some circumstances. Moreover, even though no age interaction was found in this study, it might be the case that the circumstances under which early or late errors are most common might be different for younger and older adults. This might be the case in real-life prospective memory studies where older adults are frequently reported to outperform younger adults. It would be interesting to discover in these studies if older adults are less likely to be late because they are more likely to be early.

Monitoring pattern. As hypothesized, the monitoring patterns produced by the participants in this study were influenced by the drop rates of the vials. At slow and medium drop rates the monitoring patterns are best described as a step function in which checking rates remain uniformly low until increasing significantly just before the pebble needs to be reset. In contrast, the monitoring pattern on the fast vial is suggestive of a linearly increasing function with a very small slope. Although these patterns are different from the U-shaped patterns found in Ceci and Bronfenbrenner's studies (Ceci et al., 1988; Ceci & Bronfenbrenner, 1985), the present study differs from those studies by using adult participants, scoring multiple trials, and requiring the participants to monitor more than one information source at a time. Other studies which have met one or more of these three qualifications have reported monitoring patterns similar to the ones found in the present study (Einstein et al., 1991; Maule, 1985; Reeves et al., 1992).

On the basis of past research (e.g., Einstein et al., 1991; Reeves et al., 1992), it had also been hypothesized that the younger adults would monitor more than the older adults in the zone immediately preceding the reset region. This hypothesis was not supported. In addition, there were no age differences in the number of times each age group checked the vials. However, recall that the older adults made more resetting errors than the younger adults. This means that even though both age groups monitored equally often and in the same places, the younger adults benefitted more from their monitoring than did the older adults.

Perhaps in order to reduce their number of resetting errors the older adults either need to monitor more in all zones or more in the zone preceding the reset region. (Possibly the latter is more likely because the number of checks in zone 4, but not the number of total checks, was related to monitoring accuracy.) If the older adults need to monitor more in order to maintain accuracy, it is not clear why they did not. Senders, Elkind, Grignetti & Smallwood (1966) suggest that source sampling occurs when uncertainty of a source's status exceeds a threshold. It is possible that the older adults in this study were overconfident and hence did not check as often as they needed to. However, this is contrary to the frequent finding that older adults are more cautious than younger adults (Schaie & Willis, 1986). Perhaps, instead, it is not that the older adults were less cautious but that they had a less realistic view of the task. Indeed, at the end of the study the older adults had a harder time than the younger adults recalling the frequency with which each vial was reset. In addition, more of the older adults than the younger adults reported that they completely forgot about the monitoring task for a time. It is therefore possible that more of the older adults had a hard time maintaining an internal representation of the pebbles' status. Finally, another possibility is that the older adults could not monitor enough because of a speed limitation. However, this suggestion is made cautiously because although older adults are generally found to be slower than younger adults on both motor (Kausler, 1982) and cognitive tasks (Light, 1991), the older adults in this study were not reliably slower on the concurrent lexical decision task.

Selective monitoring. Past studies have shown that people can selectively divide up their looks between three information sources dependent upon the probability with which each source needs to be reset (Hamilton, 1969; Maule & Sanford, 1980; Sanford & Maule, 1971; 1973). However, people will not monitor selectively unless there exist task demands (e.g., high reset rates, pacing) which force them to be strategic (Hamilton, 1969). In this study, the necessity of performing well on the concurrent task should have made people monitor selectively. Based on the results of vigilance studies (Maule & Sanford, 1980; Sanford & Maule, 1971; 1973) it had been hypothesized that both the younger and older adults in this study would selectively monitor the vials by devoting most of their checks to the fast vial and least to the slow vial. It was further hypothesized that the younger adults would be even more selective than the older adults.

