TO PLAN OR NOT TO PLAN: AN EXAMINATION OF PLANNING IN EVERYDAY ACTION

A Dissertation
submitted to the
Temple University Graduate Board

---

in partial fulfillment of the
requirements for the degree
DOCTOR OF PHILOSOPHY

By
Colette Seter
January, 2014

Examining Committee Members:
Tania Giovannetti, Advisory Chair, Clinical Psychology
Lauren Alloy, Examining Chair, Clinical Psychology
Jason Chein, Brain and Cognitive Sciences
Richard Heimberg, Clinical Psychology
Nancy Minniti, Temple University Hospital
Robert Weisberg, Brain and Cognitive Science

ABSTRACT

To Plan or Not to Plan: An Examination of Planning in Everyday Action

Colette Seter

Degree: Doctor of Philosophy

Temple University, 2013

Doctoral Advisory Committee Chair: Tania Giovannetti, Ph.D.

Everyday activities are necessary for independent and productive living, and errors in everyday tasks are associated with a multitude of negative consequences, from increasing stress and frustration to serious safety concerns. Current rehabilitation strategies for improving everyday functioning focus on improving deliberate planning of everyday tasks, however many fundamental questions remain regarding everyday action planning. Few studies have examined both plan formulation and plan execution during everyday task performance, included multiple traditional neuropsychological planning measures, and evaluated competing neurocognitive models of planning in one study. This study addressed several gaps in the literature by examining the extent to which individuals planned before beginning an everyday task and whether planning facilitated performance. Additionally, the study was designed to identify optimal measures of planning abilities and the neurocognitive processes that are crucial for planning skills. A sample of 92 healthy participants completed complex everyday tasks (2x3 Multi-Level Action Test; Buxbaum et al., 1998; Schwartz et al., 1998) as well as a neuropsychological battery consisting of traditional neuropsychological tests of planning (e.g., Tower Test; Delis et al., 2001) and executive functioning (e.g., Haylings Test; Burgess & Shallice, 1997), episodic memory (e.g., WAIS- IV Logical Memory; Wechsler, 2009a), and working memory (e.g., Automated Symmetry Span; Barch et al., 2009). Contrary to hypotheses, deliberate planning prior to a task did not improve performance, traditional neuropsychological measures were not significantly related to naturalistic planning variables, and neither executive functions nor episodic memory were strongly associated with planning skills. The results suggest that investigators must use caution when selecting planning variables for research and when drawing conclusions about everyday functioning from traditional neuropsychological planning measures. Further research is also needed to expand current neurocognitive models of planning to account for performance on complex everyday tasks.

ACKNOWLEDEGMENTS

I would first and foremost like to thank my advisory chair and mentor, Dr. Tania Giovannetti, for her guidance, patience, and unconditional support. She embodies all qualities of a great mentor, and without her teaching, encouragement, knowledge, and investment in my training and success, this project and my professional accomplishments would not be possible. I would also like to thank Dr. Lauren Alloy, Dr. Jason Chein, and Dr. Richard Heimberg for their scholarly input and helping me shape this study, as well as Dr. Nancy Minniti and Dr. Robert Weisberg for their support and for serving on my committee. I have been fortunate to learn and train at Temple University. I would also like to thank and acknowledge my invaluable research assistant and Temple Undergraduate, Dylan Ottemiller who dedicated countless hours to data collection and analysis. Finally I would like to thank my husband Brenton Elliott, and my parents Louise and Setko Seter, and my brother Roy Seter who have loved, encouraged, and supported me every step of the way.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Background and Significance

Planning is considered an executive function associated with the prefrontal cortex and its projections and defined as the ability to think ahead (Owen, 1997; Shallice & Burgess, 1991; Sohlberg & Mateer, 2001; Unterrainer & Owen, 2006). It is regarded as a necessary component of many cognitive and motor tasks (Owen, 1997). Luria (1978) defined planning as the "ability to organize behaviour in relation to a specific goal that must be achieved through a series of intermediate steps" (Allain et al., 2005, p. 4). More recently, authors have similarly defined planning as the process of considering the future or the consequences of actions (Burgess & Robertson, 2002), and as the anticipation of future events and goal attainment (Carlin et al., 2000). Unterrainer and Owen (2006) use the term "forward thinking" to denote planning. Consistent with prior work, this paper will conceptualize planning as forward thinking, or the deliberate consideration of an end state before the implementation of steps geared toward achieving that end state.

There are numerous reasons planning research is important. First, planning deficits are common among many patients with cognitive impairment, and neurological damage or disease. Caregivers of mixed-etiology neurological patients reported planning problems most frequently among 20 common dysexecutive symptoms (Burgess, Alderman, Evans, Emslie, & Wilson, 1998). Alarmingly, relatively few patients reported these planning problems, with the highest rate of caregiver-patient disagreement for planning problems relative to other difficulties (e.g., unconcern for social rules, etc.; Burgess & Robertson, 2002).

Second, planning deficits are believed to have a negative impact on everyday functioning (Penfield & Evans, 1935; Shallice & Burgess, 1991). There are numerous case descriptions of functional planning deficits in the literature. Luria (1980) described serious planning failures in patients with large lesions of the frontal lobes that resulted in fragmented action sequences precluding completion of most multiple-step everyday tasks. Penfield and Evans (1935) described a premorbidly high functioning woman who after neurosurgical excision of the frontal lobe (due to a right frontal glioma) could not organize and plan her daily activities (i.e. she could cook dishes individually, but she could not successfully plan and prepare an entire family meal (Shallice & Burgess, 1991; Unterrainer & Owen, 2006)). Investigators also have attempted to quantify functional planning deficits using standardized assessments. For instance, Shallice and Burgess (1991) observed everyday functioning deficits on a complex naturalistic task in three patients with frontal lobe damage, who showed minimal or no impairment on intellectual or neuropsychological tests. These patients exhibited a specific deficit in the ability to prioritize, organize, and deliberately execute a series of goals; thus, these patients had difficulty with everyday tasks that required planning/forward thinking (Burgess, Veitch, Costello, & Shallice, 2000). Schwartz and colleagues (1998) reported that compared to healthy controls, individuals with closed head injury were unable to successfully ration a limited supply of materials (e.g., not leaving enough wrapping paper to wrap a second gift). Furthermore, several investigators have shown an association between performance on neuropsychological tests of planning and everyday functioning in schizophrenia (Evans et al., 2003; Harvey, Green, Keefe, & Velligan, 2004; Kessler, Giovannetti, & MacMullen, 2007; Klapow et al., 1997; Patterson et al., 1998; Patterson, Goldman, McKibbin, Hughs, & Jeste, 2001; Semkovska, Bedard, Godbout, Limoge, & Stip, 2004; Semkovska, Stip, Godbout, Paquet, & Bedard, 2002; Seter, Giovannetti, Kessler, & Worth, 2011; Sevy & Davidson, 1995; Velligan et al., 2007). Thus, the study of planning clearly has great significance for optimal everyday functioning.

Finally, planning research is important because greater understanding of planning deficits may elucidate intervention strategies to promote everyday functioning and independent living. In fact, planning deficits are often the focus of cognitive rehabilitation approaches (e.g. Goal Management Training; Burgess & Robertson, 2002; Lawson & Rice, 1989; Levine et al., 2000; Sohlberg & Mateer, 2001). We reason that the careful characterization of planning within ecologically valid, everyday tasks might offer clear directions for future rehabilitation strategies for improving everyday functioning.

Theoretical Models of Planning

Planning is generally conceptualized as a double-level process involving *formulation* and *execution* (Grafman, 1989; Shallice, 1982). The formulation level entails the ability to mentally develop a strategy to predetermine the actions aimed at achieving a goal (i.e., the ability of forward thinking). At this level, a *plan* is formulated. The execution level involves competence in monitoring the execution of the plan to a successful outcome or end state (Allain et al., 2005). Two prominent neurocognitive models of planning-Contention Scheduling/Supervisory Attentional System and the Constructive Episodic Simulation Hypothesis- provide different accounts of the processes involved in forward thinking. Although both of these models have been supported in the literature, there is no research directly comparing these two theories and the models have been applied to planning in everyday tasks in a limited way.

The Contention Scheduling/Supervisory Attentional System (CS/SAS) model of planning (Shallice & Burgess, 1991) is based on theories of attention control and proposes that planning processes involve *executive functions* (i.e., higher-order cognitive processes). This *executive account of planning* posits that behavior in routine situations is governed by a relatively automatic control system, contention scheduling (CS), whereas a second system, called the supervisory attentional system (SAS) is necessary when deliberate *planning* is needed because there are no existing schema (i.e., knowledge representations) to guide behavior or because existing schema are not appropriate for the current task context (i.e. the task or situation is novel, unfamiliar, or there is no pre-established plan). A five-stage model of planning has developed from this executive position, with the stages of planning being defined as: goal articulation, provisional plan formulation, marker creation, marker triggering, and monitoring (Shallice & Burgess, 1991).[1] Stages 1 to 3 of this model encompass the formulation level, with stages 4 and 5 capturing the execution level of planning.

Another neurocognitive model of planning moves away from this executive function conceptualization of planning ability, and stresses the role of *episodic memory*. The constructive episodic simulation hypothesis (Schacter & Addis, 2007; Schacter, Addis, & Buckner, 2007) of planning ability proposes that remembering the past is necessary to imagine the future. This *episodic memory account* of planning posits that the simulation of future episodes and planning requires a flexible system that can recombine

---

[1] This model proposes that the initial component of planning requires the articulation of a goal. Upon identifying this goal, one must think ahead to the steps that are essential for goal attainment and in doing so, formulate a plan. In plan formulation/forward thinking one also develops markers that will aid and trigger plan execution. Finally, once the initial planning is complete, overt behavior (i.e., execution) is strongly influenced by the activation of markers and the monitoring of performance.

details from the past (Schacter et al., 2007). This hypothesis is based on the notion that episodic memory is a constructive rather than a reproductive process (Bartlett, 1932) and suggests that an important function of memory is to make information available for simulation of future events (e.g. future thinking).

Traditional Neuropsychological Tests Used to Evaluate Planning

Traditional clinical measures of planning abilities have been influenced largely by the executive theories of planning. These measures also involve abstract problems or scenarios; evaluation of planning in more ecologically valid contexts has been limited. The Tower of London task (TOL; Shallice, 1982) has been the most widely researched clinical measure of planning. The TOL, and related variants such as the *Tower Test* (Delis, Kaplan, & Kramer, 2001), evaluate the ability to solve multiple-step problems in the fewest number of moves. Solving problems in the fewest number of moves is indicative of better planning. Also, the time taken between the final task instructions and the participants' first move is thought to reflect the time spent planning the solution before engaging in the task (e.g., planning time).

Maze tracing tasks such as the *Mazes test* (Stern & White, 2003) and The *Zoo Map Test* (ZMT; Wilson, Alderman, Burgess, Emslie, & Evans, 1996), also have been used to evaluate planning in the clinic and laboratory (Lezak, Howieson, & Loring, 2004). The Mazes test requires completing seven paper-and-pencil mazes of increasing difficulty as quickly as possible, while adhering to task rules such as not going into blind alleys/dead ends. Planning ability is reflected through quantitative and qualitative scoring, including successful maze completion, rule breaks, planning time, etc. The ZMT was designed to improve the ecological/face validity of planning measures (Davalos, Green,

& Rial, 2002; Wilson et al., 1996), and it requires participants to plan a trip to the zoo according to a series of rules. Participants receive a composite raw score that reflects the number of zoo sites successfully visited as well as the number of rules violated in their performance. Planning time, or the time elapsed between instructions and task initiation, also may be calculated.

Although the TOL, Mazes, and ZMT are among the most prominent assessment measures of planning ability, questions remain about the constructs that they assess and their ecological validity. These measures have not been used together in a single study to evaluate planning ability, and it is unclear if and how they are related to each other with regard to the components of planning that they capture (i.e. plan formulation, plan execution, or both). It also is unclear if these traditional measures are related to planning in everyday tasks, although they clearly have limited face validity (c.f., ZMT). Further, some investigators have reported variability in performance even among healthy controls on tests like the ZMT, with some healthy participants scoring as impaired as patients with frontal lobe lesions, suggesting these tests may be overly complex and unrelated to real world functioning (Chevignard et al., 2000; Levine et al., 1998). These drawbacks underscore the need for planning studies that use ecologically valid everyday tasks.

Planning Abilities Assessed through Everyday Tasks

The limitations of conventional planning measures motivated my pre-dissertation research on everyday action planning (Seter et al., 2011). In this prior study, planning in everyday action was evaluated using the Naturalistic Action Test (NAT), a standardized, validated, and reliable *performance-based* measure of everyday actions (Schwartz, Segal, Veramonti, Ferrara, & Buxbaum, 2002). The goals of my prior study were to examine the extent to which planning was observed among healthy control participants and to assess whether or not these planning behaviors were also observed among people with schizophrenia, who are known to exhibit deficits in executive functions (Barch et al., 2001; Barch & Smith, 2008; Elvevåg & Goldberg, 2000; Velligan & Bow-Thomas, 1999) and everyday functioning (Evans et al., 2003; Harvey et al., 2004; Kessler et al., 2007; Klapow et al., 1997; Patterson et al., 1998; Patterson et al., 2001; Semkovska et al., 2004; Semkovska et al., 2002; Sevy & Davidson, 1995; Velligan et al., 2007). It was also of interest to evaluate whether planning was related to the ability to perform everyday tasks accurately and efficiently. First, results showed that planning behaviors during the execution of everyday tasks could be reliably coded. Second, controls demonstrated more planning behaviors than participants with schizophrenia; however, this difference was accounted for by education. People with schizophrenia spent significantly less time planning before initiating the task when planning time was analyzed as a proportion of the total time on tasks. Finally, planning variables were related to everyday task performance variables, such that increased planning was associated with greater task accomplishment and fewer errors.

In the prior study (Seter et al., 2011), given the cognitive deficits exhibited by individuals with schizophrenia, the NAT was specifically chosen to evaluate everyday functioning and planning ability because it is a relatively simple, ecologically valid measure of everyday tasks. The NAT allowed for the direct evaluation of planning abilities in relatively simple everyday tasks, such as coffee making and meal preparation. The differential planning performance of controls and participants with schizophrenia provide further motivation to evaluate planning behavior as it relates to everyday action performance. Questions still remain as to the role of planning in everyday task execution, which the current study aimed to examine in-depth.

