Strategic Flexibility and Age-Related Cognitive Change

Daniel Barulli

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2019

ABSTRACT

Strategic Flexibility and Age-Related Cognitive Change

Daniel Barulli

This series of projects aims to explore the potential role of strategic flexibility in cognitive aging, and whether this construct can serve as an effective mechanistic proxy for cognitive reserve. Study 1 introduces the task designed for this series, based on stimuli from a classic test of fluid reasoning and formatted as a task-switching paradigm to explore strategic characteristics in a structured way. This study suggests that such a task is subject to age-related effects. Study 2 introduces a redesigned version of this task, matching it more closely to existing paradigms of task-switching, and explores how covariates interact with measured performance. Study 3 draws upon an existing sample of extensive neuropsychological and neuroimaging data, and aims to describe the associations among this set of data and measures of strategic flexibility. Results overall indicate that age negatively affects strategic flexibility, but cognitive reserve may mitigate this impairment.

TABLE OF CONTENTS

List of Figures & Tables iii

Acknowledgements iv

Dedication v

Chapter 1: General Introduction 1

Introduction 1

Project Motivation 6

Project Outline 11

Chapter 2: Literature Review 16

Cognitive Reserve 16

Cognitive Strategies 17

Age-Related Strategy Differences 18

Cognitive Strategies & Executive Functions 23

Fluid Intelligence: Raven's Matrices 26

Strategy Flexibility & Task Switching 27

Chapter 3: Study One 44

Introduction 44

Methods 49

Results 57

Discussion 70

Chapter 4: Study Two 75

Introduction 75

Methods 79

Results 87

Discussion 102

Chapter 5: Study Three 104

Introduction 104

Methods 109

Results 115

Discussion 135

Chapter 6: General Discussion 138

References 141

LIST OF FIGURES & TABLES

Figure 1: Task Structure……………………………51
Figure 2: Sample Logico-Analytic Item………53
Figure 3: Sample Visuospatial Item……………54
Table 1: Study 1 Participants……………………58
Table 2: ANOVA: Global RT………………………59
Table 3: Post-Hoc Test: Global RT………………60
Table 4: ANOVA: Global ACC……………………61
Table 5: Post-Hoc: Global ACC…………………62
Figure 4: RT Across Block 1……………………63
Figure 5: RT Across Block 2……………………64
Table 6: ANOVA: Local RT………………………65
Table 7: Post-Hoc: Local RT……………………66
Table 8: ANOVA: Local ACC……………………67
Table 9: Post-Hoc: Local ACC…………………68
Table 10: ANOVA: Mixing RT…………………69
Table 11: Post-Hoc: Mixing RT………………70
Table 14: Study 2 Participants…………………87
Table 16: ANOVA: Global RT…………………88
Table 17: Post-Hoc: Global RT………………89
Table 18: ANOVA Global ACC…………………90
Table 19: Post-Hoc: Global ACC………………91
Table 20: ANOVA: Local RT……………………92
Table 21: Post-Hoc: Global ACC………………93
Table 22: ANOVA: Mixing RT…………………94
Table 23: Post-Hoc: Mixing RT………………95
Table 24: ANOVA: Mixing ACC………………96
Table 25: Post-Hoc: Mixing ACC ……………97
Figure 6: RT by Age Group ……………………98
Figure 7: RT by Block Type……………………98
Figure 8: RT by Rule Type………………………99
Figure 9: RT by Strategy Type…………………99
Table 26/27: Mixed Effects Models…………101
Table 28: Study 3 Participants………………115
Table 29: ANOVA: Global RT…………………116
Table 30: Post-Hoc: Global RT………………117
Table 31: ANOVA: Global ACC………………118
Table 32: Post-Hoc: Global ACC……………119
Table 33: ANOVA: Local RT…………………120
Table 34: Post-Hoc: Local RT………………121
Table 35: ANOVA: Local ACC………………122
Table 36: Post-Hoc: Local ACC……………123
Table 37: ANOVA: Mixing RT………………124
Table 38: Post-Hoc: Mixing RT……………125
Table 39: ANOVA: Mixing ACC……………126
Table 40: Post-Hoc: Mixing ACC…………127
Figure 10: Gray Matter by Flexibility ………129
Table 42: Regression Analyses………………131
Figure 11: Correlation Matrix ………………132
Table 43: CR Mediation…………………………134
Table 44: Flexibility Mediation………………135

ACKNOWLEDGEMENTS

Many thanks to Dr. Yaakov Stern for his mentorship during and preceding this project. This project also benefitted greatly from the input of my colleagues at the Taub Institute, particularly Drs. Christian Habeck and Jason Steffener for their advise on both statistical and conceptual methods. Thanks also to Dr. Peter Balsam and Dr. Lila Davachi, who helped to bring in perspectives diverging from my own and greatly improved the interpretation of these findings. Any errors or lapses in straightforward communication in these pages are my own.

DEDICATION

Dedicated to my mother, Barbara Barulli. I hope this makes up for the unsentimental birthday cards.

CHAPTER 1
GENERAL INTRODUCTION

Aging and dementia represent an impending public health crisis (Whalley & Smith, 2013). Cognitive reserve (CR) is a latent factor theorized to allow some individuals to better cope with the deleterious effects of aging and age-related neurological illness better than others who have not built up such a reserve, which has previously been associated epidemiologically with individual differences in exposure variables such as education, leisure activities, occupational history, and verbal IQ (Stern, 2002). As such, a thorough understanding of how such a factor operates may allow researchers and public health professionals to understand, predict, and potentially even intervene on the future prevalence and distribution of dementia and cognitive impairment within the population. While many studies of CR have been conducted, most of these fall into the realms of epidemiology on the one hand, looking at the wide-scale effects of lifestyle factors on disease prevalence, or neuroimaging on the other (see Barulli & Stern, 2013 for a review). This series of studies instead attempts to understand CR as a function of cognitive operations, specifically the differential utilization of and flexibility with alternative cognitive strategies.

Much research within cognitive aging studying compensatory mechanisms in aging focuses on the neural basis of CR and attempts to explain this construct in terms of its underlying neural mechanisms. Neural reserve is one such proposed mechanism, and can be further broken down into the distinct mechanisms of the efficiency and capacity of a given task-related neural network, where efficiency is commonly thought to mean the ability of that network to accomplish a given function with some limited amount of activation, while capacity refers generally to the ability of that network to activate to the threshold at which the function attempted can be achieved (Stern, 2009).

While the study of CR using these hypothetical mechanisms and the tools provided by neuroimaging is important and valuable, as some have pointed out terms like efficiency and capacity are often too amorphous and hypothetical to always be truly explanatory (Poldrack, 2015). "Efficiency" in particular may simply be a placeholder for several much more detailed mechanistic explanations, including a "more efficient" network reflecting: a) an entirely different set of cognitive processes being performed (i.e. a different task being conducted altogether), b) a different neural computation being performed to complete the same task, or c) less metabolic resources being demanded to perform the same neural computation for the same duration and at the same intensity. Without understanding something of the cognitive processes that are occurring during fMRI imaging, and not merely the ultimate cognitive performance, it is impossible to disambiguate between these possibilities.

Neural compensation is the second major mechanism proposed to explain CR, and this is deeply rooted in a wealth of neuroimaging literature considering the active recruitment of secondary networks and areas in order to perform some cognitive task when the primary network has become impaired or overwhelmed (Li et al., 2009; Reuter-Lorenz & Cappell, 2008). The major problem with this approach is that without additional information it cannot make firm predictions that link secondary sources of activation with performance; in some, additional activation of a secondary network may be related to better cognitive performance. In others however the additional compensatory activation is associated with diminished cognitive performance. This can be understood using the analogy of walking with a cane; those who require the use of a cane do not walk better than those who don't require its use, but they do walk better whilst using the cane than those who require one but are not using it (Stern, 2009). This analogy effectively parses secondary networks into two distinct categories: (cane-like) compensatory networks, and enhancing networks. The exact role that each network plays in an individual's performance may be heavily dependent on the unique attributes of that individual's brain structure and function.

Understanding the differences between such networks requires more than the observation of task performance and associated activation: ideally some understanding of the cognitive operations reflected by the distinct networks can be understood, with one plausible working hypothesis being that primary networks are associated with the most direct and explicit cognitive operations required to perform a given task, secondary enhancing networks are associated with additional cognitive operations that facilitate primary network function, and secondary compensatory networks are associated with some alternative set of cognitive operations that are less ideal for the given task but can perform it in a suboptimal fashion when the primary network is impaired due to damage.

The introduction of primary and secondary sets of cognitive operations into this discourse is evocative of an existing literature that has until recently been largely isolated from the world of cognitive aging: cognitive strategies. While some researchers speculate that fundamental cognitive capabilities such as attention (Roberson, 2013) or working memory (Sandry & Sumowski, 2014) may mediate between CR and overall cognitive performance, this line of argumentation about alternative networks potentially representing alternative sets of cognitive operations suggests instead that cognitive strategies may represent one of the dominant mechanism through which CR moderates the effects of age on cognitive function. Some work has already been conducted in the realm of mnemonic strategies (Woods et al., 2010), with one study showing that mnemonic strategy use moderates the effects of HIV-related dementia on verbal working memory, and such strategy use has been shown to be associated with proxies for higher CR such as verbal intelligence and socioeconomic status.

We have previously conducted two preliminary studies of cognitive reserve in the context of strategy use. In the first study (Barulli et al., 2013), a computational estimation task was used which required participants to solve 100 two-digit 2 x 2 multiplication problems using one of two estimation strategies: either rounding both numbers up to the nearest decade and multiplying (e.g. '32x56' becomes '40x60'), or rounding both numbers down to the next lowest decade and multiplying (e.g. '30x50' in the example above). This task is designed such that all problems are amendable to either strategy, however one strategy will always result in a better answer in that its result will be mathematically closer to the actual non-rounded product (e.g. in the above example, rounding down is the better strategy because the result, 1500, is much closer to the actual product of 1792).

Using this task in a sample of 20 healthy young adults (20-31 year-olds) and 18 healthy older adults (62-77 year-olds), we found that older adults performed more poorly on this task as expected, as measured using their mean percent use of the best strategy across all trials. However, proxy measures of CR in the form of verbal IQ moderated the effect of age group on strategy selection such that while verbal IQ variation did not affect the performance of the young group, those elders with greater verbal IQ performed better on the task and chose the best strategy more often than those with lower verbal IQ. These differential results for young and old hinted that CR specifically, above and beyond just IQ, may be associated with cognitive strategy selection.

Next we conducted an extension and replication of these results, incorporating a greater number of measures and testing subjects who had also received T1-weighted structural MRI scans. Performance on the cognitive control factor of the NIH EXAMINER (Kramer et al., 2014) battery was also assessed to obviate the possibility that executive abilities were mediating the relationship between age and strategy selection, which might suggest that any observed relationship with CR may be explained simply by the commonly proposed mechanistic relationship between CR and executive abilities (Hodzik & Lemaire, 2007). Basic arithmetical abilities were also included as a covariate to preclude the chance that those with higher CR simply had greater domain-specific knowledge. Even including these measures as covariates revealed a similar effect of CR proxy variables in moderating the effect of age on performance, but not for executive abilities or for arithmetic scores.

This last study also incorporated MRI measures including the cortical thickness of various regions that had previously (Steffener et al., 2014) been demonstrated to a) mediate between age and cognitive performance on a set of fluid reasoning tasks, and b) to be successfully subject to a moderation effect by CR as measured by verbal IQ and education. Using a similar moderated-mediation model, we demonstrated that in at least one of these regions–the right rostral anterior cingulate cortex–an effect of age-related differences in cortical thickness on fluid reasoning was also moderated by the strategy selection measure. Significantly, this suggests that differential strategy use may not just be associated with CR, but may behave like it in a mechanistic fashion, to at least a partial degree.

One of the major limitations of our previous research into cognitive strategies is the relatively narrow focus of both the domain (computational estimation), and the limited variability between the strategies themselves; rounding-up and rounding-down, while clearly distinct, do not engage fundamentally different sets of cognitive operations to accomplish the same end. Thus we instead sought a task with high levels of generalizability to other tasks, and centrality to a wider set of cognitive domains, that could be decomposed into a set of qualitatively distinct strategic processes.

Project Motivation

This project investigates the possibility that efficient cognitive strategy switching is one possible cognitive mechanism of cognitive reserve by using a novel adaptation of a classic fluid reasoning test set within the context of a task-switching paradigm. Ultimately this project aims to compare strategic flexibility with other proposed mechanisms of cognitive reserve, which could point the way to more effective cognitive interventions and remediation programs.

Although the cognitive mechanisms of cognitive reserve (CR) have received a disproportionately low share of attention as compared to its neurofunctional implementation, there has been some recent research into such mechanisms. Some work has hinted that the effective utilization of distinct cognitive strategies may play a role imbuing reserve (Barulli et al., 2013). While some researchers speculate that fundamental cognitive capabilities such as attention or working memory may mediate between CR and overall cognitive performance, another line of research suggests instead that a more dynamic process involving the acquisition, selection, and execution of cognitive strategies may represent the dominant mechanism through which CR moderates the effects of age on cognitive function.

Cognitive strategies represent differential cognitive steps towards solving some problem or performing some cognitive task. Two distinct strategies can rely on the same fundamental set of cognitive abilities permuted differently (e.g. navigating a spatial grid by moving laterally in one direction to the desired x position, then moving in another direction to the desired y position), can draw on largely the same set of constituent cognitive operations with one or more additional operations included (e.g. navigating a spatial grid by moving alternately in the x and y directions, with iterative monitoring that the desired position is becoming closer), or can utilize entirely distinct sets of cognitive operations (e.g. navigation using landmarks rather than allocentric turns). For this reason, alternative strategy use can make performance on the same exact task simpler or more difficult depending on the suitability of the strategy employed, and can cause the same task to rely on different neural substrates, complicating interpretations of individual differences.