Consistent with the vigilance studies, the participants in this study did selectivity monitor the vials. And the probabilities that the participants would observe each vial (.28, .34, .38) were very similar to those found in another study in our lab (.29, .34, .37) which used the same monitoring paradigm (Reeves et al., 1992). However, the probabilities in our lab's studies were much less differentiated than the optimal probabilities (.22, .32, .46) found by Hamilton (1969) for resets required in the same proportions (i.e., .10, .30, .60). There are several possible reasons the participants in this study were not more selective. One possibility is that the participants were not more selective as a result of the task demands. This explanation is based on Broadbent's (1971, cited in

Sanford & Maule, 1973) idea that people might either check a source purposefully (because it is time) or randomly (because they are unsure which source to check). The extra cognitive demands imposed by a concurrent task might increase the proportion of times when participants feel they should check but lose track of which source they should check. A higher proportion of random checks would lead to a decrease in selective monitoring. A second explanation for the low degree of selectivity is that the participants frequently chose to check all three vials together. (Three-quarters of the participants reported using this strategy.) To the participant's way of thinking, this might have been a popular strategy because three checks took only marginally more time than a single check, were arguably no more disruptive to the concurrent task, and were more likely to guard against late resets. If this explanation is true, moving the three vials to locations requiring different arm movements might cause the participants to abandon this strategy and hence become more selective. Finally, it must be remembered that the probability of observing each source is based on group means. If individual scores had been examined it would likely have been found that although some of the participants were not selective others were very selective.

Contrary to expectations, the younger adults were not more selective than the older adults. This is also consistent with a previous study done in our lab (Reeves et al., 1992). The reason why no age differences were found is not clear; however, Sanford and Maule (1973) found that younger (but not older) adults tended to become more selective as reset rates increased. In our studies the reset rates were far below the rates in any of the vigilance studies. It is therefore possible that age differences would have been observed with higher reset rates. However, caution must be taken whenever directly comparing a vigilance study to a dual-task study.

Content Recall

The participants were told to use an instance of the bird category to reset one vial, an article of clothing to reset another vial, and an type of fruit to reset the third vial. There were four items from each of these categories and four distracter items from a fourth category. The younger adults more frequently used a word from the right category to reset a vial than the older adults did. This age difference in content recall is consistent with some prospective memory studies (e.g., Dobbs & Reeves, 1991) but not with others (e.g., Dobbs & Rule, 1987). However, because content recall has been thought of a retrospective memory component of prospective memory, it might be better to compare these content recall results with the results of similar retrospective memory studies rather than with the results of all prospective memory studies. The observed age differences in content recall are consistent with retrospective memory studies which found that although both younger and older adults lose information from short term memory at the same rate, older adults always remember less than the younger adults (Kausler, 1982).

Errors in content recall were made when the participants reset a vial with a word from the wrong category. Whenever a vial was reset three incorrect categories were visible to the participants (one distracter category and two categories meant to be used with the other vials). It was hypothesized that the likelihood that each of these categories would be used in error would be related to the frequency with which it was meant to be used correctly. In this way content recall errors might be thought of as action slips which occur when high frequency actions are substituted for low frequency actions (cf., Reason, 1979). It is possible to think of the frequency of the error categories in two different ways. On the one hand, the distracter category was never meant to be used whereas the two other vial categories were meant to be used on their own vials. Therefore, one hypothesis would be that (after correcting for the number of categories) action slips would be more common to words in categories that were meant to be used (the other vial categories) than to words in the category that was never meant to be used (the distracter category). Another hypothesis would be that the frequency associated with each of the individual categories would determine how often each was used in error. The two other vial categories differ in the frequency with which they should be used on their own vials. Therefore, each can be thought of as either the high or low frequency category. Consequently, another hypothesis would be that action slips would be more common to the high frequency category items than to items of the low frequency category and more common to the low frequency category items than to the no frequency category items. The first hypothesis was confirmed: Slips to words in meant-to-be-used categories were more common than slips to words in the never meant-to-be-used category.

Output Monitoring

In addition to having to use a word from the correct category to reset a vial, the participants also had been instructed to use every item in the category before repeating an item. Younger adults were more successful at doing this than were older adults. This finding is consistent with Koriat et al.'s (1988) finding that older adults were more likely than younger adults to repeat words in a free recall task. The present study extended output monitoring age differences into the prospective memory realm.