Although it is intuitive that planning is crucial for the successful performance of everyday tasks, there remains debate over the extent to which plans are assembled prior to execution of everyday tasks, especially basic or routine tasks. Some investigators de-emphasize planning of these tasks and posit that everyday task "plans" are formulated online to allow for flexibility in achieving practical goals (Joe, Ferraro, & Schwartz, 2002). If behavior is overly planned or rigid, it may not be adaptable to variability in contexts and available materials/objects. Flexibility, rather than a rigid, plan-driven approach to everyday tasks may be more advantageous. Other researchers offer support for plan-driven everyday task behavior (Bickerton, Humphreys, & Riddoch, 2007), and anecdotal accounts suggest that deliberate planning of everyday tasks may be quite helpful for individuals with cognitive deficits.

The Current Study

This study addressed several unresolved issues in the planning literature, including whether and how planning influences everyday action performance, whether traditional neuropsychological measures of planning evaluate a unitary or multidimensional construct that overlaps with planning in everyday action, and whether current neurocognitive models of planning are useful in conceptualizing everyday action planning. Healthy individuals were asked to complete a complex and novel multitasking measure, the 2 x 3 Multi-Level Action Test (2 x 3 MLAT; Buxbaum, Schwartz, & Carew, 1997; Schwartz et al., 1998). The 2 x 3 MLAT is an everyday action task that is designed to tap planning ability, and, as it is more complex than the NAT, it is making more appropriate for research on healthy or mildly impaired populations.

Aim One

The current study has three aims. The primary purpose of this study (Aim One) was to evaluate the extent to which healthy individuals plan before executing everyday tasks, as well as the quality and impact of everyday plans on everyday task performance. This was done by administering participants the 2 x 3 MLAT (Buxbaum et al, 1997; Schwartz et al., 1998), and recording planning behaviors before and during the task. Variables reflecting performance accuracy also were collected. I hypothesized that participants would benefit from planning before engaging in everyday action tasks. Specifically, I hypothesized that individuals who took more time to formulate a plan and formulated a more detailed plan prior to initiating the task would exhibit 1) more planning behaviors (e.g., gathering materials before executing subtasks) during subsequent task execution and 2) better task performance (e.g., fewer errors, break fewer rules, and shorter completion times).

Aim Two

A second aim of this study was to evaluate and compare performance on traditional neuropsychological planning measures (e.g., Tower Test) to each other and to a measure of planning in an everyday action task. The extent to which neuropsychological planning tests are related to each other and to planning in everyday tasks is unknown. I hypothesized that traditional neuropsychological planning measures would be related to each other but reflect two components of planning—formulation (plan time) and execution (planning behaviors during the task). I also hypothesized that traditional neuropsychological planning measures and everyday action planning measures would be positively correlated.

Aim Three

A final aim of the current study was to evaluate the cognitive components of planning as proposed by two different theoretical models of planning: Contention Scheduling/Supervisory Attentional System (Norman & Shallice, 1980, 1986; Shallice & Burgess, 1991; executive functions) and the Constructive Episodic Simulation Hypothesis (Schacter & Addis, 2007; Schacter et al., 2007; episodic memory). These models of planning highlight different cognitive skills involved in planning (i.e., executive functions vs. episodic memory) and conceptualize planning accordingly. I hypothesized that the results would show support for both models, as they are not mutually exclusive. Specifically, I predicted that plan time (i.e., formulation) would be related to measures of episodic memory and plan behaviors during task performance (i.e., execution) would be related to measures of executive functions.

In summary, the current study sought to elucidate the extent to which planning takes place in healthy individuals and whether and how it benefits everyday action performance. It was anticipated that the results would provide important information regarding the ecological validity of neuropsychological tests of planning and inform neurocognitive models of planning processes, particularly in the context of complex, real world tasks. Taken together, this information could inform evaluation and treatment of everyday action difficulties in a wide range of clinical populations.

CHAPTER 2

METHOD

Participants

I originally proposed to recruit 70 participants to achieve a final sample of at least 68 participants, which would provide sufficient power (.80) to detect a medium effect size ($f^2 = .15$) with alpha set at .05 and 2 predictor variables (Aim One Analysis; Erdfelder, Faul, & Buchner, 1996). However, recruitment efforts were more successful than anticipated, and 92 individuals participated in the study. Participants were recruited from the Temple University undergraduate student community, a racially, culturally, and economically diverse population. Participants were recruited if they were age 18 or older and if English was their primary language. Exclusion criteria included: (1) no alcohol or illicit substance abuse (in the past month); (2) no history of traumatic brain injury or neurological disorder, such as cerebral vascular accident or epilepsy; (3) no current psychiatric disorders and no history of psychiatric disorders that required hospitalization; (4) no mental retardation. Non-Native English speakers were excluded because several of the study measures were verbal in nature and required participants to be fluent in English. Information pertinent to exclusion criteria was obtained from in-person interview/questionnaire prior to initiating study participation.

Recruitment

All potential participants were recruited from the Temple University psychology research system. The research study was described to individuals, and informed consent was obtained after they agreed to participate. In addition to signing an IRB-approved consent form, participants signed a separate IRB-approved form that specified that they consented to having their everyday action task performances videotaped. Participants were compensated two research credits toward their psychology course requirements.

Measures

Demographic information and exclusion criteria were confirmed through a questionnaire. Additionally, measures of everyday action, planning, other executive functions, episodic memory, and working memory were administered. These measures are described below.

Demographic/Exclusion Criteria Questionnaire

In order to collect and verify demographic and exclusion criteria information, all potential participants were administered a demographic questionnaire. Questions on this confidential measure included the following items:1. Age; 2. Gender; 3. Years of education completed; 4. Languages spoken/fluency in English; 5. Alcohol/illicit substance abuse in the past month; 6. History of traumatic brain injury or other neurological disorder or diagnosis of mental retardation; and 7. Current or history of psychiatric disorders that required hospitalization.

The 2x3 Multi-Level Action Test (of everyday action)

The 2 x 3 Multi-Level Action Test (2 x 3 MLAT; Buxbaum et al., 1997; Schwartz et al., 1998) was used to assess everyday action abilities. The 2 x 3 MLAT is a structured and challenging test of naturalistic action that is inspired by multi-step tasks like the Six Elements Test (SET; see Shallice & Burgess, 1991). The 2 x 3 MLAT has been used to study everyday action difficulties in healthy participants compared to patient groups, including people with semantic dementia, and dementia with executive deficits (Buxbaum et al., 1997), and closed head injury (CHI; Schwartz et al., 1998).

All materials needed to complete the six total tasks for the 2 x 3 MLAT were presented in a standardized array on a U-shaped tabletop. The 2 x 3 MLAT required participants to execute three everyday tasks twice (e.g., pack lunchbox, wrap present, make toast), while adhering to two rules: 1. Two versions of the same task cannot be performed concurrently or consecutively, and 2. A buzzer, located in a drawer on the tabletop must be pressed after completing each task (i.e., the buzzer must be pressed six total times). Participants were also told that the task materials were in limited supply. These rules are designed to increase the demands of the everyday action tasks and to encourage participants to plan when using the materials for the tasks. A card featuring task rules was available for participants to consult during the entirety of the task, to reduce episodic memory demands and ensure that task errors could not be attributed to failure to recall the task instructions. This card was modeled after the rule card for the modified-SET (MSET; Shallice & Burgess, 1991; Wilson et al., 1996; see Appendix).

Administration of the 2 x 3 MLAT was slightly modified for this study in order to capture the extent to which participants naturally choose to plan before engaging in the task. Participants were provided with task instructions while facing away from the test table, which was covered with a sheet, to minimize planning during instructions. Participants were asked to overtly reiterate the instructions to ensure comprehension of all of the task goals; clarification of instructions was provided as necessary. Upon completion of task instructions, participants were told to sit at the test table and the sheet was removed. At that time, participants were asked if they would like to plan or begin the task immediately (e.g., “Would you like to think through what you will do before you start or do you want to begin now?”). Those who opted to plan were further instructed to think aloud as they planned their performance (e.g., “Tell me how you intend on going about the test?” (Burgess et al., 2000). Last, all participants, regardless of whether or not they chose to plan, were instructed to complete the 2 x 3 MLAT as quickly as possible[2], and timing for the task commenced when participants touched the first item on the table.[3]

Performance was videotaped and evaluated for 1) planning prior to task execution (e.g., plan quality and planning time) and 2) planning behavior and task accuracy during task execution according to the guidelines described below. Maximum time for 2 x 3 MLAT administration was approximately 15-20 minutes.

Planning Prior to Task Execution

Plan quality. This variable evaluated the detail and accuracy of generated plans (Burgess et al., 2000; van Beilen, van Zomeren, van den Bosch, Withaar, & Bouma, 2005) utilizing a coding scheme that is particular to the task requirements of the 2 x 3 MLAT. Plans were evaluated for *detail* based on the number of *tasks* (e.g., making toast, wrapping a gift, and packing a lunchbox), *steps*, and test *rules* provided in the plan. Scores for plan detail ranged from 0 to 57 points, with higher scores indicating more detailed plans. Please see Appendix for scored task steps. Participants’ plans were also scored for plan accuracy, and each inaccurate step resulted in the deduction of 1 point from the final score. *Total score for plan quality was plan detail - plan accuracy (maximum score = 57)*. Plan detail was the sum of the number of tasks (e.g., make toast) and rules (e.g., press buzzer after

[2] Time pressure has been shown to encourage action errors in healthy control participants (Giovannetti, Schwartz, & Buxbaum, 2007; Mattson & Baars, 1992). The goal of this manipulation was to increase task difficulty and thus variability among healthy participants.

[3] Participants who opted to plan before beginning the task were instructed to work as quickly as possible *after* they stated their entire plan. Thus, all participants were provided with this instruction just before they began the 2 x 3 MLAT.

making toast) mentioned as well as the total number of specific task steps (e.g., put bread in toaster) generated, and was scored according to the scheme described in Table 1.

Table 1: *Coding scheme for plan detail*

Tasks (maximum points)	**Maximum Number of Steps**
Toast (1)	5
Present (1)	7
Lunchbox (1)	9
Total x 2= **6**	Total x 2= **42**
Rules (maximum points)	
Buzzer press (6)	
Not completing two subtasks in a row (1)	
Short supplies (3)	
Total= **9**	
Accuracy of Plan	-1 for each inaccurate item

Planning time. This variable was calculated as the time elapsed between the time a participant began and stopped dictating his/her plan. This variable reflects how planning time is captured in neuropsychological tests of planning (Allain et al., 2005; Shallice, 1982; Unterrainer, Rahm, Leonhart, Ruff, & Halsband, 2003; Unterrainer & Owen, 2006; Wilson et al., 1996). Planning time did not have an upper limit.

Planning and Accuracy During Task Execution

Online planning behavior. Planning behaviors during task execution have been operationalized as behaviors that suggest consideration of future steps (Seter et al., 2011). In other words, this variable encompasses all behaviors that garner evidence of forward thinking (e.g., gathering items before engaging in a task or subtask). These behaviors are not essential for successful task completion; they are conceptualized as evidence for a plan-driven *approach* to the task. The concept behind the planning behavior coding scheme is similar to the evaluation of planning strategy/task approach on published tests of planning where participants are scored for their ability to think ahead and plan their actions before actually performing them e.g., TOL, SET/MSET, ZMT). Given the nature of the 2 x 3 MLAT tasks, the planning behavior codes for this study were specific to the basic everyday tasks that compromise the test (see Table 2). *Total online planning behavior* ranged from 0 to 100 points.

Table 2. *2 x 3 MLAT online planning behavior coding rubric (adapted from Seter et al., 2011)*

Toast (maximum online planning points for each toast prepared= 21 and 15, respectively; 36 total points)	-Gathers materials in workspace before putting anything on toast/bread (2 points for gathering with 1 point added per object; maximum 8 points) or -Scans for objects before beginning the subtask (scanning is coded only for overt head movement to left and/or right side of table; do not rely on eye movement (1 point per side of table scanned; maximum 2 points) -Counts number of slices in bread bag (1 point) -Takes out 2 slices of bread and puts aside 1 to use later (1point, only toast 1) -Begins MLAT by toasting bread (1 point, only toast 1) -Engages in planning behavior for toast or other task while waiting (1 point) -While waiting for toast takes lid off butter container (1 point), opens jelly jar (1 point), butter knife (1 point), puts knife in jelly jar (1 point) -Uses separate knives for butter and jelly (1 point) -Leaves lid off butter container (1 point, only toast 1), off jelly jar (1 point, only toast 1) -Leaves knife on butter container or lid (1 point, only toast 1),

Table 2, continued

	leaves knife in jelly jar or lid (1point, only toast 1)
Gift Wrapping (maximum online planning points for each gift wrapped= 11, 22 total points)	-Gathers materials in workspace before selecting the gift (2 points for gathering with 1 point added per object; maximum 8 points) or -Scans for objects before beginning the subtask (scanning is coded only for overt head movement to left and/or right side of table; do not rely on eye movement (1 point per side of table scanned; maximum 2 points) -Measures wrapping paper before wrapping gift by bringing gift and wrapping paper together in the workspace (1 point) -Pre-cuts scotch tape (i.e., cuts pieces and places hanging along the table, their hand, etc, and then uses the tape; 1 point) -Cuts gift-wrap once (1 point)
Lunchbox (maximum online planning points for each lunchbox prepared= 24 and 18 respectively, 42 total points)	-Gathers materials in workspace before beginning the sandwich (mustard or bologna on bread), drink (before juice bottle opened), or the snack (before cookies collected) (2 points for gathering plus 1 point added per object; maximum 14 points) or -Scans for objects taken before beginning the subtask (scanning is coded only for overt head movement to left and/or right side of table; do not rely on eye movement (1 point per side of table scanned per subtask; maximum 2 points) -Counts number of slices in bread bag (1 point) -Takes out 4 slices of bread and puts 2 aside for later (1 point, only Lunch 1) -Leaves open lunchmeat container (1 point, only Lunch 1) -Leaves open mustard container (1 point, only Lunch 1) -Leaves knife in mustard container or on lid (1 point, only Lunch 1) -Measures foil to wrap sandwich before cutting (1 point) -Checks if cap fits the thermos before adding juice to thermos (1 point) -Leaves open juice bottle (1 point, only Lunch 1) -Prepares container for cookies before cookies are removed from package (1 point) -Tears larger piece of foil in two to use for snack later (1 point, only Lunch 1)

2x3 MLAT task accuracy. The following dependent variables were obtained to evaluate overall performance: Task Errors, Rule-Violation Errors, and Total Task Execution Time.

Task errors. This variable represents the total number of errors committed across all six tasks, and it does not have an upper limit.[4] A range of error types were coded, including instances when a task step was not performed (i.e., omission error), instances when a step was performed inaccurately (i.e., commission error), and instances when an extra, irrelevant step was performed (i.e., action addition error). Please see Table 3 for the scoring taxonomy that was used to classify the 2 x 3 MLAT *task errors*. This measure also captured when participants did not leave enough materials to complete the second item of a task.