Using a novel paradigm culled from the neuropsychology of fluid intelligence, we here study the potential of healthy younger and healthy older participants to flexibly adapt their cognitive strategy use to a particular problem, and investigate the relationship between this capacity and their performance on various other cognitive tests. For each study, two groups of young (20-40) and old (60-80) subjects were enrolled from either an online "workforce" population (Amazon's Mechanical Turk workforce) or from within a larger study conducted at Columbia University Medical Center, the Reference Ability Neural Network Study (Stern et al., 2014). This latter study is the source for all neuropsychological performance values with the exception of the main dependent variables (the strategic flexibility measures), as well as select MRI-derived cortical thickness and brain volume measures. Using these data, we intend to explicate the relationship between strategy flexibility and other neuropsychological constructs, as well as its relationship with neuroanatomical biomarkers.

Using these subjects as well as their existing data, we aim to derive a measure of strategic flexibility from a modified version of a single and well-understood fluid reasoning task, using tools adapted from the domain of task-switching. Our working definition of "cognitive strategy" is: a meta-cognitive permutation of cognitive processes selected from memory (either implicitly or explicitly), with the intention of achieving a desired goal in the most optimal manner possible. "Strategic flexibility" in turn refers to the capacity of an individual to alternate between such permutations across time and over rapidly changing task demands or goals. While we assume cognitive strategies are themselves the products of an individual's own experiences with a task domain or exposure to explicit knowledge/training about alternative approaches, they are not necessarily bound to any particular cognitive process. One strategy may encompass various cognitive processes (e.g. spatial navigation and arithmetical comparison) so long as it is permuted with an intention of achieving a particular task. For this reason, investigating strategies directly can be methodologically challenging, since every instantiation may be unique across a task while simultaneously involving various disparate and neuroanatomically distinct cognitive processes. Localizing the neural substrates of different cognitive strategies in future research may well be possible with sufficient spatial and temporal imaging resolution, but for the purposes of this project we chose to instead analyze the more global and stochastic property of strategic flexibility.

Just as different individuals have differing levels of measured intelligence, executive ability, and task-switching capacity, so we assume they have differing levels of flexibility when it comes to selecting appropriate cognitive strategies. Whereas some may have ample experience with a problem domain and reliably produce an optimal strategy but falter when given a new and slightly unfamiliar problem, others may have little experience at all with any domain tested yet be capable of maintaining an adequate level of strategy use across the two domains. Hence the ability to apply optimal strategies in some circumstances does not necessarily entail a high level of strategic flexibility. Expert chess players would be expected to generate highly optimized strategies for most games of chess, but they would not be expected to necessarily do the same during interspersed games of Monopoly. Those with more general experience playing boardgames however, while not always generating optimized chess strategies, could be expected to perform with greater ”flexibility”--formulating overall decent strategies in both games. Thus generality of strategy exposure is likely to play an important role in an individual’s strategic flexibility.

Owing to the conceptual importance of this general exposure, CR is expected to share critical similarities with strategic flexibility. Both presumably originate with greater levels of intellectual engagement, both are definitionally characterized by compensatory ability in the wake of impairment (be it via pathology or unfamiliarity with a task domain), and both (we speculate) may modulate the effects of brain states on measured performance. Intuitively, CR can be thought of as differential lifestyles which can bring about more or less familiarity with a wide range of cognitive strategies, which in turn is reflected in differential levels of strategic flexibility.

We hypothesize that higher levels of CR will be associated with greater levels of such flexibility as represented by fewer switch costs when utilizing one strategy type following another. These switch costs can come in the form of greater reaction times and diminished accuracy following a strategy switch. Further, we expect that this relationship between CR and strategy switch costs will hold even while controlling for fluid intelligence as measured using other reasoning tasks.

We also present evidence of the moderating role that strategic flexibility plays in the relationship between age-related differences in cortical thicknesses and cognitive performance within several cognitive domains, including reasoning, speed of processing, and memory. We hypothesize that strategic flexibility moderates the effect on cognitive performance of age-related differences in cortical thickness of several brain regions previously identified as being related to CR.

The major significance of this series of studies lies in its extension of the preexisting strains of research on cognitive strategies and their potential role as a cognitive mechanism of CR. By methodologically narrowing in on the measures of strategy use (by embedding them within a rigid task-switch paradigm) while broadening the cognitive range of available strategies, this project aims to demonstrate that strategic flexibility is a stable source of interindividual difference with close ties to CR. Showing that even the purest tests of fluid intelligence are amenable to different strategies, and understanding how the choice of these different strategies can be influenced negatively by age but may be preserved by lifetime exposures like education, cognitive psychologists could suggest a powerful mechanism through which environmental exposures can operate seemingly independently of mechanisms like neural reserve or neural compensation. This would not only reopen the door to much basic behavioral research in the field of cognitive aging, but it would also suggest an area ripe for cognitive intervention in the form of cognitive strategy training.

Project Outline

In the current set of studies we use a modified version of the Raven's Progressive Matrices (RPM) test (Raven, 2003). Much prior work has been done investigating the structure of the RPM, and many researchers have suggested that there are various dimensions to this task which may be amenable to distinct cognitive strategies. Kirby and Lawson (1983) for instance suggested that 2 broad strategies could be employed, depending on the particular trial: so-called Gestalt strategies, utilizing visuospatial operations, and so-called analytic strategies, utilizing logical and sequential operations.

Carpenter, Just, and Shell (1990) later suggested that the vast majority of RPM problems could be solved using just five abstract rules, some of which may be categorized as primarily analytical in nature and some as primarily visuospatial. Based partially on this and partially on the work of Hunt (1974), DeShon, Chan and Weissbein (1995) suggested that all RPM problems were either of the visuospatial reasoning type, or of the analytical reasoning type. Furthermore, they identified two distinct sets of rules corresponding to the two strategies (e.g. object addition/subtraction for visuospatial problems, or quantitative pairwise progression for analytical problems). Based on the pattern of performance observed in test-takers who apply appropriate versus inappropriate rules to specific items, the two general types of matrix can be said to encourage the use of different strategies; i.e. applying one set of rules to a problem leads to performance decrements relative to applying another set of rules. Furthermore, such performance dissociations have been linked to individual differences, e.g. with males in one study outperforming females on RPM items requiring the application of analytic rules but no difference in visuospatial RPM items (Mackintosh & Bennett, 2005).

We intend to utilize this property of RPM and RPM-like problems by introducing an 80-item computerized Strategic Raven's Task (STRATA). The task features 4 blocks of 20 trials, with 2 blocks dedicated exclusively to a single strategy, visuospatial strategy items or logico-analytic strategy items, respectively; and 2 blocks having equal numbers of each strategy type items. Problem order for these blocks was randomly determined but fixed for all participants. Mirroring closely the procedure used in the computational estimation studies mentioned above, participants were first given instruction in the two distinct strategies (visuospatial and logico-analytic) by exposing them to trials utilizing each type of rule in a forced-strategy. Next they were given the STRATA and instructed to choose only one strategy for completing the problem. Solution times and accuracy of item completion were recorded.

Stimuli for this task come from software for automatically generating 3-by-3 RPM-like problems (each with 8 possible solutions) based on various well-understood and quantifiable task parameters (Matzen et al., 2010). The particular stimuli adopted here were extensively piloted and equalized for difficulty within blocks: all trials across all blocks average .75 accuracy in piloting (either during an initial norming study as reported below, or during piloting during task development at Columbia.)

All logico-analytic strategy trials require the use of one of three basic rules: OR, wherein subjects must add the figures together across rows or down columns to generate the missing figure; AND, wherein participants must generate the missing figure using only those elements that are present in both other cells in the row or column; and XOR, where subjects must take only those elements that are unique to one cell and the other and add them to the missing cell.

All visuospatial strategy problems require the use of three more basic rules that attend to two simultaneous element features: size and number, wherein subjects complete the missing cell based on the size and number of elements that should be expected in the missing cell (e.g. the two complete rows may have 1, 2, and 3 elements of sizes small, medium, and large, but the incomplete row has only 1 and 2 elements of size small and medium, suggesting the missing cell should contain 3 large elements); orientation and size of elements; and shape and number of elements.

In the first study, we administered an initial version of the STRATA to investigate its psychometric properties. Two groups of MTurk workers (young and old) were administered the task. We hypothesized that, consistent with previous research on task-switching, younger adults would outperform older adults both in reaction time (RT) and in accuracy/percent errors made (PE). Even more critical, this study intended to validate the task, and the notion of "strategy-shifting"; i.e., we hypothesized that subjects would overall perform worse (higher RT and PE) on trials wherein they needed to employ the alternative strategy to the one used in the immediately antecedent trial. This matches to the metric of local switch-costs in the task-switching literature. Furthermore we also expected that subjects would display so-called mixing costs in this task; i.e. subjects would display higher RTs and PEs in non-switch (using the same strategy as the one they had used on the prior item) trials within a mixed-strategy block when compared to trials during a pure-strategy block.

In the second study, we revised the STRATA to reflect some of the results of the initial study. Here we imposed an "alternating-runs" paradigm, culled directly from the task-switching literature, onto the task to make the expectancy of each strategy greater for participants. We also revised the set of items to make them all roughly equal in difficulty (with around .75 accuracy in piloting results). To control for the impact of more classical task-switching effects rather than the intended strategy-shifting effects, we included within the pure-strategy blocks a matching alternating-runs pattern of rule-switches requiring subjects to employ a different set of cognitive operations while still relying on the same fundamental cognitive domain. Finally, we randomized the order of block presentation to control for order effects in which strategies were first utilized. Additionally, we administered a cognitive questionnaire intending to probe the participants for their overall mental status as well as various exposure variables such as education and occupational history, providing us with a proxy for CR. Our hypotheses for this study were 1) strategy-switching costs would be induced over and above task(rule)-switch costs, 2) older adults would experience greater strategy-switching costs than younger adults, and 3) our CR proxy would moderate the effect of age on performance such that older adults with higher CR scores would perform more akin to younger adults than those with low CR scores.

In the third and final study, we administer the STRATA to a group of participants who had previously been studied within the context of a broader neuroimaging study at Columbia University Medical Center. Again two groups of young and old subjects participated, but in addition to their performance scores we also utilized a rich existing set of data from these subjects, including a lengthy neuropsychological battery, a more detailed composite measure of their estimated CR, and their values for various structural MRI measures such as cortical thickness and regional brain volumes. Here we expected to replicate the prior behavioral effects observed in studies 1 and 2, but also explore the relationship between their STRATA performance and their neuropsychological scores as well as MRI measures. Specifically, we expected that STRATA performance could serve as a substitute for CR in a series of moderated-mediation models where previously CR had differentially altered the effect of age-related cortical thickness (in a set of specific regions) on fluid reasoning performance.

CHAPTER 2
LITERATURE REVIEW

Cognitive Reserve

Cognitive Reserve (CR) is latent variable which is meant to account for the frequently-observed discrepancy between an individual's actual level of cognitive functioning, and their expected level given their neurological status (Barulli & Stern, 2013). Such discrepancies have been observed and related to this concept in cases of Alzheimer's, traumatic brain injury, multiple sclerosis, HIV-related dementia, normal aging, and many others (Stern, 2002).

Numerous individual lifetime exposure variables have been found to be protective against the effects of each of these neurological insults on cognitive status, and because of their close features they are often grouped together to form a composite CR variable. Such exposure variables include years of formal education, literacy levels, complex occupational status, engaging leisure activities, and higher socioeconomic status. The common feature shared by such variables is that they either directly or indirectly lead to greater levels of intellectual engagement across time.

The exact mechanisms through which sustained intellectual engagement leads to preserved cognitive functioning among neurologically at-risk individuals are widely speculated, but so far inconclusive. However, several key links have been drawn between CR and patterns of neurofunctional activation measured using a variety of neuroimaging modalities. Famously, CR is often predictive of the use of compensatory network activations among an at-risk population (e.g., elders), meaning patterns of task engagement not observed in younger participants. In many cases, such patterns are associated with diminished performance on the task at hand, but this is not always the case (Steffener & Stern, 2009).

Cognitive Strategies

Alternative patterns of fMRI-related network activations may not necessarily reflect compensatory activity. Sanfratello et al (2014) for instance point out that the common practice of deriving group-wide activation patterns in neuroimaging studies often do not reflect the individual patterns observed among the subjects who comprise the sample of these studies, often as a function of confounding variables such as strategy use (c.f. Aine et al, 2011). Such a limitation of neuroimaging studies is especially problematic when groups are compared in terms of their mean activation differences, such as when younger and older subjects are determined to rely on different neural substrates to perform the same task; one especially problematic confound is the increased variability among older samples, which these authors speculate could derive in part from cardiovascular risk factors not captured by typical exclusion criteria (e.g. high blood pressure or type II diabetes).

Another probable source of variation is alternative strategy use among the elders, whose brains may have been afflicted by various sources of age-related neurological deterioration, as well as altered by experience-dependent changes reflective of shifts in cognitive strategy or style. Sanfratello et al. (2014) reason that if differences in neural activation patterns can reliably be seen among the more homogenous group of younger participants, they would almost certainly generalize to a greater variety of neuroimaging samples including older adults. To that end they used magnetic encephalopathy to study functional activation among healthy young participants engaged in a spatial working memory task, relying on self-reported strategy use after the MEG session. Using cluster analysis, subgroups of the sample were constructed reflecting different predominant strategy use but without any a priori assumptions about the number of strategy clusters that would be found. Results indicated that two predominant strategy clusters emerged, one broadly reflecting a verbal strategy group (indicated by phrases such as "I thought of a word or a phrase"), and one broadly reflecting a visuospatial strategy group (indicated by phrases such as "I did keep the digit locations in mind the way they were presented"); participants utilizing a verbal strategy performed better on the working memory task, as well as the co-administered California Verbal Learning Test.