The above measure has been referred to as on-line output monitoring because the data were collected on this measure as the participants were performing the vial task.

When the vial task was over, the participants were given a recognition test for the items they had used, and thus a measure of off-line output monitoring was also obtained. The off-line output monitoring measure indicated that the older adults forgot about a third of the words they had used whereas the younger adults only forgot about 9% of the words they used. This age difference, and even its magnitude, is consistent with the findings of Koriat et al. (1988, study 2) who also gave their participants a recognition test (for the items they had recalled). In addition to having more trouble in recognizing the words they had used, an investigation of intrusion errors further indicated that the older adults were less successful than the younger adults at correctly rejecting the words they had not used. Failing to reject a word that had been seen but not used was more common than failing to reject a word that had not been seen. However, intrusion type did not vary as a function of age.

One additional task was given which is related to output monitoring. The participants were shown a picture of the vials in which the areas above the reset regions were divided up into four equal-sized zones. The participants were asked to estimate how many times they had checked on a vial when its pebble was in each of those four zones. The analysis of these data indicated that not only do people greatly underestimate the number of times they checked each vial, but that people are also incapable of indicating how their monitoring pattern changed across the vial region. This suggests that the observed step function and linearly increasing monitoring patterns are more likely the result of automatic than conscious processes.

Metaknowledge

At the end of the vial task the participants were asked a series of questions to assess what they had learned about the task demands. These metaknowledge questions assessed temporal, spatial and rate-based knowledge.

The general conclusion that can be drawn from these measures is that although the younger adults sometimes were significantly better than the older adults, neither age group was able to demonstrate that they had learned much about the task demands. On the temporal measures both age groups greatly underestimated the length of the task and the time it took for the pebbles to fall. In addition, the participants performed much more poorly on the rate measures than expected. When asked how many of the first 10 resets occurred on each of the vials the participants were not good at reproducing the true frequencies of 1-3-6. This is surprising because Maule and Sanford (1980; Sanford & Maule, 1973) found that both their young and old participants could make quite accurate approximations. (Only a few of the participants in this study could correctly state all three frequencies and most of those were younger adults.) In addition, although all of Maule and Sanford's participants correctly rank ordered the reset frequencies, at least one-fourth of the participants in this study were unable to do this. Perhaps the reason why the participants in the present study had such poor knowledge of the reset frequencies can be attributable to a difference in attentional demands. Maule and Sanford's studies were vigilance tasks in which the participants were able to devote all of their attention to monitoring; however, in this study the participants had to divide their attention between monitoring and performing a lexical decision task. This might also explain why the participants had poor metaknowledge for all aspects of the task. (No conclusions could be drawn from the spatial measure due to the high degree of variability in the data.)

Not only did people have poor knowledge of the task demands but they expressed their poor knowledge by a tendency to reduce extreme variables. Not only did people underestimate the time of the task and the time for each pebble to fall, but they also underestimated the number of times they had checked each vial and the number of times they had used the wrong category to reset a vial. Some of these measures indicated that older adults underestimated more severely than younger adults. Moreover, the participants in this study often showed a tendency to underestimate the ratios between the three vials in their estimates. This was the case with the estimated breakdown of the first 10 resets and also with the estimated distribution of total resets. The older adults were especially likely to reduce the spread between the three estimated frequencies. The fact that people underestimated the ratios between the three vials indicates that they did not have a good conscious knowledge of the reset rates. This is supported by the fact that the participants were not nearly as selective in monitoring the vials as had been expected and, further, that many of the participants were not even able to rank order the vials by reset rate.