Rule-violation errors. This is a measure of the number of times the 2 x 3 MLAT rules were violated (i.e., when the same task-type was performed consecutively, when the buzzer press was omitted or pressed at an inappropriate time). This variable did not have an upper limit.

Total task execution time. This variable reflects the total time taken to complete all six 2 x 3 MLAT tasks. Time to completion was calculated as the time elapsed between the point at which the participant initiated the first task to the point at which the participant performed no further action.

[4] Task errors, especially omission errors, and accomplishment of task steps are highly correlated (Schwartz et al., 1998). As such, task errors capture similar information to steps accomplished; for the purposes of this study, only task errors were utilized as a dependent variable.

Table 3. *2 x 3 MLAT scoring taxonomy for task errors* (adapted from Buxbaum et al., 1997; Schwartz et al.,1998)

Error Category	Examples
Omission	Fail to put gift into box
Sequence	Anticipation-omission: seal thermos before filling; close lunchbox before packed Reversal: apply butter, then toast bread Perseveration: apply butter perseveratively (e.g., >19 seconds)
Object substitution	Wrap gift with tissue paper (instead of wrapping paper)
Action addition	Action not interpretable as step in task; e.g., cut gift box, pack extraneous items into lunchbox
Gesture substitution	Correct object used with incorrect gesture; e.g., place lid on thermos with press rather than screw motion
Grasp- spatial misorientation	Misorientation of the object relative to the hand or to another (reference) object; e.g., grasp wrong end of scissors (misoriented relative to hand)
Spatial mis-estimation	Spatial relationship between two or more objects incorrect; act otherwise well-executed; e.g., cut paper much too small for gift
Tool omission	Spread jelly with finger (instead of knife)
Quality	Inappropriate or inexact quantity (spatial or volume); e.g., fill thermos with juice to point of overflow.

Assessment of Knowledge for 2 x 3 MLAT Tasks

Task knowledge for the everyday action tasks was evaluated by a brief *self-report questionnaire* (e.g., Likert-type) regarding familiarity and frequency of performing each 2 x 3 MLAT task. This measure was developed for this study to examine whether task frequency/familiarity influenced planning behavior (see Appendix).

Traditional Neuropsychological Planning Measures

Delis-Kaplan Executive Function System- Tower Test (D-KEFS Tower Test)

The D-KEFS Tower Test (Delis et al., 2001) is a variant of the Tower of London task (TOL; Shallice, 1982), a well-established and widely used test of planning ability (Pantelis et al., 1997; Unterrainer, et al., 2003, 2004; Unterrainer & Owen, 2006). The

Tower Test assesses complex strategic planning and decision-making skills by requiring participants to move five-disks of different sizes between three equivalent holding pegs to obtain a specific disk arrangement (Nuechterlein et al., 2004). The initial arrangement of the disks for each trial is standardized. Participants are instructed to complete the tower (i.e., match the goal arrangement) in the fewest number of moves possible following two rules: 1. Only one disk may be moved at a time, 2. A larger disk cannot be placed on top of a smaller disk. These rules are presented along with a picture of the goal arrangement for each trial. Tower Test trials increase in difficulty with the number of disks to arrange (i.e., increasing from 2-5 disks over a total of 9 trials); number of minimum moves increases from 1 move to 26 moves. Trials also increase in the time limit for completion (i.e., increasing from a limit of 30 seconds to 240 seconds). Administration is discontinued after three consecutive item failures. Planning is measured by the number of moves taken to complete the Tower Test problem, with lower number of moves indicating better planning. The time taken between the final task instructions and the participants' first move is thought to reflect the time spent planning the solution (e.g., planning time). Administration of the Tower Test required approximately 15 minutes. Below are details of the dependent variables specifically examined in this study:

Planning time. This variable represents the average total planning time across completed Tower Test items.[5] Planning time is defined as the time elapsed (in seconds) between the final task instructions and the participants' first full move of a disk (i.e., when a disk is completely lifted off a peg and placed on

[5] The D-KEFS manual describes this variable as "Mean 1st Move Time" and Tower Test scoring includes normative data for this variable (e.g., raw scores converted to scaled scores, mean= 10).

another peg and let go by the participant). There was no upper limit for this variable.

Achievement score. This variable is scored on a scale of 0 to up to 4 points for each item. Zero represents failing the item (not completing it in the range of specified moves and within the time limit). A score of one was assigned when the participant completed the items in the time allotted. Additional points (up to 4 for some items) are assigned for completing the item in few moves. More points are assigned for completing the item in the fewest number of moves possible. The upper limit for this variable was 30 points.

The Tower Test can be used to assess planning and problem-solving abilities for adults 8-89 years of age. The test has good (i.e., moderate) internal consistency for its nine trials across age groups (*r* ranges from 0.43 to 0.84; for 20-29 year olds *r* is 0.62) and test-retest reliability (average test interval was approximately 25 +/- 12.8 days; *r* = .38- .51, *r* for all ages = 0.44). Support for the construct validity of the Tower Test has been reported by its weak and nonsignificant relations with a measure of episodic memory (i.e., California Verbal Learning Test- II; *r* = 0.00 to 0.22 for Tower Test total achievement score and CVLT-II variables; Delis et al., 2001).

Neuropsychological Assessment Battery (NAB)- Mazes

Maze tracing tasks like NAB Mazes, are used to illicit and evaluate planning ability. NAB Mazes consists of seven paper-and-pencil mazes (i.e., items) of increasing difficulty, which examine planning, impulse control, and psychomotor speed (Stern & White, 2003; White & Stern, 2003). Participants must complete each maze as quickly as possible, while following maze-completion rules (i.e., no crossing solid boundary lines, and making a continuous line without picking up the pen from the page). Participants are allowed to backtrack if they get stuck in a dead end or make a mistake. Points are earned for successful maze completion within an allotted time limit. Bonus points are awarded for quicker maze completion. Qualitative features of performance are also tracked, including errors (i.e. crossing solid maze lines), long response latencies, impulsive or quick starts, and haphazard approaches. The task is discontinued after scores of zero on three consecutive items. Administration of NAB Mazes required approximately 15-20 minutes. The following dependent variables were considered in this study:

Planning time. This variable represents the total planning time across completed maze items. Planning time was defined as the time elapsed (in seconds) between the completion of task instructions and initiating the first mark on the page where marked "Start." There was no upper limit for this variable.

Task errors. The total number of errors committed during each maze, including instances of backtracking. There was no upper limit for this variable.

NAB Mazes is part of a comprehensive and integrated assessment battery with 33 tests spanning six neuropsychological modules (Stern & White, 2003; White & Stern, 2003). The NAB is a standardized measure with strong psychometric properties that is appropriate for use in adults 18 to 97 years of age. NAB Mazes is a subtest included in a protocol of multiple tests of executive functions. NAB Mazes has a correlation of .71 with the executive functions module (correlation with other modules ranges from -.10 to .14; White & Stern, 2003).

Behavioral Assessment of the Dysexecutive System (BADS)- Zoo Map Test

The Zoo Map Test (ZMT; Wilson et al., 1996) is part of the BADS test battery; it was designed to improve the ecological/face validity of traditional planning measures, such as the TOL and Mazes tasks (Davalos et al., 2002; Wilson et al., 1996). The ZMT requires participants to plan a route to visit several sites at a zoo using a map, in accordance with a series of rules (e.g., the order of attractions, the number of times paths on the map may be used, etc.). In Condition 1 of the test, participants are required to independently formulate and implement a plan (plan formulation + plan execution). In Condition 2, participants are required simply to follow a pre-formulated plan for visiting the zoo attractions (plan execution only). See Appendix for example of the Zoo Map Test. The test yields a raw composite score for each condition based on the time taken to plan and execute the plan, the number of zoo sites successfully visited, and the number of rules violated. The raw composite score for both conditions has a maximum of 16 points and no minimum; the raw score is converted to a profile score that ranges from 0 to 4. Planning time, or the time elapsed between instructions and task initiation, is also calculated. Administration for ZMT required 5-10 minutes. The following dependent variables were considered in this study:

Planning time. This variable is the time elapsed between the end of instructions and initiation of the first line drawn on the map. Total planning time across Conditions 1 and 2 was utilized in this study. There was no upper limit for this variable.

Total number of errors. A count of the total number of errors and rule breaks committed in both ZMT conditions, including: using a path more than once, deviating from the path, failing to make a continuous line, and visiting inappropriate places. There was no upper limit for this variable.

The ZMT can be used to assess plan formulation and execution abilities for adults. The test has high inter-rater reliability (*r* ranging from 0.90 to 1.00 for sequence and error scoring for both Conditions 1 and 2). Test-retest reliability after 6-12 months of initial testing is moderate (*r* equal to 0.39). The ZMT also successfully distinguishes between control and patient populations, with and without age correction (mean profile score for controls equal to 2.44 +/- 1.13, and mean profile score for patient group equal to 1.97 +/- 1.41). Poor performance on the ZMT is also highly correlated ($r = -0.46$) with caregiver/collateral ratings of executive problems on a dysexecutive questionnaire in the BADS (Wilson et al., 1996).

Executive Function Measures

Behavioral Assessment of the Dysexecutive System (BADS)- Action Program Test

The Action Program Test (APT; Wilson et al., 1996) is a novel problem-solving task that requires developing a plan of action to solve the problem. The APT is adapted from Klosowska (1976) and is part of the BADS test battery (Wilson et al., 1996). The task for the APT is to remove a cork at the bottom of a tall tube using the available materials on the tabletop while adhering to certain task rules (i.e., the assembly, beaker, and tall tube cannot be lifted, and the beaker lid cannot be touched with fingers). See Appendix for an illustration of the APT.

To solve the problem participants need to subsequently work out that the key to solving the problem is to use the water to make the cork float to the top of the tall tube and then work out how to get the water out of the beaker and into the tube. The overall solution is five-steps. There is no time-limit on this test, but if a participant is unable to initiate any steps after 120 seconds, the participant is prompted—the examiner removes the lid with the wire hook and states and encourages the participant to continue. If after another 120 seconds the participant is unable to progress through the next step from the current step they are on, the participant is again prompted with the examiner completing the next step (e.g., attaching the screw top to the container) and encouraging the participant to continue. Prompting continues if necessary until the cork is retrieved. Participants receive a point for every step completed independently (range 0 – 5 points). Administration of the APT required approximately 5-10 minutes. The following dependent variable was considered in this study:

APT Raw Score. This score reflects the number of steps independently completed when solving the APT problem. The score ranges from a minimum of 0 points to a maximum of 5 points in one-point increments. Participants received one point for each of the following: 1. Removing lid from beaker using wire hook, 2. Attaching screw top to container, 3. Filling container with water, 3. Pouring one container of water in tube containing cork, and 4. Pouring second container of water into tube containing cork.

The APT has been used with adults to assess practical problem solving and shows high inter-rater reliability (*r* equal to 1.00, indicating absolute agreement between two testers). Test-retest reliability after 6-12 months of initial testing also is high (*r* equal to 0.67). The APT also successfully distinguishes between control and patient populations, with and without age correction (mean profile score for controls equal to 3.77 +/- 0.52, and mean profile score for patient group equal to 3.18 +/- 1.15). Poor performance on the

APT is also moderately correlated (r = -0.37) with caregiver or collateral ratings of executive problems on a dysexecutive questionnaire in the BADS (Wilson et al., 1996).

Haylings Sentence Completion Test (Haylings Test)

The Haylings Test (Burgess & Shallice, 1997) was designed to be sensitive to deficits in executive function, and specifically is purported to capture inhibition, initiation speed, and response suppression (Bielak, Mansueti, Strauss, & Dixon, 2006). The test consists of two sections, each comprised of 15 sentences with a missing last word. In Section 1, participants are read each sentence and required to verbally generate a word that correctly completes the sentence as quickly as possible (e.g., “He mailed the letter without a…(stamp)”). In Section 2, participants are required to generate a word that is incorrect and completely unrelated to the sentence (e.g., “The captain wanted to stay with the sinking…(banana)”). Thus, it is purported that in this latter section participants first inhibit the strongly activated response (e.g., ship) before generating a new unrelated word (e.g., banana) (Bielak et al., 2006). Participants are not allowed to use the same word for all responses, and are told this after their first repetition. Participants’ performances are evaluated based on response latency scores for both sections (i.e., the time elapsed between the examiner finishing reading the sentence and the participant responding) and error scoring for responses in Section 2. Administration of the Haylings Test took about 5 minutes. The following dependent variable was considered in this study:

Total A and B errors. This is a measure of the number of responses on Section 2 that break the response rule and plausibly complete sentences or are related to the sentences in some way; it ranged from 0 to 15.

The Haylings Test has been utilized in various populations (e.g., bipolar disorder (Yatham et al., 2010); traumatic brain injury (Hewitt, Evans, & Dritschel, 2006)) and is appropriate to use in adults ages 18 to 80 years (Burgess & Shallice, 1997). Norms were established on 121 control participants and 77 clinical participants and consist of age-related cut-off scores, as well as age and pre-morbid IQ cut-off scores (Burgess & Shallice, 1997). In a study attempting to establish norms for older adults (ages 53 to 90, n = 457; Bielak et al., 2006), participants responded significantly more quickly to sentences in Section 1 compared to 2. Overall, younger participants had shorter response latencies and fewer errors, and the inter-rater reliability among three scorers was 96.0%, based on a random sample of 20 Haylings Tests.

Brixton Spatial Anticipation Test (Brixton Test)

The Brixton Test (Burgess & Shallice, 1997) is an implicit rule-learning and following test (i.e., a test of logical abstraction). Participants are presented with a visual stimulus comprised of an array of ten circles (e.g., two rows of five circles) numbered one to ten; one circle is blue in color on each page. Participants are required to work out where the blue circle will be located on the following page; simple logical rules govern the location of the next blue circle. Participants are presented with one page at a time and point to the location of the next circle before the page is turned based on the pattern or inferred rule; the test consists of 56 stimuli (see Appendix). The rule for the location of the blue circle changes without warning throughout the course of administration, and participants are required to respond accordingly. Scaled scores ranging from 1 to 10 are calculated based on the total number of errors made on the test. Administration of the

Brixton Test required approximately 5-10 minutes. The following dependent variable was considered in this study:

> *Total number of errors.* This variable represents the total number of incorrect responses across 55 trials. On trials where the rule changes, responses are marked as correct if they would be correct had the rule not changed on that trial (Bielak et al., 2006; van den Berg et al., 2009). The maximum number of errors is 55.