For their MEG results, participants utilizing the verbal strategy showed greater activation in the right medial temporal lobe, and slightly more than half of this group also showed activity in the left MTL. In the visuospatial strategy group, left MTL activity was also observed but no bilateral activation could be seen, and the right occipital cortex showed greater activation, and this group also had many more significant correlations between performance on neuropsychological tests and white matter tracts as measured fractional anisotropy using DTI. Of particular significance were the posterior commissure tract, which correlated with working memory performance, as well as the uncinate fasciculus (UNC); while the former has been associated with cerebellar connections to various regions of cortex as well as with working memory performance, the latter connects portions of the limbic system with the frontal cortex and was here found to correlate in its integrity with performance on the Rey Complex Figure Task (ibid). Like Kherif (2009), who found that a subset of participants engage in a particular reading strategy that activates the posterior cingulate and precuneus, but that these activation patterns are often washed out of group averaging, Sanfretello et al. argue that individual differences in strategic approaches must be taken into account for accurate generalization of neuroimaging findings. These results also belie the common practice in aging literature to label any atypical neurofunctional activation patterns as "compensatory"; instead, they may be a function of strategy use as observed here within a young sample.

Age-Related Strategy Differences

Lemaire (2010) reviewed some evidence of the existence of age-related cognitive strategy changes. He notes that while the majority of research on cognitive aging focuses on so-called quantitative factors that may affect cognitive decline across the lifespan, such as processing speed or working memory capacity, some researchers emphasize instead the equal importance of more qualitative factors such as cognitive style or cognitive strategy use. Dunlosky & Herzog (2001) for instance showed that young and old adults differ in their approach to solving Paired Associates-style problems. The two groups used a total combination of four distinct strategies, such as using mental imagery to link the words presented, creating a sentence that includes both words, or simply relying on rote repetition. While both groups had access to the same total range of strategies, the relative distribution of these strategies differed across groups; whereas older adults relied more heavily on sentence generation, younger adults relied more heavily on rote repetition (while both groups relied equally often on mental imagery).

Other researchers have supported the finding that older adults spontaneously access a fewer number and diversity of strategies. Duverne and Lemaire (2004) showed that in a study of arithmetical inequality verification, younger adults relied broadly on two classes of strategies–an exhaustive and approximate verification strategy–whereas older adults relied predominantly only on the exhaustive strategy. Such findings have implications for more ecologically valid reasoning tasks, such as the case of consumer decision-making; Johnson (1990) showed that when searching information about potential cars to purchase, young adults more frequently use a compensatory strategy (wherein less salient attributes can overwhelm the valence of the most salient attribute, for instance the comfort and color of a car leading to a decision to purchase despite its price) while older adults rely on a on the mentally non-compensatory strategy (focusing exclusively on the most salient attribute without consideration of other potential features).

The presence and magnitude of age effects depends strongly on the strategy and task utilized, with some strategies leading to equal performance in young and old groups. In a magnitude estimation task requiring participants to estimate the number of dots presented on the screen, Gandini et al. (2008) showed that an anchoring strategy (wherein a participant counts a group of dots and adds the number of groups on screen, plus some estimation of remaining ungrouped dots) led to increases in reaction time for old compared to young subjects, but use of a perceptual strategy (wherein participants rely on their memory of prior numerosities, matching the currently presented array of dots to some remembered quantity) did not lead to any differences in RT. This particular study also found different underlying neural networks for each strategy. Strategy adaptivity, or the adjustment of strategy used in order to best fit task relevant parameters and increase performance, has been shown to be present in task domains ranging from arithmetic, to reasoning, to serial recall (Lemaire, 2010).

In an associative recognition test, Patterson and Herzog (2010) show that young adults utilizing an interactive imagery strategy (wherein they form a mental picture of the two words in each pair presented) perform better than older adults using this strategy, but when both groups instead executed an individual-plus-interactive-imagery strategy (wherein they first form a mental image of each word, then combine these images into pairs) performance was held constant. And Faulkner (1983) used a sentence-picture verification task to show that the linguistic strategy (relying on the verbal description of the picture) yielded age differences in performance while a pictorial strategy did not.

Thus strategy use, like neurofunctional activation patterns, must be interpreted with caution; it is not always the case that application of an alternative strategy is compensatory in nature. People have been shown to vary their strategy use in response to the unique characteristics of the task at hand, the instructions of the given task, or even their own perceived competence (Schunn & Reder, 2001). To more systematically interpret how strategy use affects performance, Lemaire (2010) gives a useful categorization scheme for the factors of cognitive strategy use that may influence the underlying cognitive operations: 1. Strategy repertoire, or the quantity of strategies known to the individual; 2. Strategy distribution, or the frequency at which each particular strategy is used; 3. Strategy execution, or the speed and accuracy at which a particular strategy is employed; and 4. Strategy selection, or the cognitive processes affecting the choice of a particular strategy for a particular cognitive demand.

Age effects have been found along the dimensions of both strategy repertoire and strategy selection in basic arithmetical tasks, such as complex addition problems. The prior research above has consistently shown that in comparison to young adults, healthy older adults both employ a lower variety of strategies (meaning a smaller strategy repertoire), and that they select the optimal strategy for a particular item less often than do young adults (meaning impaired strategy selection) (Lemaire, 2010).

Along the dimension of strategy distribution, Cohen and Faulkner (1983) demonstrated that whereas younger adults use one particular distribution of four strategies in a mental rotation task, older adults use an entirely different distribution on average. Hartley and Anderson (1983) had a similar finding in regards to inductive reasoning strategies. Studies such as these typically rely on verbal self-reports of the strategy employed, but in some cases other methods of ascertaining strategy use such as eye tracking were also employed to the same effect (e.g. Gandini et al., 2008).

Lemaire (2010) suggested the findings about differences in strategy repertoire could be explained either by older adults simply knowing fewer strategies than younger adults (perhaps as a function of cohort effects deriving from changes to education policy), or by an age-related restriction on this set of strategies older adults are comfortable using as a process of some higher-order cognitive factors such as executive functioning. The fact that on a group level young and old participants use the same total array of strategies in many tasks suggest that the latter explanation is more likely.

Along the dimension of strategy execution, Siegler and Lemaire (1997) designed a paradigm of choice/no choice wherein one condition forces participants to use a particular strategy across all items while another condition allows participants to choose the strategy they want to use; such a paradigm has been shown to be effective for controlling for other dimensions of strategy use like repertoire, distribution, and selection. This method has been used to show that older adults are poorer in strategy selection in domains ranging from computational estimation to tasks requiring participants to select the optimal virtual pond at which to fish to garner the greatest number of catches.

Along the strategy selection dimension, computational estimation tasks, like the one described in the introduction, have frequently been employed that can allow simple classification of the correct versus incorrect strategy (Lemaire, Arnaud, & Leclere, 2004), also finding age-related performance declines. This effect has been numerously replicated and has been shown to be exacerbated in Alzheimer's patients (Lemaire & Leclere, 2014). However, the effect of age on the various factors of strategic performance has only been very partially investigated, and as noted some work has shown that protective factors like CR may modulate this relationship (Barulli et al., 2013).

Cognitive Strategies and Executive Functions

In order to better understand the interaction with age, it's necessary to consider the relationship between more proximate age-related cognitive components and strategy use. Mediation analyses have been used to show how much variation in strategy performance between age groups can be accounted for by other abilities, with executive functions accounting for 82% of age-related variance in strategic approaches to an episodic memory task (Bouazzaoui et al., 2010), processing speed accounting for 70% of age-related variance in arithmetical strategy use (Duverne & Lemaire, 2004), and more recently executive function accounting for large portions of variance specifically in strategy selection and repertoire (Hodzik & Lemaire, 2011). Such effects might be deriving from the impairment to strategy execution, or from a compensatory pattern of strategy selection (or any number of other potential combinations). Bouazzaoui (2010) for example has suggested that older adults rely more heavily on strategies that are less reliant on internal memory (e.g., the ample use of taking notes instead of relying on memory), which would make the importance of capacities like working memory particularly stand out.

Using a taxonomy of executive functions derived from Miyake (2000), dissociating inhibitory function (i.e. preventing the execution of pre-potent responses), cognitive flexibility (i.e. shifting between multiple tasks sets), and updating (i.e. monitoring and encoding new information), executive function has been ontologically parsed into its constituent components. Much prior research in the cognitive aging literature exists to support the notion that with increasing age, decrements in all three of these executive functions can be observed. In a pair of experiments designed to test the mediational role of age-related changes to executive function on strategy repertoire and strategy selection, Hodzik and Lemaire (2011) tested a sample using measures of shifting (using the Trail Making Test) and inhibition (using the Stroop test). Prior research by Lemaire (2010) had shown that older adults tend to repeat the same strategy across multiple problems, even when problem characteristics change and thus would be better suited to a new strategy. They reason that this observation would be explained by changes in cognitive flexibility and inhibition across age, limiting the ability of older adults to flexibly adopt a new and more optimal strategy. Using a test consisting of two-digit addition problems, along with self-reported strategy use immediately following each trial, these authors showed that in a sample of 40 young and 38 older adults the latter significantly use fewer strategies than young adults. While verbal reports indicated that the total range of strategies (nine reported in total) was represented by both age groups, younger adults use an average of three of these strategies whereas older adults use an average of 2.1, and that this difference was exacerbated with increasing item difficulty. However, using a composite of the executive scores within a mediation analysis reduced the age-related variance in performance by 63%; controlling for executive function, the age effect on performance was no longer significant. Furthermore, younger adults also solved problems faster and more accurately than the older adults as expected, and had better executive function scores.

In another experiment designed to investigate the relationship between executive functioning and strategy selection, these researchers expected to find age-related decreases in performance that would again be mediated by executive functions. They used a computational estimation task in which participants are asked to solve two digit multiplication problems (e.g. 57 x 49) by employing only one of two unique strategies– either rounding up (60 times 50) or rounding down (50 times 40); such problems have the advantage of having a mathematically optimal solution (i.e. in the above example, rounding up would yield an estimate of 3000 which is only 207 units away from the actual answer, in comparison to the rounding down strategy which would yield an estimate 793 units off). In a sample of 40 young and 40 old participants, age affected performance as measured using the mean percent use of the best strategy (with younger subjects using the best strategy 87.7% of the time, and older subjects using the best strategy only 77.3% of the time). Again testing for the mediation effect of an executive function component, age-related variance in performance was found to be reduced by 39%; while the main effect of age on performance still remained after controlling for executive functions, executive function did explain a relatively large proportion of the performance decrements.

Such prior literature on the relationship between age, executive functions, and strategy use suggests that the presence of mediators is a serious possibility, and just as strategy use must be taken into account when considering neurofunctional results, so must basic cognitive abilities be considered within the context of strategy selection and execution.

Fluid Intelligence: Raven's Matrices

While the prior literature on cognitive strategies has begun to paint an overall consistent picture of age-related decline, many of the "strategies" studied lack obvious ecological validity; while rounding down versus rounding up undoubtedly engage different underlying cognitive operations, their existence within the same narrow domain of arithmetical computation belies the argument that differential strategy use may be immensely important in accounting for all sources of age-related cognitive variance. To investigate this claim more fairly, a far more domain-general type of task amenable to strategic manipulation is required, and fluid intelligence is one potential such domain.

The Raven's Progressive Matrices task (RPM) is widely considered to be a gold standard for measuring fluid intelligence because of its high correlations with other measures of cognitive processing, and high correlations with measures of intellectual achievement (Korte and Raven, 1982). The pattern of individual differences observed on the Raven's task also correlate highly with such differences observed in other complex cognitive tests (Jensen, 1987), and even correlates with biomarkers such as nerve conduction velocity (Reed and Jensen 1992). Each item of the RPM consists of a 3 x 3 matrix wherein the bottom right entry is missing and must be selected from a set of eight potential responses.

Despite its wide use and psychometric centrality, the question of whether or not the task is best represented by a single underlying cognitive component or multiple such components was a matter of extensive debate. Early exploratory factor analyses (Dylan, 1981) suggested that RPM performance is best modeled as a function of two distinct factors, an “addition/subtraction” factor and a “detection of pattern progression” factor. However, a later confirmatory factor analysis (Arthur and Whor, 1993) suggested that a two-factor model did not sufficiently improve model fit over a one-factor model to warrant the conclusion that two distinct factors are represented by the task.

Hunt (1974) suggested that two general problem-solving algorithms could be applied to the majority of APM problems: first, a visual strategy (relying on perceptual reasoning, continuing perceptual patterns such as lines, and superimposition of elements onto one another; and second an analytic strategy that applies logical operations to features of the problem. While these two strategies can be simultaneously employed for a large number of RPM problems, such simultaneous application has been shown to lead to the verbal overshadowing effect, wherein concurrent verbal processing can impair spatial processing of visuospatial stimuli (Schooler & Englster-Schooler, 1990). These authors argued that when a propositional representation is available to solve a problem, it is much more likely to be used; furthermore such propositional processing has led to decreased performance on other visual-spatial tasks such as face recognition tasks (1990), mental rotation tasks (Brandimonte, Hitch, & Bishop, 1992), as well as visual insight problems

(Brandimonte & Gerbino, 1993), but does not lead to decreases in performance on primarily propositional or verbal problems.