Planning

At the end of the task the participants were asked which of a series of strategies they had used to keep track of the pebbles. The strategies that were used by most of the participants were figuring out the fastest and/or slowest pebbles (rate), visually imaging the position of the pebbles (spatial) and checking all three vials together (non-differentiated monitoring). The strategies that were used least frequently were using one pebble to estimate the others' positions (spatial), estimating the time for one or more pebbles to fall (temporal), forgetting about monitoring but having the need to do so suddenly spring into awareness (nonconscious monitoring) and being prompted by the lexical decision task (external reminder). It had been expected that spatial and temporal would be the most likely strategies adopted. It was, therefore, interesting that although temporal seemed to be an unpopular strategy and the conclusion regarding spatial strategies is ambiguous, rate strategies (which combine spatial and temporal components) were very popular. Rate strategies also paid off in performance because people who did not use the strategy of trying to figure out the vial with the slowest drop rate were usually unable to identify the slow vial at the end of the task. The other popular strategy (non-differentiated monitoring) was somewhat surprising because it is a poor strategy in terms of optimizing the number of checks. However, as has been discussed before, it may be an attractive strategy to the participants because it maximizes the amount of information that can be obtained in one (brief) break from the lexical decision task.

Age differences appeared only for the "strategy" of nonconscious monitoring. This is indicative of the need to monitor leaving conscious awareness. Compared to the younger adults, the older adults were either more likely or more willing to admit that they had allowed the task to leave awareness. Interestingly, also, the few young adults who did report this happening also tended to report that they had received an idiosyncratic monitoring prompt from the lexical decision task. This also suggests that the need to monitor had left conscious awareness for a time.

Finally, just over half of the participants reported that they changed their strategies, presumably as a result of learning. This is somewhat surprising because no indication of learning was found for monitoring pattern or selectivity. However, these analyses looked for differences between the first and second half of the task. Perhaps learning effects might have been observed over a shorter time frame.

General Age Findings

The purpose of this study was to investigate age differences in the components underlying prospective memory. In the above paragraphs the findings have been discussed separately for each component. Now it is time to consider what the findings as a whole can tell us about age differences in prospective memory performance. Five different prospective memory components were investigated in this study; however, because the study was designed around the vial monitoring paradigm, better assessments could be taken of components that could be measured while the participants were monitoring (i.e., monitoring, content recall, and output monitoring) than components that had to be measured after the monitoring task was over (i.e., metaknowledge and planning). Because metaknowledge and planning were measured "off-line" and because no important age differences were found for either of these components, this general discussion of age differences will focus on monitoring, content recall and output monitoring.

The younger adults in this study outperformed the older adults on monitoring, content recall and output monitoring. For all of these components, the older adults made more errors than the younger adults. As has already been discussed, each of these findings is consistent with its own small body of research. The fact that older adults were deficient on all three of these components is also consistent with some self-report data collected by Lovelace & Twohig (1990). The older adults in Lovelace and Twohig's (1990) study reported that they had experienced an age-related increase in memory failures for remembering to bring up an intended point in conversation, remembering what they intended to do or say, and remembering whether or not they had done something. The first type of memory failure seems to be a monitoring failure, the second type a content recall failure and the third type an output monitoring failure. Thus, the self-perceptions of Lovelace & Twohig's (1990) participants are consistent with the present study's findings that there is an age-related decrease in accuracy on all three of these prospective memory components.

Although monitoring, content recall and output monitoring can be equated in terms of the observed age differences, the meaning and implications of these age differences are quite different. Monitoring the environment for a performance cue is the most distinctive component of prospective memory. In fact, many researchers consider the need to remember to respond as a defining distinction between prospective and retrospective memory (e.g., Winograd, 1988). In contrast, content recall bears many similarities to retrospective memory tasks in that, like in retrospective memory tasks, information is recalled after a response command. Finally, a failure in output monitoring (not remembering what you have and have not done) again seems to reflect a quite different kind of failure than a failure in content recall (not knowing what to do) or a failure in monitoring (doing something at the wrong time).

If prospective memory performance is measured only in terms of whether or not a task is completed, then the distinctions between these components are not important: Performance will only be assessed in terms of task success or failure. However, if prospective memory performance is looked at in terms of components, then it can be seen that the overall success or failure of the PM task is dependent on the success or failure of each component. The success or failure of each component, in turn, is influenced by a different set of task demands. Monitoring performance is probably influenced by the difficulty of the concurrent tasks and the distinctiveness and number of the monitored for cues. Content recall performance is probably effected by the length and complexity of the content. Finally, output monitoring performance probably varies depending on whether the task to be performed is habitual or episodic.