The Brixton Test has been utilized in various populations (e.g., traumatic brain injury (Hewitt et al., 2006); stroke, diabetes mellitus, MCI/early dementia, schizophrenia, and Korsakoff's patients (van den Berg et al., 2009)) and is appropriate to use in adults ages 18 to 80 years (Burgess & Shallice, 1997). Normative data from 121 healthy participants and 77 clinical participants have been published along with age-corrected and IQ-corrected cut-off scores (Burgess & Shallice, 1997). Normative data for older adults (e.g., ages 53 to 90, n = 457; Bielak et al., 2006) show more errors are associated with older age, less education, and being female. Similar results were obtained in a subsequent study establishing normative data for healthy older adults (e.g., ages 55 to 92 years, n = 283; van den Berg et al., 2009). Test-retest reliability based on a sample of 83 healthy older adults who completed the Brixton Test between 6 to 48 months apart was lower (r = 0.61 (van den Berg et al., 2009) than previously reported in the original normative sample (r = 0.71; Shallice & Burgess, 1997). Construct validity analysis has also shown that the Brixton Test is more related with a measure of executive functioning (i.e., the Trail Making Test) than with measures of episodic memory or processing speed (van den Berg et al., 2009).

Episodic Memory Measures

Wechsler Memory Scale 4th Edition, Logical Memory I and II (WMS-IV- LM I, II)

The WMS-IV- LM I, II is a learning and memory test for structured verbal information (i.e., short stories). LM I and II are part of the WMS-IV battery (Wechsler, 2009a). The examiner reads participants stories (e.g., story A, B, or C) that they are required to recall verbatim immediately and after a delay (see Table 4); participants are also administered a yes/no- question-recognition trial about details of the stories after the recall trials. Participants receive a cue during LM-II if they have difficulty with initiating story recall. Please see Table 4 for an outline of the LM I and II story administration as well as scoring information.

Table 4. *Maximum raw score values for all trials (in the order they are administered) in the WMS-IV- LM I, II tests (Wechsler, 2009a)*

	Logical Memory- I	LM- I Maximum Raw Score	Logical Memory- II (after a 20-30 minute delay)	LM-II Maximum Raw Score
Ages 16-69 years	Story B	---	Story B Delayed Recall	25
	Immediate Recall Story B	25	Story C Delayed Recall	25
	Story C	---	(Delay recall total score)	50
	Immediate Recall Story C	25	Story B Recognition	15
	(Immediate recall total score)	50	Story C Recognition	15
			(Recognition total score)	30

Each story item is scored as either a 0 or 1, based on scoring-criteria guidelines (Wechsler, 2009a). Administration of the WMS-IV, LM I, II takes 5-10 minutes, followed by a 20-30 minute delay, and then 10 minutes for administration of delay recall and recognition trials. The following dependent variable was considered in this study:

Delayed memory (LM II) raw score. This variable reflects the number of story details correctly recalled following a 20-minute delay. Raw scores ranged from 0 to 50.

The WMS-IV- LM-I, II is a rigorously standardized measure with good psychometric properties (Wechsler, 2009b). The WMS-IV is appropriate for people ages 16 to 90; the normative sample includes a total of 1,400 participants. Both LM-I and LM-II exhibit good internal consistency across all age groups (adult battery: *r* range .77 to .86 for LM-I, .80 to .90 for LM-II, average *r* .82 and .85, respectively; older adult battery: *r* range .83 to .88 for LM-I, .85 to .89 for LM-II, average *r* .86 and .87, respectively). Both tests also exhibit good test-retest reliability (adult battery *r* .74, .71, older adult batter *r* .79, .77; testing interval ranged from 14 to 84 days, mean 23 days). Both LM-I and II also correlate most strongly with the WMS-IV memory indices, including auditory, immediate, and delayed memory (*r* ranges from .63 to .83), and LM-II strongly falls on the auditory memory factor (.57). LM I and LM II also both show correlations with other episodic memory tests (e.g. RBANS memory index, CVLT-II) (Wechsler, 2009b).

The Autobiographical Memory Interview (AMI)

The AMI (Kopelman, Wilson, & Baddeley, 1990) is a semi-structured interview measure of personal retrograde episodic memory. It consists of three sections probing memory for childhood, early adult life, and recent life (e.g., names of three teachers or friends from high school, home address during your first job, month in which a holiday or journey took place in the last year). Participants are read the questions aloud and required to verbally respond, with responses recorded verbatim. Question-items in each of the three sections are scored from .5 to 3 points and are tallied into personal semantic total scores; each section also includes three time-period specific autobiographical incident questions (e.g., recall of an event while in high school; maximum three points each). The AMI scoring manual outlines questions that can be substituted if certain questions/prompts are not applicable to participants (Kopelman et al., 1990). Corroboration with family member or close friends for memory accuracy is not required for scoring this test. The AMI scoring rubric is outlined in Table 5.

Table 5. *AMI scoring guidelines*

Section	Personal Semantic- Maximum Points	Autobiographical Incident- Maximum Points
Childhood		
Period before school	5	3
First school	8	3
Main secondary school	8	3
Early Adult Life		
Career	8	3
Wedding	9	3
Children and meeting someone new	4	3
Recent Life		
Present institution	8	3
Previous institution	8	3
Last Christmas or Thanksgiving	2	---
Holiday or journey	3	3
Total Score Summary		
Section A: Childhood	21	9
Section B: Early adult life	21	9
Section C: Recent life	21	9
Maximum Total	63	27

Administration of the AMI took approximately 15-20 minutes. The following dependent variable was considered in this study:

Percent autobiographical incidents score. This variable is the percentage of the total score for autobiographical incidents items across all sections of the AMI over the total points possible for the questions answered. There are 9 questions worth 3 points each for a maximum of 27 points; higher percentage indicated better performance.

The AMI has good psychometric properties for use in adults. To establish inter-rater reliability, three independent raters scored written descriptions of recalled memories, and correlations between rater-pairs ranged from 0.83 to 0.86 (Kopelman et al., 1990). The AMI distinguishes between healthy and patient (i.e., amnestic patients) participants, and correlates with other measures of remote memory. The test also shows a temporal gradient for retrograde memory that matches other measures' findings that the most distant memories are spared when there is disruption of retrograde memory (e.g., Ribot's law; Kopelman et al., 1990). Finally, there is evidence to support the accuracy of autobiographical memories produced in studies that have shown strong correspondence between responses and reports by participants' relatives (e.g., average 93.1 percent accuracy across all sections).

Working Memory

(Automated) Symmetry Span (ASymSpan)

Symmetry span is a measure of visuospatial working memory capacity. An automated version of symmetry span provides for easy administration of this measure (Barch et al., 2009; Schrock & Engle, 2005). Like other classic complex span tasks (e.g., Operation and Reading Span; Unsworth, Heitz, Schrock, & Engle, 2005), symmetry span requires participants to "process" information while "maintaining" some other type of information in working memory. The ASymSpan task requires participants to recall the sequence of red squares that are presented individually in a 4x4 grid (e.g., the maintenance/storage component) while performing a symmetry judgment task (e.g., the processing component) (see Appendix; Barch et al., 2009; Martens & Johnson, 2009). The symmetry judgment task requires participants to judge whether or not individually presented figures, consisting of 8 x 8 grids with blackened squares, are vertically symmetrical (see Panels 1 and 2 of the task's illustration in the Appendix; Barch et al., 2009; Martens & Johnson, 2009).

The ASymSpan alternates between the symmetry task and the memory task. When the symmetry-task figure appears on the screen, participants have to judge whether it is symmetrical about the vertical axis as quickly as possible. (The figure is symmetrical in approximately half the trials.) The next screen displays a 4 x 4 grid with one red square, and its location is to be remembered. This pattern continues until the participant is prompted to provide the sequence of observed red-squares on the 4 x 4 grid (Martens & Johnson, 2009). To account for individual differences in speed, the symmetry judgment task is practiced first, and the maximum time allowed to solve the symmetry judgment in test trials is set to equal 2.5 SDs above (i.e., slower) each participant's mean time for the practice trial (Barch et al., 2009; Unsworth & Engle, 2007).

To maximize task engagement and effort, participants are also given feedback regarding accuracy and are encouraged to remain above 85% accuracy. Twelve sets of trials are completed, and each set contains two to five symmetry-memory combinations (Martens & Johnson, 2009). Scores are immediately available upon completion of the task. Administration of the ASymSpan takes 20 minutes. The dependent variables of interest for this study include:

ASymSpan absolute score. This variable is the traditional scoring method for the symmetry span test, and it denotes the sum of all perfectly recalled trials for the sequence of red squares (e.g., location and number of squares in the sequence). Scores ranged from 0 to 42.

The ASymSpan test has good psychometric properties. Large-sample latent-variable studies examining the validity of complex span tasks have shown that complex span tests, like the ASymSpan test: 1. exhibit construct validity by strongly correlating with each other (r= .54 for automated symmetry and operation spans in nearly 4,000 college-aged participants and r=.52 in approximately 750 participants aged 18-98), 2. display criterion-validity by accounting for variance in fluid intelligence and reading comprehension, 3. meet Cronbach alpha cut-offs for internal consistency (.80 for symmetry span), 4. meet cut-offs for satisfactory reliability with test-retest correlations (r= .77), and 5. show a developmental trajectory, such that scores increase throughout childhood and decrease in older relative to younger adulthood (Barch et al., 2009; Kane et al., 2004; Unsworth, Redick, Heitz, Broadway, & Engle, 2009). This measure, like other complex span measures has been extensively studied in healthy young adults (Conway et al., 2005).

Procedure

All participants were tested at Temple University, Main Campus. After participants signed the consent forms, they were administered the 2x3 MLAT followed by neuropsychological and cognitive measures. All measures were administered to participants individually, over a single 2-hour session. The order of the neuropsychological and cognitive measures was quasi counterbalanced across subjects to minimize order effects. Participants completed the measures in one of five predetermined orders. See Appendix for a schedule of each session. The 2x3 MLAT performance was videotaped for subsequent analysis by the investigators.

Statistical Analysis

This study used a prospective observational design. All statistical analyses were conducted with Statistical Package for Social Sciences (SPSS) software, Version 19. Significance levels for all statistical analyses were set at .05 and were two-tailed. More information about the statistical analyses is provided below before the discussion of each aim.

CHAPTER 3

RESULTS

Demographics

The sample consisted of 92 individuals, 44 (47.8%) of whom were male. Participants had a mean age of 21.80 years (SD = 4.74) and had completed an average of 13.46 years of education (SD = 1.04). More than half of the sample 59.8% (*n* = 55) identified as Caucasian, 14.1% identified as African American (*n* = 13), 6.5% (*n* = 6) identified as Latino/Hispanic, 16.3% (n = 15) identified as Asian, and 3.3% (n = 3) identified as Multiracial. Participants were distributed across the five administration versions, such that 18 received administration version 1, 17 received version 5, and 19 each received versions 2 through 4.

Interrater reliability for 2x3 MLAT planning and execution variables

Inter-rater reliability was evaluated in a randomly selected subsample of 13 participants. After inter-rater reliability was calculated, final scores for this subsample were determined by reconciling discrepancies. This process included discussion between the coders and review of videotapes when necessary. Following the discussion, the coders agreed on a final code/score.

2x3 MLAT Plan

2x3 MLAT Plan Quality Score was very strongly and significantly correlated between the two raters ($r = 0.997$, $p < .01$). With respect to scoring Plan Quality, the raters demonstrated 100% agreement in coding tasks (Plan Task Points Cohen's $\kappa = 1.0$, perfect agreement), 99.63% agreement in coding steps (Plan Step Points Cohen's $\kappa = .9315$, almost perfect agreement), and 98.29% agreement in coding rules (Plan Rules

Points Cohen's κ = .8481, almost perfect agreement) (Landis & Koch, 1977). Mean Total 2x3 MLAT Planning Times between the two raters differed by approximately 1 second (Rater 1 *M* = 24.38, *SD* = 37.83; Rater 2 *M* = 23.85, *SD* = 37.32) and were very strongly and significantly correlated ($r = .99$, $p < .01$).

2x3 MLAT Execution

The raters demonstrated 96.89% agreement in coding online planning behaviors on the 2x3 MLAT (2x3 MLAT Online Planning Behaviors Cohen's κ = .9317, almost perfect agreement; Landis & Koch, 1977). Mean Total 2x3 MLAT Completion Times between the two raters differed by less than 1 second (Rater 1 M = 827.62, *SD* = 180.75; Rater 2 *M* = 827.54, *SD* = 827.54, 181.07) and were very strongly and significantly correlated ($r = 1.00$, $p < .01$). Further, the mean 2x3 MLAT Rule Violations between the two raters did not differ (Rater 1 *M* = 2.23, SD = 3.24; Rater 2 *M* = 2.23, SD = 3.24), nor did the mean differ with respect to CES (Rater 1 *M* = 6.77, *SD* = 5.43; Rater 2 *M* = 6.77, *SD* = 5.26). 2x3 MLAT Rule Violations coded by the two raters were very strongly and significantly related ($r = 1.0$, $p < .01$), as was CES errors ($r = .94$, $p < .01$).

Hypothesis Testing

Aim 1: Examine whether and how planning facilitates or impedes everyday action performance

Preliminary Analyses and Analysis Plan

Normality of planning and performance variables data was assessed and tested with the examination of histograms and boxplots, skewness and kurtosis, and the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. 2x3 MLAT Total Time, 2x3 MLAT Online Planning Behaviors, and Task Frequency Rating were normally distributed. 2x3 MLAT Total Errors was normally distributed when transformed using square-root transformation (skew statistic = .50). 2x3 MLAT Total Planning Time for just the participants who planned (i.e., Planners) was normally distributed when transformed using square-root transformation (skew statistic = .79). 2x3 MLAT Plan Quality Score for the Planners was trimmed to make it normally distributed (skew statistic = .81). Task Familiarity Rating was not normally distributed and unable to be transformed; non-parametric analyses were used to evaluate this variable.

Of all 92 participants, 48.9% (n=45) chose to plan before engaging in the 2x3 MLAT. Given this distribution of Planners and Non-planners, two groups were formed and between group analyses were conducted for variables assessing everyday action performance. Between group analyses between Planners and Non-planners were conducted using independent samples t-tests or Mann-Whitney U tests. Cohen's d calculations were used to estimate effect sizes of between group analyses (0.2 = small, 0.5 = medium, 0.8 = large; Cohen, 1988).

To examine relations between planning variables and action performance variables, correlations were conducted using the subsample of Planners (n = 45). Pearson Correlations (r) were performed to analyze relations between 2x3 MLAT Total Planning Time and 2x3 MLAT Plan Quality Score variables, and all 2x3 MLAT dependent variables and self-report measures.

All non-parametric analyses were also run using parametric tests, and the results never differed between statistical methods. Also, all analyses were run using both transformed and non-transformed variables, and the results did not differ. To improve interpretability of the findings, means and standard deviations of everyday action performance and planning variables presented in the text and tables represent the non-transformed data; statistical results are provided using the transformed variables.