Carpenter, Just, and Shell (1990) further sought to provide a detailed taxonomy of the types of rules required by the RPM. Using behavioral data of college students who took the test, and relying on such metrics as verbal reports, eye fixations, and error patterns in the different problem types, these authors designed two computer simulations to model performance on the RPM. One of the simulations is designed to model the performance of a median college student in their sample (the FAIRRAVEN model); and another is designed to model the performance of one of the top performers in their sample (the BETTERRAVEN model). These models differ in two respects: 1. The better model can form more abstract relational representations, and 2. the better model has a greater working memory capacity.

These authors' investigation into the structure of the RPM problems yielded a taxonomy of five different rule types, with each item differing in the necessary combination of these rules. The first rule is the constant in a row rule, wherein the same elements will occur throughout a row but change down a column. The second rule is the quantitative pairwise progression rule, wherein adjacent cells differ according to some quantitative increment, be it in size, position, or number. The third rule is the figure addition or subtraction, wherein a figure from one column is either superimposed onto the figure from another column, or subtracted from it. The fourth rule is the distribution-of-three-values rule, wherein three distinct values of a particular category (e.g. figure type) are distributed throughout a row. The fifth rule is the distribution-of-two-values rule, wherein two values from a particular category are distributed throughout a row while the third value is null.

These rules are usually equally applicable if applied row-wise as if they were applied column-wise, but behavioral analysis suggests that most test-takers rely on a row-wise strategy of rule application. Determining which figural elements correspond to which rule is a source of cognitive complexity in the RPM; it is termed by these authors “correspondence finding”. Two common heuristics are employed to complete this process of correspondence finding: first is the matching-names heuristic, wherein participants assume that figures having the same name (such as line) correspond to each other and are governed by the same rule set; the second heuristic is the matching-leftovers heuristic, wherein participants assume that if all but one of the features in adjacent cells have been grouped then the remaining features correspond to the same rule set. This, along with the pure number of rules to be applied, are the two main theoretical contributions to the difficulty of each problem. In two series of experiments on a small sample of undergraduate students performing a sub-selection of the RPM items, the authors used verbal reported strategies as well as eye tracking to monitor the application of rules to each item.

Behaviorally, error rates are highly correlated with reaction times at .87, suggesting that these problems are not subject to a speed-accuracy trade-off but that participants will take additional time for more difficult problems. The strongest task-specific feature predictive of error rates was the number of rules that had to be applied to a problem, suggesting a strong influence of goal management and working memory on performance. The main finding from this experiment was that problem-solving was fundamentally incremental in nature, with participants decomposing the problem into smaller sub-problems, including processing one rule at a time, and within each individual rule processing each adjacent pair of cells in an incremental fashion before moving on to considering additional rules. This pattern of processing was universal among the small sample, and was not a source of individual differences. This general pattern was borne out both by the verbal reports, and by the eye tracking data. To test their hypothesis that goal management and working memory capacity are critical to performance on this task, the authors conducted an additional study in which they tested young subjects on both the RPM problems, as well as the Tower of Hanoi, a task heavily reliant on breaking goals into subgoals and managing the subgoals and memory. The correlation between the error rates on these two tasks was very high (.77).

Using models to simulate a test-taker of the RPM, the authors included a set of algorithms roughly grouped into three categories: perceptual analysis, concept analysis, and response generation and selection. The fourth component that differed between these two simulations was working memory. Perceptual analysis functions to encode information about the figures, conduct the correspondence finding between rules and figural elements, and compare figures in adjacent cells, resulting in a pattern of pairwise similarities. Conceptual analysis by contrast attempts to determine the rule type to be applied to each problem, and this operates directly on the set of pairwise similarities computed earlier. The response generation and selection mechanism then applies the set of computed rules to generate the missing cell, and searches the offered answer set for a matching entry. This method of generating a potential response before searching solution alternatives is in line with the behavioral performance of higher scoring subjects, which differs from performance among lower scoring subjects who often rely on a response elimination strategy (searching alternatives to eliminate based on incremental application of the rules, rather than directly generating a potential solution and then matching it to the answer choices). This FAIRRAVEN model's performance closely matched that observed in the authors' sample of undergraduates, with a point biserial correlation between the error rate for each problem and the dichotomous score (solved or unsolved) from the model's performance being .67.

The BETTERRAVEN model made several changes to the FAIRRAVEN model which improved performance, allowing it to match top human performers. First and foremost, a goal monitor was added as a separate module that allows for recursive sub-levels of the goals to be flexibly taken into account, and backtracked when either a particular subgoal is reached or a particular rule is found to be invalid. On the perceptual level, more abstract algorithms for correspondence finding were included, and on the conceptual level rule types were prioritized such that they could be tested in a serial fashion. These changes resulted in this model being able to solve all but two of the problems. By simulating "lesions", or limits to specific portions of the new model, the authors were further able to demonstrate that performance was most heavily reliant both on the goal monitor, and on the inclusion of new abstraction rules (such as the distribution of two rule); performance was intermediate when these were degraded in some piecemeal fashion.

Following Hunt (1974) and Carpenter et al. (1990), DeShon, Chan, and Weisbein (1995) decomposed all previously identified rules into sets of distinct visuospatial rules and verbal-analytic rules, and attempted to induce an experimental dissociation to these two problem types. Visual-spatial rules consisted of superimposition of item elements, superimposition with cancellation, object addition/subtraction, movement, rotation, and mental transformation. Verbal-analytic rules and their taxonomy refers to constant-in-a-row, quantitative pairwise progression, distribution-of-three-values, and distribution-of-two-values. In a large sample of undergraduate participants, they tested for the impact of concurrent verbalization on item performance, after systematically labeling each APM problem as predominantly visuospatial or predominantly verbal analytic; as expected concurrent verbalization decreased performance in the primarily visual-spatial problems, but not in the verbal analytic problems.

While verbal strategies may be a source of error for some subjects in some conditions, there is evidence that they can also bolster performance. Some research has suggested that the visual appearance of items in the RPM can affect their difficulty, with so-called contextual matrices designed by Richardson (1991) improving children's performance on the reasoning problems when abstract and unidentifiable shapes are replaced with socially meaningful contexts. Thus instead of elements that were unfamiliar and difficult to name, common items such as furniture or toys were used instead. To test the impact of element salience further, Meo et al. (2007) devised a paradigm wherein either European letters or invented letters were used as elements, appearing in non-overlapping or overlapping conditions. The authors found that as expected the greatest decrements in performance occurred when item elements are both unfamiliar and overlapping, and when elements are either unfamiliar or overlapping there are moderate decrements to performance as compared to when elements are nonoverlapping and familiar. These authors suggest that this effect is driven by the potential strategy of applying verbal tags to item elements, which reduces working memory load; such verbal tags are possible with familiar and non-overlapping items, but not necessarily in other conditions.

Verguts & De Boeck (2002) attempted to determine if continual use of the same rules during RPM performance biases individuals towards them in subsequent items. They suggest further that RPM performance might be subject to the sequence effect (Sweller & Gee, 1978) wherein if problems are ordered from easy to hard, they are much easier to solve than in the reverse ordering. Using a talk-aloud methodology, which according to Erikson and Simon (1984) is useful to determining how people solve complex tasks so long as it is employed during the problem-solving process and not after, they speculated that induction of the appropriate rule occurs during RPM performance early with easy items, and that these rules are then biased towards repetition in later, harder items. The authors suggest a formal model of rule induction wherein the prior activation of a rule biases it's later application to the problem, weighted by some individual learning parameter. They furthermore include a parameter for a lag effect (a preference for items that were shown recently versus items that were shown before). In one of their experiments utilizing newly created RPM-type problems and providing explicit feedback on performance, learning effects are much stronger than in the other experiment, wherein the actual RPM problems were presented and no explicit feedback was provided. This suggests that explicit information about rules can greatly increase the learning rate in problem-solving tasks such as the RPM. Explicit instruction therefore can improve overall task performance, as mediated by appropriate strategy selection and/or execution.

Other evidence exists that prior knowledge may affect strategic approaches to the RPM. Schulze, Beauducel, and Brocke (2005) speculated that figural reasoning tasks containing semantically meaningful objects would have higher crystallized intelligence (Gc) loadings than figural reasoning task containing only abstract objects. Despite its high G loadings, the RPM has been criticized under the supposition that not all test-takers will be equally familiar with the figural elements included within it, a fundamental assumption of fluid intelligence testing. They note that fluid intelligence (Gf) is often measured using figural reasoning tasks while Gc is often measured using verbal tasks, but both of these tasks could have a confounding influence of prior exposure relevant to their measured constructs; in response they recommend an analysis model wherein verbal, numerical, and figural abilities are separately estimated for Gc and Gf, and controlled in analyses. The researchers designed non-abstract figural reasoning scales and hypothesized that these would load onto Gc. Using confirmatory factor analysis, these authors demonstrated that figural reasoning tasks designed with abstract elements loaded differently on Gf and Gc factors than those designed with concrete elements. These results show that while the RPM may be a particularly strong measure of Gf, the impact of Gc on an individual's performance must be carefully considered.

In support of Carpenter's (1990) finding that in general participants employ serial rule induction on the RPM, often considering more basic rules before moving on to more complex rules. Primi (2001) decomposed problem-solving behavior into three basic stages: first, creating a mental representation of the problem and problem rules (sometimes called encoding and inference, perceptual and conceptual analysis, pattern comparison/decomposition, or transformational analysis and rule generation); second, recognition of similarities between the problem rules in a novel analogous situation (sometimes called mapping, perceptual and generalized conceptual analysis, or rule comparison); and third, applying rules to generate an appropriate solution representation and or select an appropriate presented answer (sometimes called application, comparison-response, or response generation and selection). He notes that research utilizing structural equation models has previously linked Gf to the central executive components or the controlled attention component of working memory. He also attempted to disambiguate the sources of complexity factors in RPM problems, and theorized that this factors may derive from 1. The number of elements in a problem, 2. The number of transformations or rules required for that problem, 3. The types of rules, and 4. Perceptual organization. Salthouse (1994) divided working memory into three separate components: 1. Storage capacity, 2. Processing efficiency, and 3. Coordination effectiveness. While the latter two components have been shown to be highly correlated with fluid reasoning tasks, the former is less critical. Bethel-Fox (1984) investigated analogical reasoning problems and suggested that two distinct strategies could be applied: constructive matching wherein a mental representation of an answer is generated and compared to existing options, and response elimination wherein partial solutions are generated based on isolated elements of the problem and incorrect answers are systematically ruled out. Constructive matching is the more dominant strategy in simple items and is used more often by high Gf participants.

Perceptual organization involves the grouping of visual perceptions into a Gestalt, for instance grouping by proximity, similarity, continuity, or common region. It has been found to either increase or decrease the complexity of a given problem (Primi, 1995). A rather intangible property of specific items here might be termed harmony, referring to the aesthetic elegance of a particular figural grouping that can make certain items less difficult. In a study of 313 undergraduates, this author related the amount of information in an item to individual differences in goal management, variables like rule complexity and perceptual organization to individual differences in selective encoding and abstraction; systematically controlling for each of these four variables in a repeated measures design showed that abstraction significantly affects item complexity while goal management does not.

Beyond strategy and stimuli-specific consideration, executive functions can also be evoked to account for differences in RPM performance. Working memory capacity has such a strong correlation with fluid intelligence that some researchers have even suggested that they are reflective of the same underlying construct (Martinez, 2011). Unsworth and Engel (2005) tested Carpenter's hypothesis that working memory would become more critical as the RPM increased in item difficulty, finding instead that working memory correlates equally to RPM items requiring fewer rules as those requiring multiple rules. Wiley et al. (2011) instead proposed and tested the interference/distraction model, finding evidence that the application of new rules on items is most strongly associated with working memory capacity; however, as noted by Harrison (2015), these conclusions may have been confounded by methodological limitations, including a fixed order in their first study making the potential for idiosyncratic characteristics of the particular items potentially problematic, as well as a small sample size in their second study. Furthermore in an attempted replication with 99 subjects, Engel (2005) was not able to replicate their findings. In another experiment utilizing a new set of RPM-type problems with varying item orders set to counterbalance the exposure to novel and repeated rules, Harrison (2015) found that the correlation between working memory capacity and novel rule problems was lower (.36) than that between working memory capacity and repeated rule problems (.50), supporting instead the learning efficiency account. Moreover, using a median split of fluid intelligence composite scores, he found that this relationship did not differ between low Gf and high Gf participants. Such results suggest that Gf should be highly predictive of the ability to disengage from irrelevant information stored in memory, and to instead apply the relevant rule to any repeating item.

These investigations into the RPM and RPM-type problems suggests 1) that there are at least two distinct classes of strategies which might be applied to the overall set of RPM items, reflecting a broadly verbal-analytic strategy on the one hand, and a broadly visuospatial one on the other; and 2) that there are a variety of factors that can affect performance on this task, but that unstructured experiments can easily lose sight of what manipulation is making a difference to performance. In order to gain tractable ground on the problem of decomposing Gf tasks into discrete strategies, a well-controlled and extensively studied paradigm would be ideal. The next section will discuss the potential for investigating cognitive strategies within the context of task-switching paradigms.

Strategy Flexibility and Task Switching

While many computational models of strategy selection have been proposed (including the ACT-R model, the RCCL model, the adaptive decision-maker model, the SSL model, and that SCADS model, etc.) all of them assume that strategies are selected on a problem by problem basis. However, only one such model (the SCADS model) allows for the potential impact of a strategy interruption mechanism, wherein a participant may choose a particular strategy but may realize mid-execution that an alternative strategy would be preferable. While there is little direct empirical support for such an idea, one study of two-digit multiplication problems using either a mental strategy or a calculator strategy to achieve the answer found that in a subset of items participants could be seen to initiate cursor movement towards the on-screen calculator before abandoning this approach and using a mental solution strategy (Walsch & Anderson, 2009). This suggests that for some items, the initial strategy selection can indeed be abandoned and a new strategy adopted. From a theoretical perspective, such within-item strategy revision may be strongly related to the executive function components of inhibition and cognitive flexibility, both of which have previously been seen to mediate the effect of age on strategy selection.