Prospective memory failures might occur because a person missed the signal to perform the task (e.g., did not look at the clock at 3 pm), forgot what to do (e.g., could not remember what to get at the store) or forgot if the task had been completed (e.g., could not remember if the front door had been locked). Because the older adults in this study were deficient on all three components, an implication of this research would be that older adults will be generally more likely than younger adults to make prospective memory failures. Indeed, this seems to be true for prospective memory studies conducted in the lab (e.g., Einstein et al., 1991, Study 1; Reeves et al., 1992; Schonfield & Shooter, cited in Harris, 1984; West, 1988, Study 2), but not for prospective memory studies conducted in the real world (e.g., Martin, 1986, Study 2; Moscovitch & Minde, cited in Moscovitch, 1982; Poon & Schaffer, 1982, cited in West, 1988). Perhaps the reason why older adults do better in real world studies is that in these studies people are able to design their own memory aids to compensate for their memory failures. The construction of memory aids depends on metaknowledge and planning, neither of which demonstrated meaningful age differences in this study. Perhaps, in daily life, older adults are more conscious of the possibility of memory failures and hence are better at compensating for them than are younger adults.

Conclusion

The goal of this study was to demonstrate the effectiveness of investigating prospective memory age differences in terms of underlying components. The study has demonstrated that age differences can be found in some of the components of prospective memory without being found in all components, that some measures within a component can reveal age differences when others do not (i.e., planning, metaknowledge), and that age differences can be found in outcomes without being found in the associated processes (i.e., monitoring). All of this supports the need to study prospective memory not in terms of the ability to perform one task but rather in terms of performance on each of the components of the task. More research needs to be done both with this paradigm and with others before any firm conclusions can be made regarding age differences in prospective memory performance. This study has provided a start.

References

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. Journal of Experimental Psychology Monograph, 80(3, part 2).

Beal, C. R. (1985). Development of knowledge about the use of cues to aid prospective retrieval. Child Development, 56, 631-642.

Ceci, S. J., Baker, J. G., & Bronfenbrenner, U. (1988). Prospective remembering, temporal calibration, and context. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory: Current research and issues (Vol. 1, pp. 360-365). Chichester: John Wiley & Sons.

Ceci, S. J., & Bronfenbrenner, U. (1985). "Don't forget to take the cupcakes out of the oven": Prospective memory, strategic time-monitoring, and context. Child Development, 56, 152-164.

Devolder, P. A., Brigham, M. C., & Pressley, M. (1990). Memory performance awareness in younger and older adults. Psychology and Aging, 5(2), 291-303.

Dobbs, A. R., & Reeves, M. B. (1991, June). Prospective memory, aging, and the workplace. Paper presented at the annual meeting of the Canadian Psychological Association, Calgary, Alberta.

Dobbs, A. R., & Rule, B. G. (1987). Prospective memory and self-reports of memory abilities in older adults. Canadian Journal of Psychology, 41(2), 209-222.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. Journal of Experimental Psychology: Learning, Memory, & Cognition, 16(4), 717-726.

Einstein, G. O., McDaniel, M. A., Cunfer, A. R., & Guynn, M. J. (1991). Aging and prospective memory: Examining the influences of self-initiated retrieval processes and mind wandering. Unpublished manuscript, Furman University.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). Manual for Kit of Factor Referenced Cognitive Tests. Princeton, NJ: Educational Testing Service.

Feinstein, A. R. (1990). On white-coat effects and the electronic monitoring of compliance. Archives of Internal Medicine, 150, 1377-1378.

Hamilton, P. (1969). Selective attention in multisource monitoring tasks. Journal of Experimental Psychology, 82, 34-37.

Harris, J. E. (1978). External memory aids. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory (pp. 172-179). London: Academic Press.

Harris, J. E. (1980). Memory aids people use: Two interview studies. Memory and Cognition, 8, 31-38.