Between Group Analyses

As shown in Table 6, the groups were not significantly different with respect to demographic variables and test administration version.

Table 6. *Participants' demographic characteristics*

	Non-planners (n = 47)		Planners (n = 45)		Analysis	
	M	*SD*	*M*	*SD*	χ^2	*t-value*
Age	21.23	4.08	22.40	5.33		1.18
Education Completed	13.43	1.02	13.49	1.08		.290
Sex (%women)	57%		47%		1.07	
Ethnicity					4.37	
Caucasian	66%		53%			
African American	9%		20%			
Hispanic	4%		9%			
Asian	19%		13%			
Multiracial	2%		4%			
Administration Version					3.34	
1	17%		22%			
2	23%		18%			
3	26%		16%			
4	15%		26%			
5	19%		18%			

Everyday Action Performance Between Non-planners and Planners

There was no significant difference between Non-planners and Planners on the total number of errors made during the everyday action task (i.e., CES and 2x3 MLAT Rule Violations) ($t(89) = -1.24$, $p = .22$). There was also no significant difference in the 2x3 MLAT Total Time required to complete the task between groups ($t(88) = -.43$, $p = .67$). Finally, there was no significant difference between the groups in total 2x3 MLAT Online Planning Behavior points ($t(88) = -.86$, $p = .39$). See Table 7 for a summary of these results. Note, due to tape malfunction, two participants in the Non-planners group had incomplete 2x3 MLAT data.

Participant's Frequency and Familiarity Ratings also were evaluated. Planners reported significantly lower Frequency Ratings than Non-planners, indicating that Planners had performed the 2x3 MLAT tasks less often than Non-planners ($t(90) = 2.03$, $p = .046$). See Table 7 for a summary of results. The groups did not differ on Familiarity Ratings; both groups rated the tasks to be highly familiar.

Table 7. *2x3 MLAT planning, performance, and task ratings across both groups*

	Non-planners				Planners (n= 45)			Analysis		
	n	*M*	*SD*	*range*	*M*	*SD*	*range*	*t-value*	*p*	*Cohen's d*
2x3 MLAT Total Errors (points)	46	9.87	5.93	0-29	11.09	5.76	3-26	-1.24	.22	.21
2x3 MLAT Total Time (seconds)	45	783.33	198.09	398-1310	801.89	209.28	427-1416	-.43	.67	.09
2x3 MLAT Online Planning (points)	45	29.00	7.86	8-43	30.51	8.85	9-51	-.86	.39	.18
Task Frequency Rating	47	5.79	1.53	3-9	5.16	1.46	2-9	2.03	.046*	.42
Task Familiarity Rating	47	9.66	2.12	4-12	8.89	2.47	4-12		.14^	.34

^Mann Whitney U Test was performed because data were not normal and unable to be transformed.

Relations between Planning Variables and Everyday Action Performance within the Group of Planners

Relations between task planning and everyday action performance were examined within the group of participants who opted to plan before beginning the 2x3 MLAT (Planners; $n = 45$), to evaluate whether more planning behaviors conferred a performance advantage. The mean 2x3 MLAT Planning Time among the Planners was 44.11 (28.98) seconds and mean 2x3 MLAT Plan Quality Score was 5.44 (4.01) points. Planning Time and Plan Quality Score were positively and significantly related ($r = .60$, $p < .01$). Correlation analyses between Planning Time and Plan Quality Score and all 2x3 MLAT dependent variables, including, 2x3 MLAT Total Errors, 2x3 MLAT Total Time, and 2x3 MLAT Online Planning Points were non-significant (Table 8). Planning Time and Plan Quality Score were also not significantly related to participants' report of familiarity (e.g., Task Familiarity Rating) and frequency (e.g., Task Frequency Rating) of engaging in 2x3 MLAT tasks (Table 8).

Table 8. *Correlations (r- values) between 2x3 MLAT Planning Time and Plan Quality Score for the Planners group (n = 45)*

	Planning Time	Plan Quality Score
2x3 MLAT Total Errors (points)	-.06 (p= .69)	.03 (p= .87)
2x3 MLAT Total Time (seconds)	-.12 (p= .43)	-.14 (p= .37)
2x3 MLAT Online Planning (points)	-.01 (p= .98)	.08 (p= .60)
Task Frequency Rating	-.05 (p= .75)	-.08 (p= .59)
Task Familiarity Rating	.01 (p= .96)	-.05 (p= .75)

Aim 1 Summary

Planning in everyday tasks was examined to determine whether planning facilitated or impeded performance. Consistent with our hypotheses, plan formulation was reliably coded using a novel scoring system developed for this study. Planning time and planning behaviors during everyday task performance were also evaluated and reliably coded using a scoring system amended from a prior study (Seter et al., 2011). Approximately half the sample chose to plan before initiating the everyday tasks. However, there was no significant difference between Planners and Non-Planners in everyday action performance as measured by total errors, and completion time. Planners and Non-Planners also did not differ in online planning behaviors. Interestingly, Planners reported engaging in the everyday tasks used in the MLAT less frequently than Non-planners, but both groups reported the tasks to be equally familiar.

Among the Planners, pre-planning variables including planning time and the quality of plans were fairly variable, with some participants spending very little time generating a terse outline of their intentions whereas others spent up to two minutes detailing the specific actions that they were going to perform. Contrary to prediction, however, even among the Planners there were no significant relations between pre-planning variables and everyday action performance (i.e., errors, completion time) or online planning behaviors. There were also no significant relations between pre-planning variables and participants' ratings of familiarity or frequency of engaging in everyday tasks.

This is the first study to our knowledge to directly assess both plan formulation and planning behaviors during the performance of everyday tasks. Because this approach is novel, it was important to establish that plan quality and time variables could be extracted reliably from participants' dictated plans. Further, planning variables were again extracted reliably from behaviors during task execution (Seter et al., 2011). This novel approach to evaluating planning for both formulation and execution stages directly during everyday tasks, holds promise for improving the ecological and face validity of planning assessments.

Contrary to prediction and prior studies (Phillips, Wynn, McPherson, & Gilhooly, 2001; Seter et al., 2011; Unterrainer et al., 2004), participants in our sample did not benefit from planning prior to or during everyday action tasks. Interestingly, Non-planners reported greater frequency with engaging in the 2x3 MLAT everyday action tasks, indicating the possibility that individuals who opted to take time to plan the task did so to compensate for their inexperience with the tasks. Together, the findings suggest that task planning is not unequivocally beneficial to everyday action performance, at least for cognitively healthy participants.

Aim 2: To evaluate and compare outcomes on traditional neuropsychological planning measures to each other and to planning in everyday action tasks.

Preliminary Analyses and Analysis Plan

Normality of neuropsychological planning test data were assessed and tested with the examination of histograms and boxplots, skewness and kurtosis, and the Kolmogorov-Smirnov and Shapiro-Wilk tests of normality. Following these analyses, extreme outliers were removed from NAB Mazes Planning Time and Tower Test Planning Time, resulting in the loss of nine total participants for subsequent analyses. Data for four more participants were not included in the principal component analysis (PCA) because of missing data from the neuropsychological measures due to administration error (1 participant's NAB Mazes, and 1 participant's NAB Mazes and Zoo Map, and 2 participants' Tower Test). Tower Test Planning Time, Tower Test

Achievement Score, and 2x3 MLAT Online Planning Behaviors were normally distributed. Distributions for NAB Mazes Planning Time, NAB Mazes Total Errors, Zoo Map Total Planning Time, and Zoo Map Total Errors improved following square-root transformation (skew statistic before transformation = 6.32 to 10.98; following transformation = 3.10 to 5.26).

To examine whether six planning variables obtained from three different traditional neuropsychological tests of planning could be reduced to fewer meaningful component scores, a principal component analysis (PCA) was conducted with two variables from each of the three neuropsychological planning tests (e.g., 1- NAB Mazes Total Planning time, 2- NAB Mazes Total Errors, 3 -Zoo Map Total Planning Time, 4 -Zoo Map Total Errors, 5 -Tower Test Total Planning Time, and 6 -Tower Test Achievement Score). Extracted components were saved as factor scores using the regression method. Notably, the PCA was also run using non-transformed variables, and including the extreme outliers; the pattern of the results was not different. All of the results reported here include transformed variables without the extreme outliers.

To examine the relations between the extracted component planning scores and measures of everyday action planning (i.e., 2x3 MLAT Planning Time and 2x3 MLAT Online Planning Behaviors), Pearson correlation analyses were performed. Because half the sample engaged in planning prior to starting the 2x3 MLAT, the 2x3 MLAT Planning Time could not be normalized for the entire sample. The dichotomous 2x3 MLAT Planning Time variable (i.e. Planning Group), reflecting whether participants planned or did not plan, was analyzed using point-biserial correlation analyses.

Principal Component Analysis of Neuropsychological Planning Variables

Seventy-nine participants were included in the principal component analysis (PCA) of the 6 variables from the neuropsychological planning tests. Orthogonal rotation (varimax) was applied. The Kaiser-Meyer-Olkin measure verified the sampling adequacy for the analysis, KMO= .53 (Field, 2005), and all KMO values for individual items were > .50, which is at the acceptable minimum of .5 (Field, 2005). Bartlett's test of sphericity chi-square (15) = 38.681, *p* =.001, indicated that correlations between items were sufficiently large for PCA (Field, 2005). Two components had eigenvalues over Kaiser's criterion of 1 (Field, 2005) and in combination explained 51.11% of the variance. Table 9 and Figure 1 show the factor loadings after rotation.

Table 9. *Rotated component loadings for the six neuropsychological planning variables*

Variables	Component 1	Component 2
Zoo Map Total Errors	-.063	**.779**
Tower Test Achievement Score	-.144	**-.692**
Tower Test Total Planning Time	**.616**	.122
NAB Mazes Total Planning Time	**.748**	.266
NAB Mazes Total Errors	**-.601**	**.430**
Zoo Map Total Planning Time	**.522**	-.333

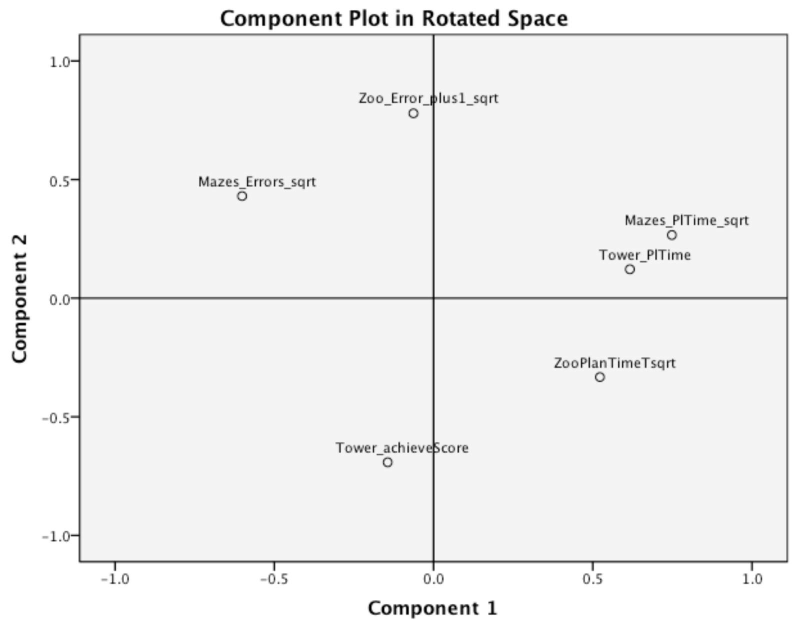

Figure 1. Rotated component loadings plot for the six neuropsychological planning variables (Zoo_Error_plus1_sqrt = Zoo Map Total Errors; Mazes_Errors_sqrt = NAB Mazes Total Errors; Mazes_PlTime_sqrt = NAB Mazes Total Planning Time; Tower_PlTime = Tower Test Total Planning Time; ZooPlanTimeTsqrt = Zoo Map Total Planning Time; Tower_achieveScore = Tower Test Achievement Score).

Upon inspection of the PCA results, two planning variables/items (e.g., NAB Mazes Total Errors and Zoo Map Planning Time) loaded on both components. A second PCA without these two variables was run, and the results were similar but the component scores were easier to interpret (n = 80; KMO = .52 and all KMO values for individual items >.50; Bartlett's test of sphericity chi-square (6) = 13.58, p = .035; two components with eigenvalues > 1 explaining 64.03% of the variance). Table 10 shows the component loadings after rotation. Component 1 represents Planning Accuracy and component 2 represents Planning Time (see Figure 2). Factor scores for these two components were saved to use in subsequent analyses.

Table 10. *Rotated component loadings for four neuropsychological planning variables*

Variable	Component 1	Component 2
Zoo Map Total Errors	**.804**	-.032
Tower Test Achievement Score	**-.756**	-.114
Tower Test Total Planning Time	-.132	**.831**
NAB Mazes Total Planning Time	.241	**.750**

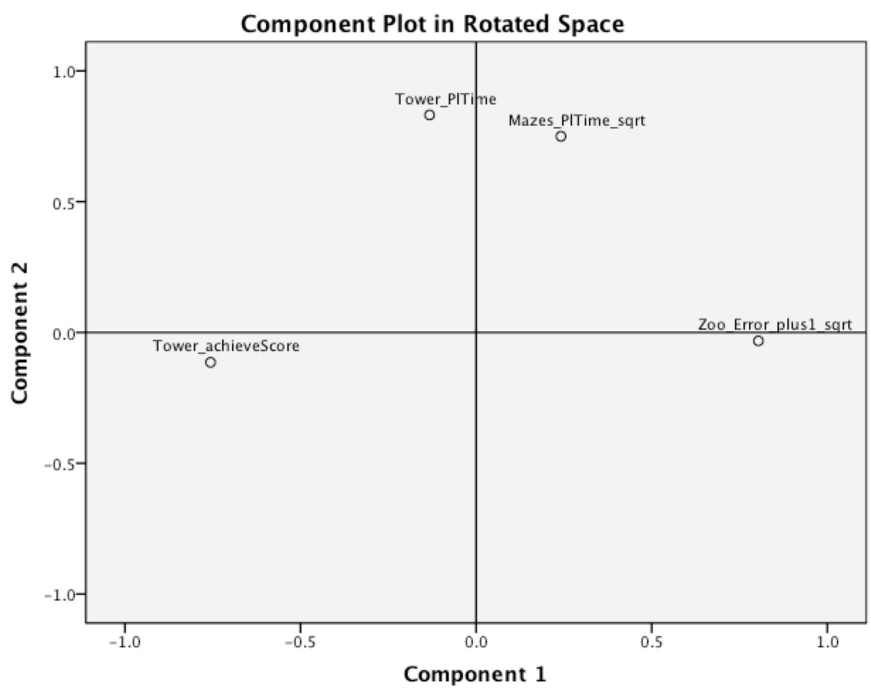

Figure 2. Rotated component loadings plot for four neuropsychological planning variables (Tower_PlTime = Tower Test Total Planning Time; Mazes_PlTime_sqrt = NAB Mazes Total Planning Time; Zoo_Error_plus1_sqrt = Zoo Map Total Errors; Tower_achieveScore = Tower Test Achievement Score).