In two experiments utilizing the same computational estimation task described above, Ardiale and Lemaire (2012) hypothesized that such within-item strategy revisions would result in greater accuracy despite potentially greater reaction time, and that younger adults would engage in this practice more than older adults due to the effects of age on executive function. In their sample 37 young and 37 older adults they found that older adults did indeed engage in fewer within-item strategy revisions, that the strategy revisions resulted in greater overall RTs, but that age did not affect the magnitude of these RTs, suggesting that (similar to local switch-costs within the task-switching literature) the executive processes responsible for within-item revisions are largely age invariant. In the second experiment utilizing a forced strategy cue (indicated by the font color of the stimulus) wherein half of the trials required participants to switch strategies and half did not, greater strategy switch costs were observed for older adults than younger adults. The age groups also differed in which variables were indicative of switch cost asymmetries, with the younger participants showing larger switch costs when switching from the poor strategy to the best strategy and older adults showing larger switch costs when switching from the rounding up strategy to the rounding down strategy, possibly reflective of the differential difficulty of these two strategies. The authors suggest that these findings could be accounted for by the impact of processes like priming or tacit reconfiguration processes.

These results were replicated by Taillan, Ardiale, & Lemaire (2013) in another experiment which collected both the mean percentages of strategy switches and the strategy switch costs, at least in heterogeneous problems (meaning problems in which the unit digits of the two numbers being multiplied were alternatively below and above five, making the selection of the best strategy more computationally intensive). This could indicate that older adults avoid switching strategies because of the increased cognitive load of doing so. They also found that switch costs when switching from the harder strategy to an easier strategy were greater than the reverse, conflicting with between-items strategy switches previously observed. Moreover, switch costs were found to be shorter in within-item strategy switching than in between-item switches observed in the past.

Based on similar findings by Luwell et al. (2009), Lemaire and Leclere (2013) used the well studied computational estimation paradigm described above to study the phenomenon of sub-optimal strategy repetition, wherein people are more likely to use the same strategy on a particular problem when they have previously used that strategy in the immediately preceding problem, even when the problem characteristics suggest a better strategy could be employed. This conflicts with most formal models of strategy selection, which suggest that strategies are chosen on a problem by problem basis; instead these findings suggest that strategy use is also influenced by the sequence of strategies used.

These authors suggest that this phenomenon may result from the additional cognitive load necessitated to change strategies, and that there may be an underlying bias to engage in the same strategies unless the advantages of switching to an alternative strategy are great enough. Assuming this, they hypothesized that participants should be even more likely to repeat a strategy on a given trial if they had used the same strategy on two preceding trials consecutively. Given that strategy selection is at least in part mediated by executive control functions, and that executive control functions decline with age, they also hypothesized that older adults would show a greater tendency toward strategy repetition than younger adults.

In a sample of 100 participants (50 young and 50 old) as expected, in the two-prime condition (meaning participants had used the same strategy for two preceding trials) both groups of subjects repeated the same strategy more often (71%) than in the one prime condition (56%). This effect was exacerbated for the rounding up problems, which are considered slightly more difficult. Also as expected, older participants repeated strategies more often than younger participants, but not by a great extent (66% versus 62%); again however this effect was exacerbated by the problem type–older adults were especially more likely to repeat strategies in the two prime condition following a rounding-up problem (which is slightly more difficult than rounding-down problems). In terms of strategy execution, repetition of a strategy tended to decrease reaction time in both young and old adults.

This initial exploration into strategy shifts across trials suggests that the familiar task-switching paradigm can be successfully applied to the domain of cognitive strategies. As Rogers and Munsell noted (1995), the precise definition of what constitutes a “task” within this paradigm is lacking. The most frequently employed tasks within the task-switching literature are basic single step stimulus-response tasks. Another common feature is the use of bivalent tasks, where in certain stimuli can be associated with both tasks the participant completes (e.g, classifying some array of letters as vowel/consonant in one task, or as lowercase/uppercase in another). Moreover, the response mappings to these bivalent stimuli can sometimes be the same (e.g. pressing a left arrow key in response to a stimulus such as "e", corresponding to either the rule "vowel" or the rule "lowercase"); when this is the case such trials are called congruent. Another potential characteristic of S-R mappings that can influence test switching behavior is the relative difficulty of applying such mappings; ranging from easy (e.g. for overlearned responses such as in word reading) to difficult (e.g. for arbitrary mappings, such as pressing a left arrow key in response to a vowel).

This raises the question however of where exactly a "strategy" ends and a new "task" begins. While much task switching literature treats the "task" as a monolithic construct, some evidence exists to suggest that tasks must be much more precisely defined in order for researchers to understand their underlying cognitive processes. Ravizza and Carter (2008) argue that many studies conflate shifts of visual-spatial attention or perceptual switching with shifts of contextual rules or rule switching. For instance, in task-switching paradigms that rely on a visual cue such as the color or shape of an object, effective switching would require both a shift in visual-spatial attention away from one set of features to another, but would also entail implementation of the appropriate set of response rules. They point out that terms such as "set shifting", "task switching", and "attention switching" are often used almost interchangeably and without precise definition, but that they may legitimately refer to different underlying phenomena. Previously Allport (1994) attempted to identify if there were differences in switch costs incurred on the basis of shifts in stimulus dimensions, semantic categories, cognitive operations, and response modes, and found that all of these incurred similar switch costs. However, Ravizza and Carter argue that this paradigm was investigating primarily mixing costs which are known to be driven by additional working memory load. They hypothesized that perceptual shifting should only be seen when the competing stimulus set is present, necessitating a switch in visual-spatial attention, but that rule switching would be present even in the absence of such stimulus interference. They further hypothesized that target repetitions would lead to a lower switch cost in perceptual switching but not in rule switching.

Much of the neuroimaging literature is mixed about whether the neural substrates of these varying types of shifting differ; Rushworth (2001) was able to find subregions of the medial frontal cortex as well as the parietal cortex (2002) that were distinctively associated with perceptual shifting, but Wager (2005) reported very weak dissociations when directly comparing these alternative forms of task switches. To explicitly test for the neural dissociability of perceptual switching and rule switching, Ravizza and Carter utilized an odd-man-out design relying on sets of letters and shapes presented simultaneously. Switch trials were designed such that the response rule differed from the one participants had followed in the immediately preceding trial. Behaviorally, switch costs to RT were greater in the perceptual switching condition than the rule switching condition. In a separate study of 14 young participants utilizing the same test design run within the context of fMRI imaging, the authors report that a region of the left dorsolateral prefrontal cortex was sensitive to rule shifts but not to perceptual shifts, while a region in the right superior parietal cortex was sensitive to perceptual shifts but not to rule shifts, and likewise for a region of the right premotor cortex. Furthermore, greater activity in the left dorsolateral prefrontal cortex is also predictive of better performance in the rule shifting condition, while greater activity in the right superior parietal cortex predicted diminished performance in the perceptual shifting condition. To further support the neural dissociability of these different types of shifting, these researchers also conducted a small meta-analysis of neuroimaging studies probing perceptual shifting but not rule shifting, finding that none of the studies found significant correlations in the dorsolateral prefrontal cortex as would be expected if indeed this was only associated with rule shifting. Similarly, in a sample of neuroimaging studies with rule shifting tasks that do not engage perceptual shifting, lateral regions of the prefrontal cortex are universally engaged (although the peak region of activity varies widely from study to study).

Given this pattern of findings, combining alternative fluid reasoning strategies with a task-switching paradigm represents a novel yet highly tractable way to investigate the relationship between age and strategic flexibility, and may suggest further studies depending on the contributions of factors such as working memory.

CHAPTER 3
STUDY ONE

Introduction

The task-switching paradigm refers to common experimental design wherein participants perform a series of distinct tasks on a series of similar items in succession. A cost in performance (typically measured in reaction time or accuracy) is expected when the participant must perform a task on a given trial that differs from the task she was asked to perform on the immediately preceding trial; this is termed the switch cost. Switch costs are predominantly classified into one of two types: a global switch cost/mixing cost, referring to the difference in performance between a non-switch trial in a pure task block (meaning a task block in which each item necessitates performing the same task) and a non-switch trial in a mixed block (meaning a block in which the tasks alternate in some fashion). A local switch cost instead refers to the difference in performance between a switch trial in a non-switch trial within a single mixed block. While global switch costs (sometimes called mixing costs) are theoretically related to goal maintenance in the central executive as well as added load to working memory capacity, local switch costs instead are theoretically related to the executive process which deactivates recent test sets and activating a new and more relevant tasks at (Monsell, 2003). Cray and Lindenberger (2000) confirmed the dissociability of these two types of costs using structural equation modeling.

Stimuli for the study of strategy-switching in Raven's-like problems were generated using a software package described by Matzen et al (2010). These authors designed a program to systematically generate new Raven's-like matrix problems, allowing total control over a variety of parameters of each item. They also normed a sample of these problems generated using this software and compared the behavioral characteristics to the classic RPM items. Previous research on the RPM has identified certain problem features that must be taken account when generating new items. The difficulty of problems depends highly on the number of rules or relations that must be utilized in combination to solve them, with zero-relation problems being matrices in which no element transformations occur and the answer is a simple one-to-one match to a shape that is repeated in the item, one-relation problems being those in which one rule dictates the pattern of changes a test-taker must implement to achieve the solution, etc.

Using some of the classic taxonomies developed by Carpenter et al (1990), these authors identified two broad sets of problems within the RPM: problems involving object transformations, and logic problems. While logic problems utilize rules such as conjunction ("AND"), disjunction ("OR"), and exclusive disjunction ("XOR"), object transformation problems instead focus on systematic alterations to features like the shape, size, shading, numerosity, and orientation of figural elements. Each of these object transformations can be manipulated to apply in a particular direction within the matrix, either row-wise, column-wise, diagonally (from either the top left to lower right of the matrix or from the lower left the top right of the matrix), or in an outward progression beginning from the top-left corner. Using Carpenter's taxonomy, such rules could be directly mapped onto existing rule types, such that horizontal and vertical transformations would be equivalent to the constant-in-a-row rule or quantitative progression rule, diagonal transformations match the distributional-of-three-values rule, and outward progression would be a combination of two quantitative progression rules.

The Sandia matrix generation software was written in Java (J2SE JRE:6), and the software uses a recurring set of six basic shapes as feature elements–ovals, rectangles, triangles, diamonds, trapezoids, and T-shapes. The shapes were selected because of their easy identifiability, as well as their susceptibility to all of the previously listed rules. In addition to generating the matrix problems themselves, the software also generates a 2 x 4 matrix of potential answer choices for each problem, containing the correct answer in a position specified by the user as well as seven incorrect distractors generated based on models of the structure of the RPM distractors. The distractors are generated by a random combination of the following set of procedures: the correct answer has a random transformation applied to it; an incorrect trait is drawn from the shapes in the generated matrix; an incorrect shape is drawn from the shapes in the generated matrix, and randomly transformed; a transformation is randomly applied to a previously generated distractor; randomly sampled surface features from the generated matrix are combined; the shape is generated with novel surface features that do not appear the matrix. The software is designed to allow generation of object transformation problems with zero, one, two, or three relations

To test the items generated using their software, the authors generated a subset of 840 representative problems for investigation in a norming study. The problems were designed such that each particular combination of relation and direction were tested. The authors hypothesized that one-relation matrices would be the easiest to solve, and that higher-order relation matrices would be more difficult, with three-relation matrices necessitating diagonal application of rules or outward progression being more difficult than those requiring row-wise or column-wise application of rules. They also speculated however that logic problems would be the most difficult to solve, both because of the more abstract nature of the rule involved, and because the overlapping nature of these problems makes each individual shape less salient. In a sample of 80 undergraduate participants (52 female), the authors administered both the RPM task as well as a subsample of the generated matrices so that each participant received a different selection of 20 problems from the list of 840 matrices. Results indicated similar levels of reliability (Cronbach's alpha = .73 for the RPM problems, and equal to .76 for the generated matrices) and correlation between accuracy for the two sets of problems was .69, with an attenuation-corrected correlation of .93. Analysis of the object transformation problems indicated that there were no significant differences in accuracy performance on the basis of the type of transformation applied, so changes to shape, shading, size, orientation, and number had relatively equivalent difficulty. Direction of transformation was also assessed, with no significant differences in accuracy being found between row-wise or column-wise transformations, or between these horizontal transformations and diagonal transformations; however outward progression problems were significantly more difficult in both one-relation and two-relation problem types. In particular, outward transformations involving shading were significantly more difficult than other outward transformation problems. The same pattern held for three-relation problem types.

Analysis of generated logic matrices revealed that the average accuracy is quite low (37.9%), and significantly different than the average accuracy on the RPM logic problems (53.6%). The "OR" problems had the highest accuracy (.46), while the "AND" and "XOR" problems had equivalent accuracies (.34); further analysis revealed that the only specific rule type that significantly differed in accuracy from the RPM problems was the "OR" condition, and this was driving the entire effect. The authors speculate that this might be due to the fact that the “OR” problems in the RPM are visually very simple, whereas the figural elements in the generated matrices tend to overlap to a much greater degree and thus reduce the salience of each individual element.