Harris, J. E. (1984). Remembering to do things: A forgotten topic. In J. E. Harris & P. E. Morris (Eds.), Everyday memory, actions and absent-mindedness, (pp. 71-91). New York: Academic Press.

Harris, J. E., & Wilkins, A. J. (1982). Remembering to do things: A theoretical framework and an illustrative experiment. Human Learning, 1, 123-136.

Howard, D. V., Shaw, R. J., & Heisey, J. G. (1986). Aging and the time course of semantic activation. Journal of Gerontology, 41(2), 195-203.

Intons-Peterson, M. J., & Fournier, J. (1986). External and internal memory aids: When and how often do we use them? Journal of Experimental Psychology: General, 115(3), 267-280.

Istomina, Z. M. (1986). The development of voluntary memory in children of preschool age. In U. Neisser (Ed.), Memory observed: Remembering in natural contexts, (pp. 349-365). San Francisco: Freeman.

Jackson, J. L., Bogers, H., & Kerstholt, J. (1988). Do memory aids aid the elderly in their day to day remembering? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory: Current research and issues (Vol. 2, pp. 137-142). Chichester: John Wiley & Sons.

Kausler, D. H. (1982). Experimental psychology and human aging. New York: John Wiley & Sons.

Kerr, S. (1992, April). Prospective memory and aging: Older subjects burn their breakfast in a simulated cooking task. Poster presented at the Fourth Biennial Cognitive Aging Conference, Atlanta, Georgia.

Koriat, A., & Ben-Zur, H. (1988). Remembering that I did it: Processes and deficits in output monitoring. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory: Current research and issues (Vol. 1, pp. 203-208). Chichester: John Wiley & Sons.

Koriat, A., Ben-Zur, H., & Nussbaum, A. (1990). Encoding information for future action: Memory for to-be-performed versus memory for to-be-recalled tasks. Memory and Cognition, 18, 568-578.

Koriat, A., Ben-Zur, H., & Sheffer, D. (1988). Telling the same story twice: Output monitoring and age. Journal of Memory and Language, 27, 23-29.

Kreutzer, M. A., Leonard, C., & Flavell, J. H. (1975). An interview study of children's knowledge about memory. Monographs of the Society for Research in Child Development, 40, (1, Serial No. 159).

Kucera, H., & Francis, W. N. (1967). Computational analysis of present-day American English. Providence, RI: Brown University Press.

Lay, C. H. (1988). Procrastination and everyday memory. In M. M. Gruneberg, P. E., Morris, & R. N. Sykes (Eds.), Practical aspects of memory: Current research and issues (Vol. 1, pp. 453-458). Chichester: John Wiley & Sons.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. Annual Review of Psychology, 42, 333-376.

Lovelace, E. A., & Twohig, P. T. (1990). Healthy older adults perceptions of their memory functioning and use of mnemonics. Bulletin of the Psychonomic Society, 28(2), 115-118.

Madden, D. J. (1988). Adult age differences in the effects of sentence context and stimulus degradation during visual word recognition. Psychology and Aging, 3(2), 167-172.

Martin, M. (1986). Ageing and patterns of change in everyday memory and cognition. Human Learning, 5, 63-74.

Maule, A. J. (1985). The importance of an updating internal representation of the environment in the control of visual sampling. The Quarterly Journal of Experimental Psychology, 37A, 533-551.

Maule, A. J., & Sanford, A. J. (1980). Adult age differences in multisource selection behaviour with partially predictable signals. British Journal of Psychology, 71, 69-81.

Maylor, E. A. (1990). Age and prospective memory. The Quarterly Journal of Experimental Psychology, 42A, 471-493.

McMillan, T. M. (1984). Investigation of everyday memory in normal subjects using the subjective memory questionnaire (SMQ). Cortex, 20, 333-347.

Meacham, J. A., & Kushner, S. (1980). Anxiety, prospective remembering, and performance of planned actions. Journal of General Psychology, 103, 203-209.

Meacham J. A., & Singer, J. (1977). Incentive effects in prospective remembering. The Journal of Psychology, 97, 191-197.