In conclusion, the six original neuropsychological variables were meaningfully reduced to two component scores that were comprised of four variables. Two of the original six variables did not clearly load on either of these two components and will be evaluated separately in all subsequent analyses (see Table 11).

Table 11. *Extracted components and corresponding neuropsychological planning test variables*

Component 1- Planning Accuracy	Zoo Map Total Errors Tower Test Achievement Score
Component 2- Planning Time	Tower Test Total Planning Time NAB Mazes Total Planning Time
Independent Neuropsychological Planning Variables	NAB Mazes Total Errors
	Zoo Map Total Planning Time

Relations between Neuropsychological Measures of Planning and Everyday Action Planning Variables

The 2x3 MLAT Planning Time variable and the 2x3 MLAT Online Planning Behaviors variable were not significantly correlated (r(91) = .091, p = .39), suggesting that planning before the start of the 2x3 MLAT did not influence whether or not individuals took a planful approach to performing the task. Therefore, these variables were analyzed separately in relation to the neuropsychological variables. As shown in Table 12, the correlation between 2x3 MLAT Planning Time and Component 1 (Planning Accuracy on neuropsychological tests of planning) was statistically significant. However, the correlation was in the opposite direction than expected, with Planners (on the 2x3 MLAT) showing lower scores on neuropsychological tests of Planning Accuracy.[6] All other correlation coefficients revealed weak and nonsignificant relations between 2x3

[6] The variables comprising Component 1, Zoo Map Total Errors and Tower Test Achievement Score were examined in relation to Component 1 to determine the variable's variance. They were positively and negatively related, respectively, indicating that a greater Component 1 score demonstrated worse planning accuracy. Thus, the positive relation between Component 1 and 2x3 MLAT Planning Time indicates that as Component 1, planning accuracy, gets worse, MLAT planning time increases.

MLAT Planning Time and Online Planning Behaviors and neuropsychological planning tests (see Table 12).

Table 12. *Correlation Analyses showing relations between 2x3 MLAT Planning variables and neuropsychological planning test variables*

	N	2x3 MLAT Planning Time/Pre-Plan Yes-No r (*p*)	N	2x3 MLAT Online Planning Behaviors r (*p*)
Component 1 (Planning Accuracy)	80	.25 (.02)*	78	.19 (.09)
Component 2 (Planning Time)	80	.01 (.96)	78	-.12 (.31)
NAB Mazes Total Errors	82	.053 (.64)	80	.17 (.14)
Zoo Map Total Planning Time	81	.18 (.12)	79	-.032 (.78)

Aim 2 Summary

Traditional neuropsychological tests of planning were compared to one another and to planning in everyday action tasks to determine whether these variables reflected a single or multidimensional construct. PCA analyses on six planning variables from the three neuropsychological tests indicated two components, Planning Accuracy and Planning Time, as well as two independent variables of planning. Amongst Planners on the 2x3 MLAT, the relation between the Planning Accuracy component variable and planning time on everyday action tasks was significant but negative (i.e., in the opposite direction than expected). No other relations between planning variables from traditional neuropsychological tests of planning and planning in everyday action were significant.

To the best of our knowledge, no other study has included multiple neuropsychological tests of planning in a single study to determine whether they examine similar or different constructs. These results suggest some overlap among planning variables derived from traditional neuropsychological tests of planning, and this overlap was consistent with the hypothesis that the measures would reflect two components of planning, time (i.e., formulation) and execution/accuracy (i.e., planning behaviors during the task; execution). However, neuropsychological measures and everyday action did not associate as predicted. Taken together, these results support a multidimensional view of planning where plan formulation and planning behaviors (i.e., execution) should be evaluated separately. The results also suggest that neuropsychological measures of planning may not be useful in understanding planning in everyday tasks.

Aim 3: To explore relations between measures of executive functions (e.g., SAS), episodic memory, and working memory and measures of planning

Preliminary Analyses and Analysis Plan

Normality of data from tests of executive functions (Action Program Test; Haylings Test; Brixton Test), episodic memory (Logical Memory II; Autobiographical Memory Index), and working memory (Automated Symmetry Span) were assessed and tested with the examination of histograms and boxplots, skewness and kurtosis, and the Kolmogorov-Smirnov and Shapiro- Wilk tests of normality. Logical Memory II Raw Score was normally distributed. The remaining test variables, including Action Program Test Total Time, Haylings Test A and B Total Errors, Brixton Test Total Errors, Autobiographical Memory Index % Incidence Score, and Automated Symmetry Span Absolute Score were not normally distributed. All six neuropsychological test variables were standardized and the standardized values were saved as variables; z-scores were used in subsequent analyses. Normality for all standardized variables was unable to be established (skew statistics ranged from -.46 to 6.78 even after attempts at transformation).

To evaluate the relations among the neuropsychological measures and determine whether they could be reduced to composite measures of episodic memory and executive function, correlation analyses were performed. Correlations among the three executive measures and between the two episodic memory tests were weak and non-significant, thus each measure was examined independently in subsequent analyses. Pearson and Spearman correlation analyses were conducted, and results did not differ.

To explore which neurocognitive skills are important for planning, relations between measures of episodic memory, executive functions, and working memory and measures of planning were examined using correlation analyses. If a planning measure was associated with more than one neuropsychological measure, then a regression analysis was performed to determine the multivariate relations among the variables. This occurred only for Component 2 (Planning Time); a hierarchical regression was performed with the Component 2 (Planning Time) variable as the dependent variable and the Autobiographical Memory Index and Haylings Test as predictor variables entered in separate blocks. Finally, between group t-tests were performed to examine neuropsychological measures that differed between participants who opted to plan before engaging in the 2x3 MLAT (Planners) versus those who did not plan (Non-Planners). If the groups differed on more than one neuropsychological measure, then logistic regression was used to evaluate multivariate relations between neuropsychological measures of episodic memory, executive functions and working memory and 2x3 MLAT Planning Group (i.e., Planners vs. Non-planners).

All analyses were also run using non-transformed variables, and the patterns of results were not different; however, the results reported here include transformed variables. To improve interpretability of the findings, means and standard deviations of neuropsychological test variables presented in the text and tables represent the non-transformed data.

Simple Correlation Analyses of all Neuropsychological Test Variables

Correlations among the three executive measures and between the two episodic memory tests were weak and non-significant (see Table 13). The working memory variable was also not related to the other neuropsychological variables (see Table 13). Based on these results it was determined that the variables should not be reduced or combined for subsequent analyses. All mean scores are reported at the bottom of Table 13.

Table 13. *Correlation analyses (r-values) among neuropsychological test variables*

	Executive Functions			Episodic Memory		Working Memory
	Action Program Test Total Time	Haylings A and B Total Errors	Brixton Test Total Errors	Logical Memory II Raw Score	Autobiograp-hical Memory Index % Incidence Score	Automated Symmetry Span Absolute Score
Haylings A and B Total Errors	.075 (p= .48) n= 91					
Brixton Test Total Errors	.16 (p= .12) n= 91	.037 (p= .73) n= 92				

Table 13, continued

Logical Memory II Raw Score	-.14 (p= .19) n= 91	-.12 (p= .25) n= 92	-.22* (p= .04) n= 92			
Autobiogr-aphical Memory Index % Incidence Score	-.048 (p= .65) n= 91	.062 (p= .56) n= 92	-.19 (p= .08) n= 92	.15 (p= .16) n= 92		
Automated Symmetry Span Absolute Score	-.13 (p= .22) n= 87	-.029 (p= .79) n= 88	-.19 (p= .07) n= 88	.13 (p= .22) n= 88	-.11 (p= .30) n= 88	
M	74.07	3.66	6.77	22.66	75.91	16.66
(SD)	(43.93)	(3.63)	(1.48)	(6.79)	(16.64)	(9.33)
n	*n* =91	*n* = 92	*n* = 92	*n* = 92	*n* = 92	*n* = 88
range	25-244	0-14	1-10	8-39	40.74-100	0-42

Relations Between Neuropsychological Tests and Planning Variables

Correlation analyses with all six neuropsychological tests, 2x3 MLAT Online Planning Behaviors (see Aim 1), and four neuropsychological planning test variables (see Aim 2; e.g., Component 1 (Planning Accuracy), Component 2 (Planning Time), NAB Mazes Total Errors, and Zoo Map Total Planning Time) indicated mostly weak and non-significant relations among variables (see Table 14). 2x3 MLAT Online Planning Behaviors was not significantly related to any other variable, nor were NAB Mazes Total Errors, nor Zoo Map Total Planning Time. Component 1[7] was significantly related to Brixton Test Total Errors (i.e., a measure of executive functions) in the expected

[7] The variables comprising Component 1, Zoo Map Total Errors and Tower Test Achievement Score were examined in relation to Component 1 to determine the variable's variance. They were positively and negatively related, respectively, indicating that a greater Component 1 score demonstrated worse planning accuracy. Thus, the positive relation between Component 1 and Brixton errors indicates that as Component 1, planning accuracy, gets worse, Brixton errors increase.

direction (r(80)= .35 , $p < .01$), such that as planning accuracy decreased, Brixton errors increased.

Table 14. *Correlation analyses (r-values) between neuropsychological test measures and planning variables*

	Action Program Test Total Time	Haylings A and B Total Errors	Brixton Test Total Errors	Logical Memory II Raw Score	Autobiograp-hical Memory Index % Incidence Score	Automated Symmetry Span Absolute Score
2x3 MLAT Online Planning Behaviors Score	.14 (p= .22) n= 89	.023 (p= .83) n= 90	-.087 (p=.41) n= 90	.13 (p= .23) n= 90	.057 (p= .60) n= 90	.10 (p= .36) n= 86
Component 1 (Planning Accuracy)	.01 (p= .92) n= 79	.15 (p= .19) n= 80	.35* (p< .01) n= 80	-.16 (p= .16) n= 80	.028 (p= .81) n= 80	-.14 (p= .21) n= 78
Component 2 (Planning Time)	.024 (p= .84) n= 79	.22* (p= .05) n= 80	.026 (p= .82) n= 80	-.16 (p= .16) n= 80	-.22* (p= .05) n= 80	-.12 (p= .32) n= 78
NAB Mazes Total Errors	.20 (p= .06) n= 90	-.037 (p= .73) n= 91	.14 (p= .18) n= 91	-.11 (p= .31) n= 91	.091 (p= .39) n= 91	-.16 (p= .13) n= 88
Zoo Map Total Planning Time	-.11 (p= .32) n= 89	.018 (p= .86) n= 90	-.01 (p= .95) n= 90	-.12 (p= .25) n= 90	.01 (p= .95) n= 90	.058 (p= .60) n= 87

Component 2 was significantly related to Haylings A and B Total Errors (i.e., a measure of executive functions) and Autobiographical Memory Index % Incidence Score (i.e., a measure of episodic memory). However, these relations were not in the expected direction, such that as planning time increased, errors increased on the Haylings Test and memory score decreased on the AMI. Subsequent regression analysis was performed with

Component 2 as a dependent variable and the Autobiographical Memory Index and Haylings tests as independent variables entered in two separate blocks; Autobiographical Memory Index was entered first to test our prediction that planning time would be most strongly associated with measures of episodic memory. The results showed the first block (episodic memory) was significant, but the second block accounted for additional variance and was also significant (see Table 15). Thus, the hypothesis was not fully supported, and planning time was associated both with episodic memory and executive functions. Furthermore, the relations between the neuropsychological measures and planning time were not in the expected direction.

Table 15. *Regression Analysis (Beta Values)*

Component 2 (Neuropsychological Test Planning Time)	Autobiographical Memory Index % Incidence Score (episodic memory)	Haylings A and B Total Errors (executive functions)
R^2 = .046 F = 3.80, p = .05	-.22* p = .05	---
R^2 = .11 F = 4.50, p = .01 R^2 change = .06, p = .03	-.24* p = .03	.24* p = .03

Differences between Planners and Non-Planners in the 2x3 MLAT

This analysis was performed to determine which neuropsychological measures were associated with planning before engaging in the 2x3 MLAT (i.e., everyday plan formulation). It was predicted that pre-planning (i.e., plan formulation) would be associated with measures of episodic memory, such that individuals who spent more time planning would show better episodic memory than those who spent less time planning before beginning the MLAT. Contrary to prediction, t-tests between 2x3 MLAT Planners and Non-Planners indicated no significant difference in performance on either of the two episodic memory measures. The groups also did not differ on the three executive function measures or the working memory measure (t's < 1.86, $p > .07$ for all). Inspection of the means showed Non-Planners performed slightly better than Planners on the six neuropsychological tests. Because there were no significant differences between the planning groups, a logistic regression analysis was not performed.

Aim 3 Summary

Cognitive models of planning were evaluated with the administration of neuropsychological tests measuring executive functions (e.g., SAS), episodic memory, and working memory. The six variables from these tests did not correlate and were evaluated independently. It was hypothesized that planning time, the formulation level of planning, would be related to measures of episodic memory, whereas planning behaviors, a measure of the execution level of planning would be related to measures of executive functions. As predicted, online planning behaviors were related to an executive measure. Component 1- Planning Accuracy was significantly related to the Brixton Test in the expected direction, such that as planning accuracy increased, Brixton errors decreased.

Contrary to hypotheses, planning time was related to both episodic memory and executive measures and the direction of the relations was opposite of that predicted. For instance, Component 2, Planning Time was significantly related to both the Autobiographical Memory Index (AMI) and Haylings tests. Subsequent regression analysis indicated that the episodic memory measure accounted for a significant portion of variance, but the addition of the executive function measure accounted for a significant additional portion of the variance. Further, results indicated that as Planning Time increased, percent AMI incidence score decreased and Haylings errors increased. In contrast to the original conceptualization of the relations between planning time and other cognitive variables, these results suggests that participants with better executive functions and episodic memory may require less time to plan before they begin a neuropsychological task.

To further evaluate the contribution of neurocognitive processes in planning ability, planning in everyday action was examined. However, there were no significant differences on any of the neuropsychological variables (executive functions, episodic memory, working memory) between Planners and Non-planners on the 2x3 MLAT. Consistent with the results for the Component 2 – Planning Time, however, inspection of the group means suggested that Non-planners performed slightly better on several cognitive tests. Again, this result suggests that individuals with better cognitive skills may not require extensive plan formulation before performing everyday tasks. They may be capable of formulating plans extremely rapidly or may assemble task plans online during task execution.