This study was designed to test a sample of matrices generated using this software, and ordered in a way that should induce switching costs if indeed shifts in broad strategy use result in additional cognitive load. To this effect we designed two distinct sets of stimuli: logical items (requiring use of rules “AND”, “OR”, and “XOR”) were generated, roughly corresponding to the previously identified logico-analytic strategy for solving RPM problems; object transformation items (requiring use of the color, shape, size, and number transformations) were also generated, with these assumed to correspond to the visuospatial strategy. All stimuli were generated to use only one set of these rules, and so no item can be solved by utilizing the alternative strategy (unlike a selection of items from the RPM). Additionally, a selection of the normed stimuli from Matzen’s study were also included where they fit the needed parameters, since the relative difficulty of these items had already been established. All new matrices generated were piloted by a small sample (n=20) of young adults to estimate their relative difficulty and to gauge appropriate time limits for the task. While the RPM allows participants to spend as long as required for each individual item, we imposed restrictions on time in order to more precisely measure switch-costs in the form of RTs.

Methods

Participants

Participants were recruited via Amazon's Mechanical Turk platform. Thirty-seven young (20-40) and thirty-one old (60-80) subjects took the task over a period of 2 weeks. In order to be eligible to participate, all participants had to be native English speakers and reside in the US. Subjects also completed a pre-study questionnaire asking them to report their age, gender, and years of education. The entire duration of the task was approximately 25 minutes, and subjects were compensated $3 for their participation.

Studies of MTurk samples show generally good validity and reliability. Samples culled from MTurk generally compare well to more traditional samples in terms of psychometric properties (Buhrmester et al., 2011; Behrend et al., 2011), and have been found to accurately reflect their reported demographics (Rand, 2012). Furthermore as compared to traditional convenience samples such as undergraduate students, MTurk workers display higher levels of demographic and psychometric diversity (Buhrmester et al., 2011). While this worker pool tends to be slightly more educated than the general population, and typically has a higher representation of males (Berinsky et al., 2012), MTurk samples have been widely used to replicate research findings previously demonstrated in more traditional samples.

Measures

Pre-Study Screening Questionnaire

A basic pre-study screener was administered to all participants asking them to enter their assigned subject ID, their age in years, their years of education, their biological sex (male or female), and whether English was their native language.

STRATA Design

Participants received the original form of the STRATA task. The task was programmed using PHP and displayed 4 blocks of 20 trials each. The first block consisted entirely of logico-analytic strategy items, displayed in a randomized but fixed order for all participants. The second block contained a mix of logico-analytic and visuospatial strategy trials, again randomized but fixed across subjects; there were an equal number of each problem type, and the block was designed to include 9 “switch” trials (i.e trials wherein subjects have just completed a trial of the alternate strategy type and now they are seeing the other one) and 11 “non-switch” trials. The third block contained exclusively visuospatial strategy trials but was otherwise identical in structure to the first block, while the fourth block was again mixed and identical in structure to the second block.

Figure 1 displays the structure of this task.

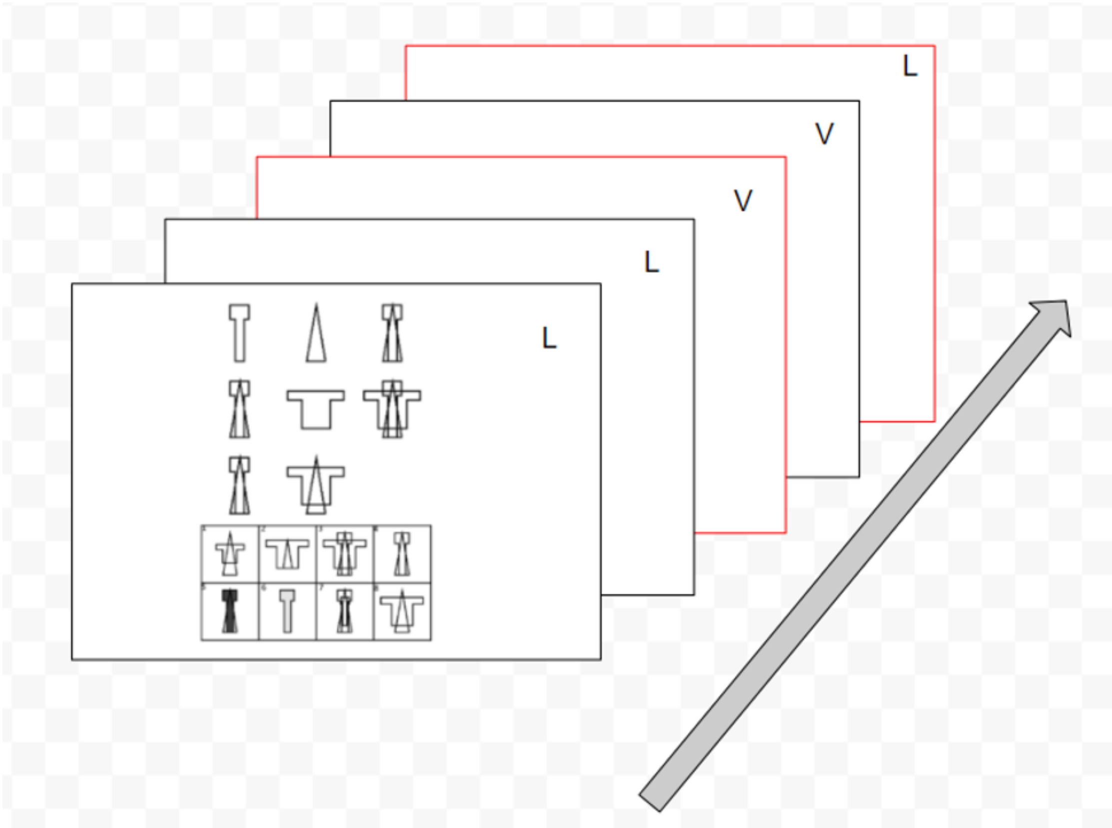

Figure 1. Stimuli ordering in a mixed block. L = logico-analytic problems. V = visuospatial problems. Red borders indicate a switch trial.

Each trial was preceded by a 500-ms inter-trial-interval (ITI), and the trial ended either when participants entered a numeric value from 1 to 8, representing their answer to the problem, or after 30-seconds elapsed without a response. In the latter case these trials were coded as “Time-Outs”. For each item, accuracy and RT (in milliseconds) were recorded.

Stimuli. Stimuli included in the norming study (Matzen et al., 2010) matching either the two-relation transformation problem type (row-wise) or the logical problem type, and demonstrating high levels of accuracy (around .75) were included in the test set. Two-relation problems consisted of every combination of object manipulation (e.g., shape +

number, size + orientation, etc.), both in order to use as many pre-normed items as possible and to test whether these combinations differed in terms of performance. This resulted in a full set of 40 visuospatial problems, and a partial set of 5 logico-analytic problems (all of which consisted of the "OR" rule). We generated an additional set of 100 logical items using the Sandia tool, and these were designed to match the structure and relative figural complexity of the normed stimuli. Piloting results showed a range of mean accuracies for these items, ranging from .8 to .15, from which we selected the top 35 most accurately answered items for inclusion in the study; the least accurate of these items still showed comparable mean accuracy (.65) to the mean accuracy of the visuospatial stimuli. Although the norming data did not include reaction time, we observed that at least in young adults, a time limit of 30 seconds per item resulted in over 90% of the items being answered.

An example of each problem type is shown in figures 2 and 3.

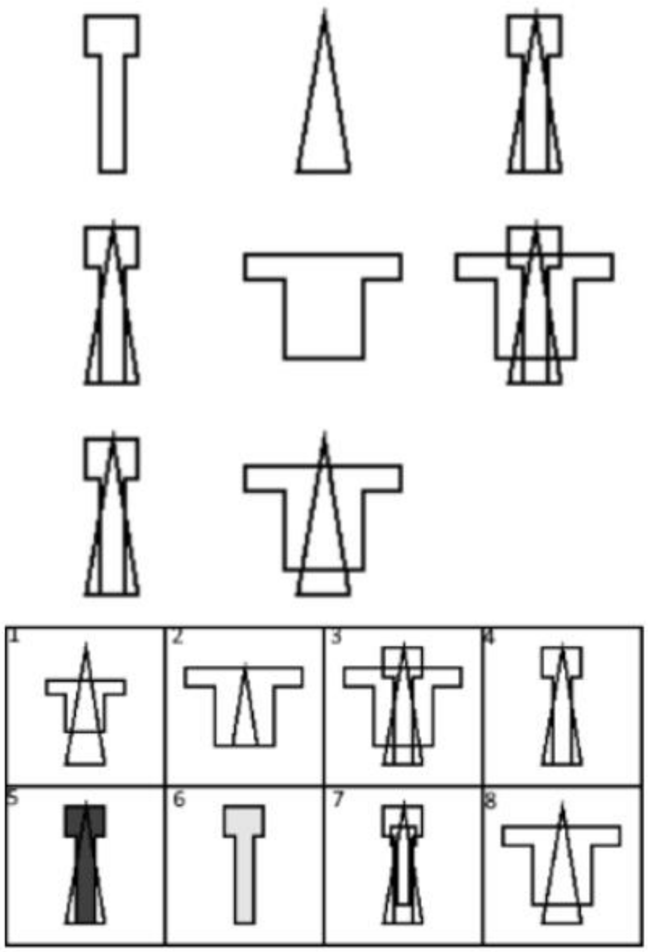

Figure 2. Sample logico-analytic problem.

(“OR” rule).

Figure 2. Sample logico-analytic problem (“OR” rule).

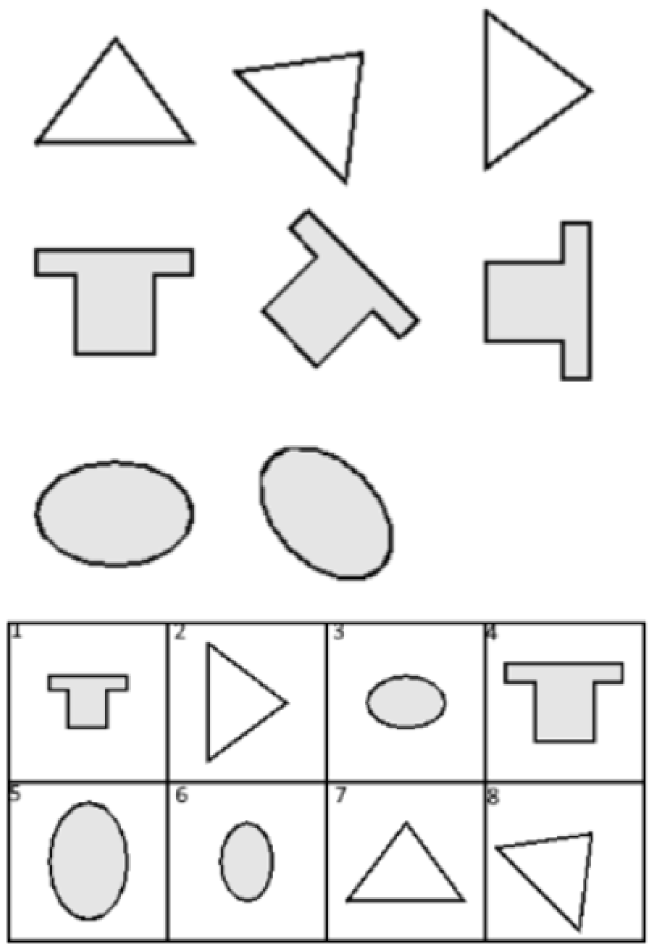

Figure 3. Sample visuospatial problem.

("Shading+Orientation" transformation).

Figure 3. Sample visuospatial problem ("Shading+Orientation" transformation).

Procedures

Participants were recruited via Mechanical Turk and linked directly to the URL for the study, where they viewed an informed consent form before completing the screening questionnaire. Following the questionnaire, participants viewed the instructions for the task. They were instructed to complete each item by finding the pattern in the matrix, and to select the answer choice that best completes the pattern. They were then given 4 self-paced practice trials, with two examples of each item type (visuospatial and logico-analytic). For each practice item, detailed feedback about the series of necessary steps to solve the item were shown to participants after each incorrect response. Following the practice, participants were instructed that they would have a 30-second time limit on each problem so they had to work as quickly as possible on each problem without sacrificing accuracy. They indicated when they were ready to commence the task with a press of the space-bar. Between blocks 2 and 3, participants were presented a screen with a visible two-minute countdown timer and told that they could rest for this time, or press a key to continue with the final two blocks. Following the last block, participants were thanked for their participation and their study compensation was automatically released.

Data Analysis

Behavioral Analysis. Behavioral analyses were conducted on two separate dependent measures: reaction time (RT) and accuracy (ACC). To ensure the non-normal distribution of the percentages would not impact the analyses, we also computed the arcsin transform of the ACC measure, and replicated each analysis using both variables (we report only results of the initial models, since these were identical to the arcsin models). The RT models used mean RT for correct trials only.

The initial set of simple models used a 2 x 2 repeated-measures ANOVA, with one within-subjects factor (either switch type or block type) and one between-subjects factor (age group). These were designed to test explicitly for differences in global (overall block-level differences, utilizing block type alone as the within-subjects factor), local (switch versus nonswitch differences within mixed blocks, utilizing switch type as the within-subjects factor within mixed blocks only), and mixing costs (mixed block nonswitch versus pure block nonswitch differences, utilizing block type as the within-subjects factor within switch trials only). The level of significance was set at .05. Prior to RT analyses, all RTs for incorrect responses were excluded from the data set. For correct responses, RTs shorter than 300 ms and longer than 2 SDs above the mean RT were also excluded; the former due to their likelihood of being anticipation errors. This resulted in the removal of 87 trials from the subsequent analyses.