Miller, J. A. (1990). Exploratory studies of prospective memory in adults. Unpublished doctoral dissertation, University of British Columbia, Vancouver.

Moscovitch, M. (1982). A neuropsychological approach to perception and memory in normal and pathological aging. In F. I. M. Craik & S. Trehub (Eds.), Advances in the study of communication and affect: Vol. 8. Aging and cognitive processes (pp. 55-78), New York: Plenum Press.

Munsat, S. (1966). The concept of memory. New York: Random House.

Orne, M. T. (1970). Hypnosis, motivation and the ecological validity of the psychological experiment. In W. J. Arnold & M. M. Page (Eds.), Nebraska symposium on motivation. Vol. 18: Current theory and research in motivation (pp. 187-265). Lincon: University of Nebraska Press.

Reason, J. (1979). Actions not as planned: The price of automatization. In G. Underwood & R. Stevens (Eds.), Aspects of Consciousness, Vol. 1: Psychological issues (pp. 67-89). New York: Academic Press.

Reeves, M. B., Dobbs, A. R., & Li, K. Z. H. (1992, April). Age differences in the prospective monitoring of temporally predictable events. Poster presented at the Fourth Biennial Cognitive Aging Conference, Atlanta, Georgia.

Sanford, A. J., & Maule, A. J. (1971). Age and distribution of observing responses. Psychonomic Science, 23, 419-420.

Sanford, A. J., & Maule, A. J. (1973). The allocation of attention in multisource monitoring behaviour: Adult age differences. Perception, 2(1), 91-100.

Schaie, K. W., & Willis, S. L. (1986). Adult development and aging (2nd ed.). Boston: Little, Brown & Co.

Searleman, A., & Gaydusek, K. A. (1989, November). Relationship between prospective memory ability and selective personality variables. Paper presented at the 30th annual meeting of the Psychonomic Society, Atlanta, Georgia.

Senders, J. W., Elkind, J. I., Grignetti, M. C., & Smallwood, R. (1966). An investigation of the visual sampling behaviour of human observers. Cambridge, Mass: Bolt, Beranek and Newman, Report No. 1246.

Somerville, S. C., Wellman, H. M., & Cultice, J. C. (1983). Young children's deliberate reminding, Journal of Genetic Psychology, 143, 87-96.

West, R. L. (1988). Prospective memory and aging. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory: Current research and issues (Vol. 2, pp. 119-125). Chichester: John Wiley & Sons.

Wichman, H., & Oyasato, A. (1983). Effects of locus of control and task complexity on prospective remembering. Human Factors, 25, 583-591.

Wilkins, A. J., & Baddeley, A. D. (1978). Remembering to recall in everyday life: An approach to absent-mindedness. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), Practical aspects of memory (pp. 27-34). London: Academic Press.

Winograd, E. (1988). Some observations on prospective remembering. In M. M. Gruneberg, P. E., Morris, & R. N. Sykes (Eds.), _Practical aspects of memory: Current research and issues_ (Vol. 1, pp. 348-353). Chichester: John Wiley & Sons.

Appendix A

Planning Questions

Notes: IV = vial; indicator = pebble

Temporal

"Did you try to estimate the time it would take for one or more of the indicators to enter its blue reset band?"

Spatial

"Did you use an indicator's position in one of the IVs to estimate the indicators' positions in the other 2 IVs?"

"Did you try to visually imagine the position of the indicators when you were not looking at them?"

Rate

"During the task, did you try to figure out which indicator was falling the fastest?"

"During the task, did you try to figure out which indicator was falling the slowest?"

Non-differentiated Monitoring

"Did you tend to check all 3 IVs together? (i.e., when you checked one did you also tend to check the other two?)"

Nonconscious Monitoring

"Did you frequently seem to forget completely about the IVs only to have the need to check an IV suddenly spring into awareness?"

External Reminder

"Did the words in the lexical decision task ever remind you to check the IVs?"

Consistency

"Did you use the same IV checking strategy throughout the whole task?"