To the best of our knowledge, this is the first study to include and examine both executive functions and episodic memory, and neurocognitive models of planning. Results show no support for the episodic memory theory. The significant relation between Component 1- Planning execution and one measure of executive functions, suggests some support for the executive account, but only for plan execution on neuropsychological measures. Neither the episodic memory nor the executive accounts were supported in analyses with measures of everyday tasks (i.e., 2x3 MLAT).

CHAPTER 4

GENERAL DISCUSSION

Although utilized in many cognitive rehabilitation approaches, the impact of planning on everyday action performance has not readily been studied empirically using ecologically valid assessments. Further, little is known about planning ability captured by traditional neuropsychological tests of planning, whether tests capture overlapping or unique planning processes, and how they correspond to planning on real-world tasks. Finally, neurocognitive models of planning are not well specified and have not been evaluated and compared in a single study. The results of the current study indicate that at least within a sample of healthy young adults: 1. Planning is not associated with better task performance, 2. Not all planning variables from traditional neuropsychological planning tests are reduced to components consistent with the double level model of planning (i.e., plan formulation and plan execution) and are not significantly related to planning in everyday tasks, and 3. Contrary to current neurocognitive models, neither executive functions nor episodic memory were strongly associated with planning ability.

Contrary to hypotheses, results from Aim 1 showed that participants who deliberately planned before performing the complex everyday tasks did not perform the task more accurately or more quickly than participants who did not plan. Nevertheless, half the sample still chose to take some time to plan before initiating the everyday tasks (i.e., Planners). Planners reported engaging with the study tasks less frequently, suggesting the possibility that they compensated for their lack of task knowledge by planning. If so, then it is possible that if Planners had been prevented from taking time to formulate a plan before task initiation, they might have performed more poorly on the everyday task. Future work is needed to empirically evaluate this possibility; however, if it were true, then it would support neuropsychological interventions that promote planning for individuals with cognitive impairment.

Results from Aim 2 partially supported the hypothesis that neuropsychological measures may reflect one of two planning stages/levels—formulation and execution. PCA results showed six variables from three neuropsychological tests of planning were reduced to the two components of the double-level panning model (Planning Time and Planning Accuracy). However, not all planning variables reflected one of these constructs. This suggests that some neuropsychological variables may capture both plan formulation and execution. The Tower of London was the only planning task from which variables reflecting each planning component could be derived. These results suggest that care should be taken when selecting variables to reflect plan formulation versus plan execution.

Results of Aim 3 were unexpected, as relatively few relations were observed among measures of executive function, episodic memory, and working memory and measures of planning. However, as predicted, plan execution for traditional neuropsychological planning tests was related to an executive functioning measure. Planning time for traditional neuropsychological planning tests was related to *both* episodic memory and executive function measures. However, these significant relations were not in the expected direction. This result suggests that individuals with lower cognitive abilities may use lengthy plan formulation as a strategy when performing challenging tasks.

Overall, this study addressed many gaps in the literature and has numerous strengths. First, this study examined a large sample of participants, especially when compared to other studies of planning. Second, this study showed that planning time, plan quality, and online planning behaviors could be reliably coded using a performance-based measure of everyday action, which is promising to improve the ecological-validity of planning assessment. With regard to Aim 1 (evaluating planning in everyday tasks), this is one of the only studies to our knowledge to exhaustively examine the formulation level of planning by measuring both planning time and plan quality. Planning time, the duration of time between the end of instructions and start of the task is often utilized as the only variable of planning prior to task execution. However, planning time is often criticized because it is unclear what participants are engaged in during the elapsed time (Unterrainer et al., 2003). Van Beilen et al. (2005) attempted to overcome this limitation by providing participants time alone (during which they were presumably forming a plan) prior to executing the Cognitive Effort Test (CET; a performance-based measure incorporating a computer task, looking up phone numbers in a phonebook, and threading three nuts down three screws). After sixty seconds the experimenter re-entered the room and prompted participants to detail how they planned to complete the tasks. However that method still raises questions about whether participants just came up with a plan in response to the prompt, the impact of memory when recalling their plans, and so on. By contrast, the *current study's* methods required participants who chose to plan to verbalize their plans in real-time. The length of their verbalization was captured as the planning time interval. This method did not require assumptions that participants were actively engaged and thinking about the tasks prior to task execution.

The results on the everyday action task are inconsistent with prior studies showing planning time to be positively associated with task performance. Seter et al. (2011) have shown more time spent planning as a proportion of total time on everyday tasks was related to better task performance in psychiatric and healthy control samples. Also, studies on traditional neuropsychological planning measures (e.g., the Tower Test and related variants, Tower of London, Tower of Hanoi, Stockings of Cambridge, etc.) have shown that directions that emphasize planning lead to improved performance in healthy participants (Unterrainer et al., 2004; Unterrainer et al., 2003). Planning time has even shown to increase as a function of task difficulty, as participants classified as "good" performers on the Tower of London demonstrated planning times that were twice as long as "intermediate" and "poor" performers, and "good" performers showed the strongest increase in planning times with increased task difficulty (Unterrainer et al., 2004). Longer planning times on similar tasks have also been associated with greater accuracy (Unterrainer et al, 2004; Phillips et al., 2001). The difference between the current and past studies is that the current study includes only healthy participants on everyday tasks. It is possible that lengthy plan formulation on everyday tasks is only beneficial when one has cognitive difficulties or is less familiar with everyday tasks.

With regard to Aim 2, this is also one of the only studies to our knowledge to examine and compare neuropsychological tests of planning to each other and to planning in everyday tasks. Little is known about the relations among traditional neuropsychological measure of planning, let alone the relation of these measures to more ecologically valid planning measures. The current results suggest caution when using traditional neuropsychological measures of planning to understand planning on everyday tasks. A drawback of the rehabilitation literature is that intervention studies do not usually include planning measures (traditional or functional) for either assessment of baseline abilities or evaluation of outcomes (Green et al., 2008; Rodewald et al., 2011). This study suggests that naturalistic planning variables may be reliably obtained and should be included alongside traditional measures of planning in future intervention studies, as these measures assess different abilities. Future research is needed to determine whether the naturalistic planning variables are more strongly associated with functional outcomes, such as occupational or academic success; however, the face validity of naturalistic measures may promote patient engagement more than traditional neuropsychological measures.

Finally, the current study is one of the only studies to our knowledge to examine proposed cognitive components of planning from different models in one study (Aim 3). The contribution of executive functions, episodic memory, and working memory were evaluated in *both* planning in everyday action and planning captured on traditional neuropsychological tests. The results of the study did not robustly support any of the planning models described in the literature. This was surprising, as the CS/SAS and episodic memory accounts of planning have influenced planning research and rehabilitation approaches. It is worth noting that although the models claim to explain planning in everyday life, current support for the models comes largely from studies using abstract tasks (e.g., puzzles). This underscores the need for studies like ours to evaluate the specific processes involved in planning ability and move theory toward more detailed descriptions that apply to complex, real-world activities.

Despite the novelty and the numerous strengths of the current study, this study also had some limitations and there are many remaining questions about the assessment of planning ability and impact on rehabilitation strategies. For instance, although participants were given a choice to plan before task execution to maximize naturalistic conditions and capture planning tendencies, it may have been beneficial to assign participants into planning and non-planning groups. Group assignment would have eliminated the confounds (e.g., task knowledge) associated with participant's natural tendency to plan or not plan. It is also possible that a more detailed assessment of plan quality (e.g., word count, extraneous task steps) may shed more light on the contribution of planning to performance, and the depth of planning that is necessary for exceptional performance (e.g., no errors, increased efficiency and speed) or poor performance. Examining whether formulated plans match executed plans may also elucidate the influence of plans on everyday functioning. For example, it is possible that individuals who followed their plan too closely or rigidly may have performed more poorly than individuals who were more flexible in their plan adherence. Alternatively, participants may have not adhered to their plan because they failed to recall the plan or had difficulty monitoring their performance. More fine-grained behavioral measures, such as eye-tracking methods, may also enhance the assessment of planning in everyday tasks by capturing the direction of attention. For example, efficient performers on tasks like the Tower Test have shown a different pattern of time-related eye-movements during the task than poor performers (Hodgson, Bajwa, Owen, & Kennards, 2000). It is possible that plans influenced efficient attentional focus, but that this was not evident in the overt behavioral measures that were collected in the current study. Further, although the 2x3

MLAT is a performance-based measure of everyday action, which greatly improves its ecological validity, it is still utilized within a laboratory setting. Although this is the recommended procedure when direct observation in naturalistic settings is not possible (McKibbin et al., 2004), it would be beneficial to conduct similar studies in familiar environments for participants (i.e. their homes or with their own materials/objects), or use video-streaming to the laboratory from participant homes when they are engaged in everyday tasks. Finally, it is difficult to know the extent to which research participants' exerted maximal effort on the study tasks. Investigators have shown that healthy college participants commonly exert sub-optimal effort on neuropsychological measures, which could create substantial error variance in test scores (An, Zakzanis, Joordens, 2012). Future research should incorporate formal measures of effort when conducting studies of healthy college participants.

Another caveat is that the extent to which these results may generalize to patient populations is unknown. Research with patient populations or people with certain cognitive impairments may capture some benefits of planning in everyday action (Seter et al., 2011), or show greater positive or negative relations between planning and performance. Moreover, research with certain homogeneous patient samples may better highlight different planning components in everyday action and neuropsychological planning tests. Research has shown discrepant results between normal and clinical populations utilizing factor reduction techniques for cognitive constructs, with assessment measures suggesting shared variance in the intact brain, but dissociation and unique variance in the damaged brain (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003). It is also important to consider that many cognitive rehabilitation approaches include strategies to improve plan formulation and deliberate planning behavior [e.g., Goal Management Training (Levine et al., 2000); Plan-a-Day (Holt et al., 2011); Autobiographical Episodic Memory Cueing Procedure (Hewitt et al., 2006)], and planning deficits are believed to be debilitating and are commonly reported across a range of patient populations (Burgess & Robertson, 2002). Thus, there is still much more to understand about planning in everyday tasks in people with cognitive deficits.

There are important lessons to be learned from this study. First, although results are contrary to prior studies (Seter et al., 2011), this is the first study to our knowledge to directly and reliably assess *both* plan formulation and planning behaviors during the performance of everyday tasks. This method holds promise for improving the ecological and face validity of planning assessments. Second, the results from Aims 2 and 3 support the need to carefully consider the selection of planning tests and to consider plan formulation and execution separately when evaluating planning ability. Moreover, since neuropsychological measures of planning and everyday action planning did not associate, it is recommended that clinicians exercise great caution when drawing conclusions about everyday functioning from neuropsychological measures of planning. Third, taken together, the results indicate the need for more studies examining the neurocognitive components of planning and more studies to evaluate the neurocognitive models of planning, which may best be evaluated using patient populations. This study is encouraging for future research to examine planning with different populations and extended methods. The future of planning conceptualization, assessment, and rehabilitation rests in research that is integrative and utilizes existing models, measures, and remediation designs.

REFERENCES

Allain, P., Nicoleau, S., Pinon, K., Etcharry-Bouyx, F., Barré, J., Berrut, G.,...Le Gall, D. (2005). Executive functioning in normal aging: A study of action planning using the Zoo Map Test. *Brain and Cognition, 57*, 4-7. doi:10.1016/j.bandc.2004.08.011

An, K. Y., Zakzanis, K. K., Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology, 27*(8), 849-857. doi:10.1093/arclin/acs085

Barch, D. M., Berman, M. G., Engle, R., Jones, J. H., Jonides, J., MacDonald, A., Nee, D. E., … Sponheim, S. R. (2009). CNTRICS final task selection: Working memory. *Schizophrenia Bulletin, 35*(1), 136-152. doi:10.1093/schbul/sbn153

Barch, D. M., Carter, C. S., Braver, T. S., Sabb, F. W., MacDonald, A., Noll, D. C., & Cohen, J. D. (2001). Selective deficits in prefrontal cortex function in mediation-naïve patients with schizophrenia. *Archives of General Psychiatry, 58*, 280-288. doi:10.1001/archpsyc.58.3.280

Barch, D. M., & Smith, E. (2008). The cognitive neuroscience of working memory: Relevance to CNTRICS and schizophrenia. *Biological Psychiatry, 64*, 11-17. doi:10.1016/j.biopsych.2008.03.003

Bartlett, F. C. (1932). *Remembering*. Cambridge, UK: Cambridge University Press.

Bickerton,W-L., Humphreys, G. W., & Riddoch, M. J. (2007). The case of the unfamiliar implement: Schema-based over-riding of semantic knowledge from objects in everyday action. *Journal of the International Neuropsychological Society, 13*, 1035-1047. doi:10.10170S1355617707071585

Bielak, A. A. M., Mansueti, L., Strauss, E., & Dixon, R. A. (2006). Performance on the Hayling and Brixton tests in older adults: Norms and correlates. *Archives of Clinical Neuropsychology, 21*, 141-149. doi:10.1016/j.acn.2005.08.006

Burgess, P. W., Alderman, N., Evans, J., Emslie, H., & Wilson, B. A. (1998). The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society, 4*, 547-558. doi:10.1017/S1355617798466037

Burgess, P. W., & Robertson, I. H. (2002). Principles of the rehabilitation of frontal lobe function. In D. T. Stuss, & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 557-572). New York, NY: Oxford University Press.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests*. Thurston, Suffolk: Thames Valley Test Company.

Burgess, P. W., Veitch, E., Costello, A. de L., & Shallice, T. (2000). The cognitive and neuroanatomical correlates of multitasking. *Neuropsychologia, 38*, 848-863. doi:10.1016/S0028-3932(99)00134-7

Buxbaum, L. J., Schwartz, M. F., & Carew, T. G. (1997). The role of semantic memory in object use. *Cognitive Neuropsychology, 14*, 219-254. doi:10.1080/026432997381565

Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., & Grafman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesions. *Neuropsychologia, 38*, 655–665. doi:10.1016/S0028-3932(99)00102-5

Chevignard, M., Pillon, B., Pradat-Diehl, P., Taillefer, C., Rousseau, S., Le Bras, C., & Dubois, B. (2000). An ecological approach to planning dysfunction: Script execution. *Cortex, 36*, 649-669. doi:10.1016/S0010-9452(08)70543-4

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Conway, A R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12*, 769-786. Retrieved from http://www.psychonomic.org/PBR/forthcoming.htm

Davalos, D. B., Green, M., & Rial, D. (2002). Enhancement of executive functioning skills: An additional tier in the treatment of schizophrenia. *Community Mental Health Journal, 38*, 403-411. doi:10.1023/A:1019860428856

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society, 9*, 936-946. doi:10.1017/S1355617703960139

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System-Technical Manual*. San Antonio, TX: The Psychological Corporation.