To analyze the effects of covariates on the dependent measures, we constructed two separate general linear models (GLM) that were analyzed in stages (heterogeneous slopes) (Kumar et al., 2008; Siegel, 1956). In each model we introduced the covariates of interest, in this case gender. The initial, heterogeneous-slopes model used a repeated measures analysis of variance (ANOVA) design with strategy type as a within-subjects factor (visuospatial versus logico-analytic problems), and added the main effects of the covariate in addition to its interaction with age group. The latter effects tested the assumption of the analysis of covariance that the effects of the covariates are equivalent at each level of the model's fixed effects, and only these equivalent effects were further inspected. In the first stage, we constructed a full model with the following predictors: age group, gender, education, age group X gender, and age group X education. Gender was controlled for because of past research which observed sex differences in strategy use, particularly within the context of the RPM task (Lynn, Allik, & Irwin, 2004; Macintosh & Bennett,2005). After performing this full model, retaining strategy-type as a within-subjects factor, we constructed a reduced model which retained only the covariate main effects, as well as any interaction terms which were statistically significant in the full model ($p < .05$).

Post-hoc analyses for all significant ANOVAs were conducted to further inspect the pattern of these relationships. These were designed within the context of a linear mixed-effects model to replicate the structure of the mixed ANOVAs and computed using the 'nlme' package in R, with either RT or ACC being predicted by age group as well as the task factor of interest. Next, the equivalent of Tukey HSDs were computed simultaneously for all linear models using R's multicomp package, with p-values Bonferroni-corrected.

Results

Participant characteristics are given in table 1. T-tests were used to ensure that the covariates were not significantly different in each group. Error rates were on average low (ranging from 21% to 32%). On average, older adults (whose mean accuracy was .68 across all trials) displayed more errors than younger adults (mean accuracy = .79). Older adults demonstrated greater reaction times depending on block type, with the largest RTs being observed in mixed blocks than in pure blocks.

Table 1

Study 1 Participant Characteristics

	Young	Old
Age	26.54 (7.81)	67.09 (7.38)
N	37	31
% Female	58	65
Education	14.09	14.51

Note. Education is reported in years attained.

Global (Block-Level) Switch Costs

To investigate the overall impact of age and strategy type, two mixed ANOVAs were conducted with age group as the between-subjects factor and block type as the within-subjects factor. These ANOVAs were conducted to determine if global differences in reaction time and/or accuracy could be observed based only on the between-subjects factor and the type of block.

RT. Analysis of global differences to reaction time (i.e. differences across all blocks) based on age group and type of block resulted in a significant main effect of block type, $F = 50.906$, $p < .05$. Age group, and the interaction of age group with block type, were not significant ($F= 1.233$, and $F = 2.21$, respectively). Results of this ANOVA are reported in table 2.

Table 2

ANOVA results: Study 1 global RT

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	66	19403675754.70	384651561.91	3329.36	.000	.98
age_group	1	66	7188783.30	384651561.91	1.23	.271	.02
block_type	1	66	21068468.28	27315297.90	50.91	.000	.05
age_group x block_type	1	66	915952.82	27315297.90	2.21	.142	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for RT. At the block-level, post-hoc analyses revealed that older adults in mixed blocks showed significantly higher RTs than younger adults in pure blocks, and this was the greatest estimated difference between subgroups. Older adults in the mixed blocks also displayed significantly higher RTs than older adults in the pure blocks. Finally, young adults in the mixed blocks also showed significantly higher RTs than young adults in the pure blocks. See table 3 for these results.

Table 3

Post Hoc Test: Global RT Differences (Mixed Blocks Minus Pure Blocks)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	297.1	429.4	0.692	1.00000
Young, Mixed Blocks - Young, Pure Blocks	626.9	149.4	4.197	0.000162 ***
Old, Mixed Blocks - Young, Pure Blocks	1252.1	429.9	2.912	0.021518 *
Young, Mixed Blocks - Old, Pure Blocks	329.8	429.7	0.767	1.00000
Old, Mixed Blocks - Old, Pure Blocks	955.0	163.8	5.831	3.3e-08 ***
Old, Mixed Blocks - Young, Mixed Blocks	625.2	430.2	1.453	0.876961

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

ACC. Analysis of global differences to mean proportion accuracy (i.e. differences across all blocks) based on age group and type of block displayed the opposite pattern of findings; here we observed a main effect of age group, $F = 5.373$, $p < .05$, as well as a significant interaction between age group and block type ($F = 7.736$, $p <.05$). Block type alone ($F =8.905$) by contrast did not result in a significant main effect as with RT. These results are reported in table 4.

Table 4

ANOVA results: Study 1 global ACC

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	66	75.46	0.40	12502.68	.000	.99
age_group	1	66	0.32	0.40	53.73	.000	.33
block_type	1	66	0.00	0.28	0.02	.891	.00
age_group x block_type	1	66	0.03	0.28	7.74	.007	.05

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for ACC. At the block-level, post-hoc analyses revealed that older adults in mixed blocks showed significantly lower accuracy than younger adults in pure blocks. Older adults in the mixed blocks also displayed diminished accuracy relative to young adults in the pure blocks, and relative to young adults in mixed blocks. Younger adults in the mixed blocks also displayed slightly greater accuracy than older adults in the pure blocks. Table 5 displays these results.

Table 5

Post Hoc Test: Global ACC Differences (Mixed Blocks Minus Pure Blocks)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	-0.06713	0.01544	-4.349	8.20e-05 ***
Young, Mixed Blocks - Young, Pure Blocks	0.03264	0.01642	1.988	0.281
Old, Mixed Blocks - Young, Pure Blocks	-0.09651	0.01807	-5.340	5.58e-07 ***
Young, Mixed Blocks - Old, Pure Blocks	0.09977	0.01768	5.643	1.00e-07 ***
Old, Mixed Blocks - Old, Pure Blocks	-0.02938	0.01793	-1.639	0.608
Old, Mixed Blocks - Young, Mixed Blocks	-0.12915	0.02003	-6.449	6.75e-10 ***

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Learning Effects. To analyze the potential for learning within a block as items progress, the average RT for each item within the two pure blocks was plotted against item index. As shown in figures 4 and 5, no evidence for learning in the form of reduced RTs during later trials is observed. Paired T-tests confirmed that no significant difference between each participant's mean RT in the first half of a block and the second half are present.

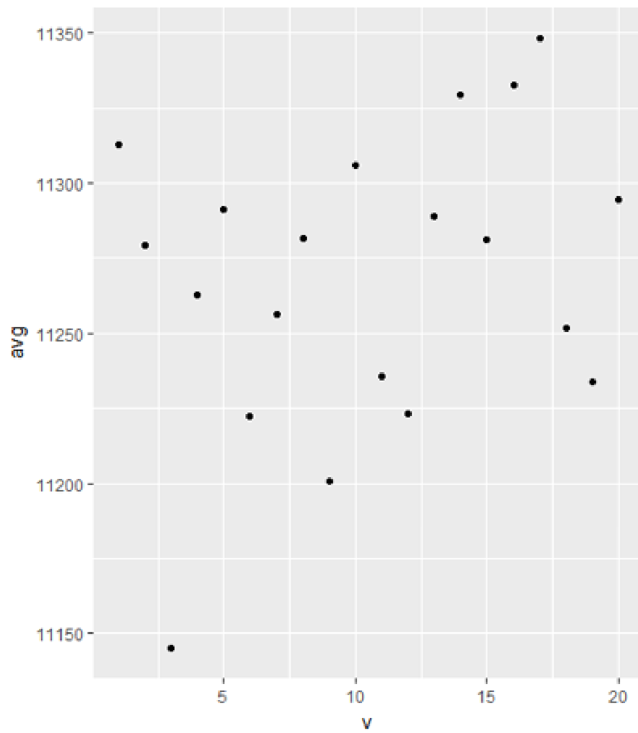

Figure 4. Average RT Across Pure Block Trials. Block 1 (logico-analytic).

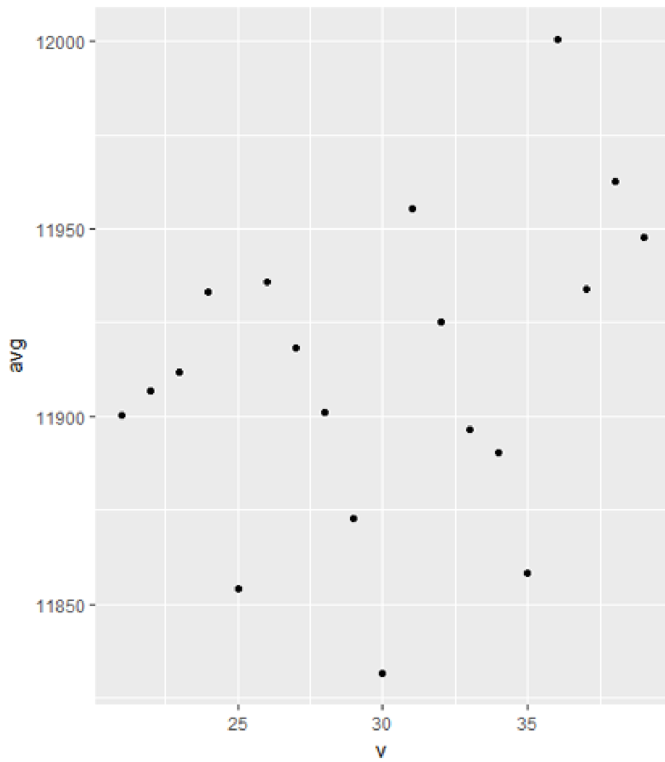

Figure 5. Average RT Across Pure Block Trials. Block 2 (visuospatial).

Local Switch Costs

To analyze the local switch costs (i.e. differences in performance between switch and nonswitch trials within the mixed blocks), mixed ANOVAs were conducted controlling for the impact of block type on both reaction time and accuracy. A 2 (age group: young or old) by 2 (trial type: switch or nonswitch) design was utilized exclusively within the mixed blocks.

Local Switch Costs: RT. Local switch-costs to RT within mixed blocks were significantly associated with trial type (switch or nonswitch), $F = 5.95$, $p < .05$. No significant main effect of age-group, or interaction effect between age group and trial type, was observed. Results of this ANOVA are described in table 6.

Table 6

ANOVA results: Study 1 local switch costs in reaction time

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	66	20546087989.83	449531768.38	3016.57	.000	.98
age_group	1	66	12940877.08	449531768.38	1.90	.173	.03
switch_type	1	66	3609338.21	40020873.72	5.95	.017	.01
age_group x switch_type	1	66	132963.30	40020873.72	0.22	.641	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for Local RT Switch Costs. At the block-level, post-hoc analyses revealed that older adults in mixed blocks showed significantly higher RTs than younger adults in pure blocks, and this was the greatest estimated difference between subgroups. Older adults in the mixed blocks also displayed significantly higher RTs than older adults in the pure blocks. Finally, young adults in the mixed blocks also showed significantly higher RTs than young adults in the pure blocks. See table 7 for these results.

Table 7

Post Hoc Test: Local RT Differences (Switch Trials Minus Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	330.3	458.7	0.720	1.00000
Young, Mixed Blocks - Young, Pure Blocks	628.4	189.1	3.324	0.00533 **
Old, Mixed Blocks - Young, Pure Blocks	1308.6	459.8	2.846	0.02659 *
Young, Mixed Blocks - Old, Pure Blocks	298.1	459.5	0.649	1.00000
Old, Mixed Blocks - Old, Pure Blocks	978.3	207.3	4.720	1.42e-05 ***
Old, Mixed Blocks - Young, Mixed Blocks	680.2	460.6	1.477	0.83860

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Local Switch Costs: ACC. Local switch-costs to ACC within mixed blocks were assessed identically to RT, and resulted in a significant main effect of age group ($F = 3.357$, $p < .05$). The main effect of trial type, and the interaction between trial type and age group, were not significant. See table 8 for detailed ANOVA results.

Table 8

ANOVA results: Study 1 local switch costs in accuracy

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	66	75.28	0.97	5139.69	.000	.98
age_group	1	66	0.56	0.97	38.57	.000	.25
switch_type	1	66	0.01	0.77	0.53	.470	.00
age_group x switch_type	1	66	0.00	0.77	0.01	.915	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for ACC. At the block-level, post-hoc analyses revealed that older adults in mixed blocks showed significantly lower accuracy than younger adults in pure blocks. Older adults in the mixed blocks also displayed diminished accuracy relative to young adults in the pure blocks, and relative to young adults in mixed blocks. Younger adults in the mixed blocks also displayed slightly greater accuracy than older adults in the pure blocks. Table 9 displays these results.

Table 9

Post Hoc Test: Local ACC Differences (Switch Trials Minus Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	0.03173	0.0289691	-4.142	0.000207 ***
Young, Mixed Blocks - Young, Pure Blocks	0.02678	0.0173981	0.442	1.000000
Old, Mixed Blocks - Young, Pure Blocks	-0.11593	0.02829	-4.098	0.000250 ***
Young, Mixed Blocks - Old, Pure Blocks	0.14326	0.02889	4.959	4.25e-06 ***
Old, Mixed Blocks - Old, Pure Blocks	0.01548	0.02924	0.529	1.000000
Old, Mixed Blocks - Young, Mixed Blocks	-0.12778	0.02506	-5.099	2.05e-06 ***

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Mixing Costs

Mixing costs (i.e. the difference in performance between nonswitch trials within mixed blocks versus nonswitch trials within pure blocks) were assessed by controlling for the trial type factor. A 2 (age group: young or old) by 2 (block type: mixed or pure) mixed design was conducted exclusively among nonswitch trials.

Mixing Costs: RT. Results of the ANOVA of RT values of nonswitch trials revealed a significant main effect of block type (F = 35.262, p < .05), but age group and age group x block type were not significant. ANOVA results are reported in table 10.