Elvevåg, B., & Goldberg, T. E. (2000). Cognitive impairment in schizophrenia is the core of the disorder. *Critical Reviews in Neurobiology, 14*(1), 1-21. Retrieved from http://www.begellhouse.com/journals/7b004699754c9fe6.html

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers, 28*, 1-11.

Evans, J. D., Heaton, R. K., Paulsen, J. S., Palmer, B.W., Patterson, T., & Jeste, D. V. (2003). The relationship of neuropsychological abilities to specific domains of functional capacity in older schizophrenia patients. *Biological Psychiatry, 53,* 422-430. doi:10.1016/S0006-3223(02)01476-2

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). Washington, DC: Sage Publications.

Giovannetti, T., Schwartz, M. F., & Buxbaum, L. J. (2007). The Coffee Challenge: A new method for the study of everyday action errors. *Journal of Clinical and Experimental Neuropyschology, 29*, 690-705. doi:10.1080/13803390600932286

Grafman, J. (1989). Plans, actions, and mental sets: Managerial knowledge units in the frontal lobes. In E. Perecman (Ed.), *Integrating Theory and Practice in Clinical Neuropsychology* (pp.93-138). Hillsdale, NJ: Lawrence Erlbaum Associates.

Green, M. F., Nuechterlein, K. H., Kern, R. S., Baade, L. E., Fenton, W. S., Gold, J. M.,… Marder, S. R. (2008). Functional co-primary measures for clinical trials in schizophrenia: results from the MATRICS Psychometric and Standardization Study. *American Journal of Psychiatry*, *165*(2)**,** 221-228.

Harvey, P. D., Green, M. F., Keefe, R. S. E., & Velligan, D. I. (2004). Cognitive functioning in schizophrenia: A consensus statement on its role in the definition and evaluation of effective treatments for the illness. *Journal of Clinical Psychiatry, 65*(3)*,* 361-372. Retrieved from http://www.psychiatrist.com/default2.asp#

Hodgson, T. L., Bajwa, A., Owen, A. M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower of London task. *Journal of Cognitive Neuroscience, 12*, 894-907. doi:10.1162/089892900562499

Holt, D. V., Rodewald, K., Rentrop, M., Funke, J., Weisbrod, M., & Kaiser, S. (2011). The Plan-a-Day approach to measuring planning ability in patients with schizophrenia. *Journal of the International Neuropsychological Society, 17*, 327-335. doi:10.1017/S1355617710001712

Hewitt, J., Evans, J. J., & Dritschel, B. (2006). Theory driven rehabilitation of executive functioning: Improving planning skills in people with traumatic brain injury through the use of an autobiographical episodic memory cueing procedure. *Neuropsychologia, 44*, 1468-1474. doi:10.1016/j.neuropsychologia.2005.11.016

Joe, W. Q., Ferraro, M., & Schwartz, M. F. (2002). Sequencing and interleaving in routine action production. *Neurocase, 8*, 135-150. doi:10.1093/neucas/8.1.135

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189– 217. doi:10.1037/0096-3445.133.2.189

Kessler, R. K., Giovannetti, T., & MacMullen, L. (2007). Everyday action in schizophrenia: Performance patterns and underlying cognitive mechanisms. *Neuropsychology, 21*, 439-447. doi:10.1037/0894-4105.21.4.439

Klapow, J. C., Evans, J., Patterson, T. L., Heaton, R. K., Koch, W. L., & Jeste, D. V. (1997). Direct assessment of functional status in older patients with schizophrenia.

American Journal of Psychiatry, 154(7), 1022-1024. Retrieved from http://ajp.psychiatryonline.org/index.dtl

Klosowska, D. (1976). Relation between ability to program actions and location of brain damage. *Polish Psychological Bulletin, 7*, 245-255. Retrieved from http://versita.metapress.com/content/121138/

Kopelman, M. N., Wilson, B. A., & Baddeley, A. D. (1990). *The Autobiographical Memory Interview*. Suffolk: Thames Valley Test Company.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310

Lawson, M. J., & Rice, D. N. (1989). Effects of training of executive strategies on a verbal memory problem resulting from closed head injury. *Journal of Clinical and Experimental Neuropsychology, 11*, 842-854. doi:10.1080/01688638908400939

Levine, B., Robertson, I. H., Clare, L., Carter, G., Hong, J., Wilson, B. A.,...Stuss, D. T. (2000). Rehabilitation of executive functioning: An experimental-clinical validation of goal management training. *Journal of the International Neuropsychological Society, 6*, 299-312. doi:10.1017/S1355617700633052

Levine, B., Stuss, D. T., Milberg, W. P., Alexander, M. P., Schwartz, M., MacDonald, R. (1998). The effects of focal and diffuse brain damage on strategy application: Evidence from focal lesions, traumatic brain injury and normal aging. *Journal of the International Neuropsychological Society, 4*(3), 247-264. Retrieved from http://journals.cambridge.org/action/displayJournal?jid=INS

Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment*. (4th ed.). Oxford, UK: Oxford University Press.

Luria, A. R. (1978). *Les fonctions corticales supérieures de l'homme*. Paris, FR: Presses Universiatires de France.

Luria, A. R. (1980). *Higher cortical functions in man* (2nd ed.). New York, NY: Consultants Bureau Enterprises.

Martens, S., & Johnson, A. (2009). Working memory capacity, intelligence, and the magnitude of the attentional blink revisited. *Experimental Brain Research, 192*, 43-52. doi:10.1007/s00221-008-1551-1

Mattson, M. E., & Baars, B. J. (1992). Laboratory induction of nonspeech action errors. In B. J. Baars (Ed.). *Experimental slips and human error: Exploring the architecture of volition* (pp. 151-191). New York: Pelnum.

McKibbin, C. L., Brekke, J. S., Sires, D., Jeste, D. V., & Patterson, T. L. (2004). Direct assessment of functional abilities: Relevance to person with schizophrenia. *Schizophrenia Research, 72*, 53-67. doi:10.1016/j.schres.2004.09.011

Nuechterlein, K. H., Barch, D. M., Gold, J. M., Goldberg, T. E., Green, M. F., & Heaton, R. K. (2004). Identification of separable cognitive factors in schizophrenia. *Schizophrenia Research, 72*, 29-39. doi:10.1016/j.schres.2004.09.007

Norman, D. A., & Shallice, T. (1980). Attention to action: Willed and automatic control of behavior. *Center for Human Information Processing Technical Report* 99. San Diego, University of California.

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behviour. In G. E. Schwartz, & D. Shapiro (Eds.). *Consciousness and Self-Regulation*. New York: Plenum Press.

Owen, A. M. (1997). Cognitive planning in humans: Neuropsychological, neuroanatomical and neuropharmacological perspectives. *Progress in Neurobiology, 53*, 431-450. doi: 10.1016/S0301-0082(97)00042-7

Pantelis, C., Barnes, T. R., Nelson, H. E., Tanner, S., Weatherley, L., Owen, A.M., & Robbins, T. W. (1997). Frontal-striatal cognitive deficits in patients with chronic schizophrenia. *Brain, 120*, 1823-1843. doi:10.1093/brain/120.10.1823

Patterson, T. L., Goldman, S., McKibbin, C. L., Hughs, T., & Jeste, D. V. (2001). USCD performance-based skills assessment: Development of a new measure of everyday functioning for severely mentally ill adults. *Schizophrenia Bulletin, 27*, 235-245. doi:10.1093/schbul/sbm014

Patterson, T. L., Klapow, J. C., Eastham, J. H., Heaton, R. K., Evans, J. D., Koch, W. L., & Jeste, D. V. (1998). Correlates of functional status in older patients with schizophrenia. *Psychiatry Research, 80,* 41-52. doi:10.1016/S0165-1781(98)00060-2

Penfield, W. & Evans, J. (1935). The frontal lobe in man: A clinical study of maximal removals. *Brain, 58*, 115-133. doi:10.1093/brain/58.1.115

Phillips, L. H., Wynn, V. E., McPherson, S., & Gilhooly, K. J. (2001). Mental planning and the Tower of London task. *The Quarterly Journal of Experimental Psychology, 54*, 579-597. doi:10.1080/713755977

Rodewald, K., Rentrop, M., Holt, D. V., Roesch-Ely, D., Backenstrab, M., Funke, J.,...Kaiser, S. (2011). Planning and problem-solving training for patients with schizophrenia: A randomized controlled trial. *BMC Psychiatry, 11*, 1-26. doi:10.1186/1471-244X-11-73

Schacter, D. L., & Addis, D. R. (2007). The cognitive neuroscience of constructive memory: Remembering the past and imagining the future. *Philosophical Transactions of the Royal Society B, 362*, 773-786. doi:10.1098/rstb.2007.2087

Schacter, D. L., Addis, D. R., Buckner, R. L. (2007). Remember the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience, 8*, 657-661. doi:10.1038/nrn2213

Schrock, J. C., & Engle, R. W. (2005). Automated symmetry span [An E-prime file]. Retrieved 08-01-2011, from http://psychology.gatech.edu/renglelab/

Schwartz, M. F., Montgomery, M. W., Buxbaum, L. J., Less, S. S., Carew, T. G., Coslett, H. B., …Mayer, N. (1998). Naturalistic action impairment in closed head injury. *Neuropsychology, 12*, 13-28. doi:10.1037/0894-4105.12.1.13

Schwartz, M. F., Segal, M., Veramonti, T., Ferrara, M., & Buxbaum, L. (2002). The Naturalistic Action Test: A standardized assessment for everyday action impairment. *Neuropsychological Rehabilitation, 12*, 311-339. doi:10.1080/09602010244000084

Semkovska, M., Bedard, M. A., Godbout, L., Limoge, F., & Stip, E. (2004). Assessment of executive dysfunction during activities of daily living in schizophrenia. *Schizophrenia Research, 69,* 289-300. doi:10.1016/j.schres.2003.07.005

Semkovska, M., Stip, E., Godbout, L., Paquet, F., & Bedard, M.A. (2002). Behavioral disorganization in schizophrenia during a daily activity: The Kitchen Behavioral Scoring Scale. *Brain and Cognition, 48*(2-3)*,* 546-553. Retrieved from http://www.elsevier.com/wps/find/journaldescription.cws_home/622798/description#description

Seter, C., Giovannetti, T., Kessler, R., & Worth, S. (2011). Everyday action planning in schizophrenia. *Neuropsychological Rehabiliation, 21*, 224-249. doi:10.1080/09602011.2010.544519

Sevy, S., & Davidson, M. (1995). The cost of cognitive impairment in schizophrenia. *Schizophrenia Research, 17,* 1-3. doi:10.1016/0920-9964(95)00025-H

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 298*, 199-209. doi:10.1098/rstb.1982.0082

Shallice, T., & Burgess, P. W. (1991). Higher-order cognitive impairments and frontal lobe lesions in man. In H. S. Levin, H. M. Eisenberg, & A. L. Bentos (Eds.), *Frontal Lobe Function and Dysfunction* (pp. 125–138). New York, NY: Oxford University Press.

Sohlberg, M. M., & Mateer, C. A. (2001). *Cognitive rehabilitation: An integrative neuropsychological approach*. New York, NY: The Guilford Press.

Stern, R. A., & White, T. (2003). *Neuropsychological Assessment Battery-Administration, Scoring, and Interpretation Manual.* Lutz, FL: Psychological Assessment Resources.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*(1), 104-132. doi:10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37(3)*, 498-505. Retrieved from http://www.springer.com/psychology/cognitive+psychology/journal/13428

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*(6), 635-654. doi:10.1080/09658210902998047

Unterrainer, J. M., & Owen, A. M. (2006). Planning and problem solving: From neuropsychology to functional neuroimaging. *Journal of Physiology-Paris, 99*, 308-317. doi:10.1016/j.jphysparis.2006.03.014

Unterrainer, J. M., Rahm, B., Kaller, C., C., Quiske, K., Hoppe-Seyler, K., Meier, C.,…Halsband, U. (2004). Planning abilities and the Tower of London: Is this task measuring a discrete cognitive function? *Journal of Clinical Experimental Neuropsychology, 26*, 846-856. doi:10.1080/13803390490509574

Unterrainer, J. M., Rahm, B., Leonhart, R., Ruff, C. C., & Halsband, U. (2003). The Tower of London: The impact of instructions, cueing, and learning on planning abilities. *Cognitive Brain Research, 17*, 675-683. doi:10.1016/S0926-6410(03)00191-5 van Beilen, M., van Zomeren E. H., van den Bosch, R. J., Withaar, F. K., & Bouma, A. (2005). Measuring the executive functions in schizophrenia: The voluntary allocation of effort. *Journal of Psychiatric Research, 39*, 585-593. doi:10.1016/j.jpsychires.2005.02.00

van den Berg, E., Nys, G. M., Brands, A. M. A., Ruis, C., Van Zandvoort, M. J. E., & Kessels, R. P. C. (2009). The Brixton Spatial Anticipation Test as a test for executive function: Validity in patient groups and norms for older adults. *Journal of the International Neuropsychological Society, 15*, 695-703. doi:10.1017/S1355617709990269

Velligan, D. I., & Bow-Thomas, C. C. (1999). Executive function in schizophrenia. *Seminars in Clinical Neuropsychiatry, 4*(1), 24-33.

Velligan, D. I., Diamond, P., Glahn, D. C., Ritch, J., Maples, N., Castillo, D., & Miller, A. L. (2007). The reliability and validity of the Test of Adaptive Behaviors in Schizophrenia (TABS). *Psychiatry Research, 151,* 55-66. doi:10.1016/j.psychres.2006.10.007

Wechsler, D. (2009a). *Wechsler Memory Scale- Fourth Edition: Administration and Scoring Manual*. San Antonio, TX: Pearson, PsychCorp.

Wechsler, D. (2009b). *Wechsler Memory Scale- Fourth Edition: Technical and Interpretive Manual*. San Antonio, TX: Pearson, PsychCorp.

White, T., & Stern, R. A. (2003). *Neuropsychological Assessment Battery- Psychometric and Technical Manual*. Lutz, FL: Psychological Assessment Resources.

Wilson, B. A., Alderman, N., Burgess, P.W., Emslie, H. C., & Evans, J. J. (1996). *The Behavioural Assessment of the Dysexecutive Syndrome*. Bury St. Edmunds, UK: Thames Valley Test Company.

Yatham, L. N., Torres, I. J., Malhi, G. S., Frangou, S., Glahn, D. C., Bearden, C. E.,…Chengappa, K. N. R. (2010). Consensus article: The international society for bipolar disorders-battery for assessment of neurocognition (ISBD-BANC). *Bipolar Disorders, 12*, 351-363. doi:10.1111/j.1399-5618.2010.00830.x