Table 10

ANOVA results: Study 1 mixing costs in reaction time

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	66	19064191699.60	347545323.28	3620.35	.000	.98
age_group	1	66	6142561.08	347545323.28	1.17	.284	.02
block_type	1	66	11330640.86	21207487.48	35.26	.000	.03
age_group x block_type	1	66	568942.33	21207487.48	1.77	.188	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Mixing Costs: ACC. ANOVA results for mixing costs to accuracy again demonstrated the opposite pattern as those for RT; block type alone was not a significant main effect, yet age group ($F = 28.58$) as well as the interaction of age group with block type ($F = 4.07$) were both significant ($p < .05$). Results of this mixed ANOVA are reported in table 11.

Table 11

Post Hoc Test: Mixed RT Differences (Pure Block Nonswitch Trials Minus Mixed Block Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	372.00	439.31	0.847	1.000
Young, Mixed Blocks - Young, Pure Blocks	497.05	42.56	11.680	<2e-16 ***
Old, Mixed Blocks - Young, Pure Blocks	1099.04	438.87	2.504	0.0736 .
Young, Mixed Blocks - Old, Pure Blocks	125.05	438.96	0.285	1.0000
Old, Mixed Blocks - Old, Pure Blocks	727.04	50.30	14.454	<2e-16 ***
Old, Mixed Blocks - Young, Mixed Blocks	601.99	438.53	1.373	1.0000

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Covariates. Both gender and education displayed low and insignificant correlations with RT and ACC.

Discussion

Results of the first study confirmed our hypothesis that strategy-shifting within a single task would result in switch costs as measured using RT. Subjects were overall less fast at making selections in switch trials, and showed greater errors during these trials. Also as expected, older adults showed even more dramatic mixing costs in error rates than young adults. Despite this, we could not find statistically significant evidence for our hypothesis that older adults are slower on switch trials than on non-switch trials within mixed blocks. While performance on mixed blocks was overall slower, particularly in the old group, than on pure blocks, RTs on the switch trials within any given mixed block did not significantly differ from those of the nonswitch trials.

One potential explanation for this however lies in the conflicting pattern of results observed for reaction time and accuracy; among all analyses assessing global switch costs, local switch costs, or mixing costs, age group and its interaction with the within-subjects factor were only significant predictors of accuracy, not reaction time differences. This pattern of findings has two potential implications that warrant investigation: 1) switch costs within the STRATA may be falling victim to a speed-accuracy trade-off, and 2) older adults may be preferentially sacrificing accuracy for lowered costs to speed. While the second implication is interesting and is itself suggestive of age-related strategic behavior, the presence of such a trade-off potentially compromises the interpretability of this work and hence will be explored more fully in study 2 below.

These results are overall consistent with existing task-switching literature; mixing costs have been repeatedly shown to be exacerbated in older adults, and this is thought to derive from age-related declines to working memory capacity. While mixing costs (or global switch costs) are taken as a reflection of the set up costs for maintaining and scheduling the two mental tasks sets, local switch costs are thought instead to reflect the inhibitory potential to deactivate a just active but now irrelevant mental task set. Since mixed blocks require maintaining two distinct tasks sets (or in this case, strategies) within working memory, mixing costs are not only frequently observed but are negatively impacted by advancing age. Local switch costs, meanwhile, which instead are thought to reflect a subject's inhibitory capacity to deactivate a now irrelevant rule and instead apply the appropriate rule, are less consistently exacerbated by age.

While there are some studies that suggests that older adults are subject to increases in both global switch costs and local switch costs (e.g., Merian, Gotler, & Perlman, 2001; Kray et al., 2002), the overall pattern of findings is more consistent with age effects on global switch costs, but only negligible effects on local switch costs. In a recent meta-analysis of 36 independent task-switching studies, Wasylyshyn & Verhaeghen (2011) found overwhelming evidence that older adults are especially prone to global switch costs but are no more impaired and local switch costs than younger adults. These authors even suggest that some of the results of the contrary may be due to specific aspects of the paradigm utilized which increased the importance of working memory; Kray et al. (2002) for instance found that local switch costs were more impacted by age than global switch costs, however here they used a paradigm which doubled the number of relevant tasks sets participants had to maintain (from 2 to 4) potentially exacerbating the working memory load within mixed blocks.

The observation of age-related mixing costs within the context of a strategy-shift within a single task is significant. Such a result extends the bridge between the two literatures of cognitive strategies and age-related group differences from both directions: first by reinforcing that classical neuropsychological instruments can indeed be decomposed into discrete strategic approaches, and second by showing that distinct strategies within even domain-general tests can be experimentally manipulated to gauge a more generalizable measure of the importance of such alternation.

One potential limitation of this study lies in the particular choices made about the task-switching sequence utilized. As noted, participants were given an unpredictable sequence of strategy-switches to apply, and these switches were signaled only by their own recognition of the pattern in the particular matrix they were presented with. Cueing paradigms, wherein subjects receive explicit cues about the type of task they will next see, often result in a smaller switch cost (Kray, Li, & Lindenberger, 2002; Kray & Lindenberger, 2000), possibly because they alleviate much of the demand to working memory capacity by offloading S-R rules. Unpredictable switch sequences have also been associated with larger switch costs (Rogers & Monsell, 1995), potentially the results of the removal of external inhibitory reinforcers which would otherwise make individuals more confident in the application of a specific rule when it is called for on a specific schedule. However, Wasylyshyn & Verhaeghen's (2011) meta-analysis did not find strong evidence that these factors significantly affected the relationship between age and task-switch costs.

Another potential limitation is the training for the task. While most task-switching paradigms rely on training prior to assessment, these vary widely in the amount and nature of the training presented. Some research has suggested that extensive practice can reduce switch costs during task-switching (Cepeda et al., 2001; Kray and Lindenberger, 2000), while some research suggests that mixing costs can be virtually eliminated with sufficient practice (Berryhill & Highes, 2009) within bivalent tasks– i.e. tasks wherein stimuli can have two response mappings. However, given that performance was held relatively constant across the duration of the tasks (including mixing costs to RT), we think the effects of practice were negligible in this study. Here in order to cut down on the time required to complete the study, participants were exposed to a small number of practice trials, but practice trials that gave explicit instructions about how to apply each strategy. It could also be argued that the explicit instruction removes some of the validity of the matrix items themselves, since they are probing the application of the appropriate rule rather than the ex nihilo generation of that rule. Future research is required to determine if knowledge of the rules to apply affects the strategic engagement ordinarily presupposed by the particular trial type.

CHAPTER 4
STUDY TWO

Introduction

The second study was designed to replicate the effects observed in the first study, and to examine the relationship between the strategic flexibility measure and a proxy measure of CR. This experimental procedure was also revised to more precisely match existing task switching paradigms. Specifically, stimuli ordering was reconfigured to correspond to the so-called alternating runs paradigm in the task switching literature (Monsell, 2003).

As Rogers and Monsell noted (1995), the precise definition of what constitutes a "task" within this paradigm is lacking. The most frequently employed tasks within the task switching literature are basic single step stimulus-response (S-R) tasks. Another common feature is the use of bivalent tasks, wherein certain stimuli can be associated with both tasks the participant completes (e.g, classifying some array of letters as vowel/consonant in one task, or as lowercase/uppercase in another). Moreover, the response mappings to these bivalent stimuli can sometimes be the same (e.g. pressing a left arrow key in response to a stimulus such as "e", corresponding to either the rule "vowel" or the rule "lowercase"); when this is the case such trials are called congruent. Another potential characteristic of S-R mappings that can influence task switching behavior is the relative difficulty of applying such mappings; ranging from easy (e.g. for overlearned responses such as in word reading) to difficult (e.g. for arbitrary mappings, such as pressing a left arrow key in response to a vowel).

The introduction of the alternating runs paradigm (as opposed to Jersild's original task switching paradigm, employing mixed blocks of the form ABAB which did not include any non-switch trials) allowed for the observation of alternation or mixing costs. Normally, subjects are slower overall during such mixed blocks than they are during pure blocks, suggesting a global cost associated with switching tasks. Rogers and Monsell (1985) argued that global costs may reflect the additional working memory load necessitated by mixed blocks, and may not be related directly to what was traditionally meant by task switching. The alternating runs paradigm attempts to correct for this potential confound by allowing the direct comparison of such block level global switch costs, to performance costs associated with switch trials but not with non-switch trials within a mixed block. Such local switch costs are closer in meaning to the original theoretical conception of task switching. This paradigm has even been applied to more complex tasks sequences, so long as they are presented in a rigidly predictable order (e.g. Gotler, Meiran, & Tzelgov, 2003; Koch, 2001, 2005, 3008; Logan,2007; Schneider & Logan, 2006).

Switch costs are commonly calculated in two distinct ways: local switch costs refer to the difference in performance measure (e.g. RT) on non-switch versus switch trials within a mixed block. These are associated with the execution of the task switch. Global switch costs, or mixing costs, are instead calculated as the difference between a performance measure on non-switch trials in a mixed block versus non-switch trials in a pure block. These are associated with factors relating to goal management and the retrieval of appropriate goals.

While working memory capacity has been shown to share a significant degree of variance with fluid intelligence (Ackerman, Beier, & Boyle, 2005), and is considered critical in many theoretical models of task switching because deactivating an irrelevant task set and reactivating a newly relevant task set imposes demands on this capacity, many studies investigating the relationship between these two components do not find a correlation (e.g. Keisel, Wendt, & Peters, 2007; Logan, 2004; Miyake et al, 2000). Moreover, some studies actually find that a higher working memory capacity is predictive of decreased task-switching performance (Draheim et al, 2015).

Draheim et al. (2015) argue that one potential reason for this lack of association lies in the methodology by which tasks switches are measured. Latency switch costs, whether on the global or the local level, do not take into account the accuracy with which individuals are performing the given task. Individual differences may exist between test-takers in terms of their calculation of the speed-accuracy trade-off; while some individuals may preference speed, others may prioritize accuracy, and vice versa. Another major methodological issue is the comparatively low reliability of difference scores (Cronbach & Furby, 1970). Based on how difference scores are calculated, it's a mathematical necessity that if the two component scores are highly correlated (which we would expect given that they are measured within-subjects and likely reflect highly similar mental processes), the reliability of the scores will be low. Given these two limitations of difference scores, the lack of correlational relationship between working memory capacity and task switching may be a function of how the latter is computed, rather than the absence of a genuine relationship.

As an alternative to difference scores within task-switching, Hughes et al. (2014) proposed a rank ordering binning procedure capable of combining speed and accuracy into a single metric. Unlike difference scores, this metric has been shown to have high reliability

(Hughes et al., 2014). Using the binning procedure, Draheim et al. (2015) re-analyzed a large set of data from 552 subjects collected by Shipstead et al. (2015). Not only did the internal consistency and reliability of task switching performance improve when operationalized this way, but the expected relationship between task switching ability and working memory capacity showed a significant correlation in the expected direction (.49, versus -.26 when measured using difference scores). In a further reanalysis of task switching and working memory data on 131 subjects collected by Oberauer et al. (2003), the binning procedure dramatically improved the amount of shared variance between the new task switching score and six individual working memory tasks. Shipstead et al. (2015) further used this binning procedure to show that Gf measures contribute significantly more unique variance over working memory capacity to the explanation of task switching scores than the reverse.

Given that the purpose of this study is to investigate expected relationships between task-switching ability and other covariates, we utilize this nontraditional method of scoring to explore how age and our lifetime exposure variables interact with strategy switching performance. We hypothesized that older adults would overall display higher bin scores (reflecting a combined measure of greater errors and higher RTs), that bin scores should be higher in mixed blocks than in pure blocks, and that our CR measures would moderate the effect of age on bin score. We further speculated that the inclusion of within-strategy rule switches would induce small changes to bin score, but that these would be lesser in magnitude than those changes induced by full strategy shifts.

Methods

Participants

Participants were recruited via Amazon's Mechanical Turk platform. Forty-eight young (20-40) and thirty-five old (60-80) subjects took the task over a period of one month. In order to be eligible to participate, all participants had to be native English speakers and reside in the US. Subjects also completed a pre-study questionnaire asking them to report their age, gender, and years of education. The entire duration of the task was approximately 30 minutes, and subjects were compensated $4 for their participation.

Three young participants and one older participant were excluded for failing to complete the task. Additionally one older participant was excluded for failing to respond to the majority (>90%) of trials. This left a final study sample of 45 young adults and 33 older adults. Informed consent was obtained at the beginning of the study.

Measures

Cognitive Status Questionnaire

Cognitive status was assessed using a combination of three separate measures: the AD8 to assess for any functional impairment which may be caused by neurological disease, the extended instrumental activities of daily living (E-IADL) to probe for lifestyle factors which may be predictive of well-preserved cognitive functioning, and an educational/occupational questionnaire based on the Lifetime Experiences Questionnaire (LEQ). These measures were administered using Qualtrics.

AD8. The AD8 was administered as a screener for any participants who may exhibit early signs of cognitive impairment, ensuring that any age effects observed derived from cognitive changes rather than Alzheimer's pathology. The form is extremely brief (8 items) and probes memory, orientation, judgment, and function, and shows good sensitivity (86%) and specificity (74%) in prior research (Galvin et al., 2005). We utilized a conservative cut-off score of 2 on this test, excluding any participants scoring higher than this.

E-IADL. The extended instrumental activities of daily living (E-IADL) extends the predictive validity of traditional IADL measures by incorporating several items related to leisure activities. While traditional IADL measures are designed to detect cognitive risk factors based on physical or cognitive impairments, the E-IADL is better equipped to detect risk factors based on activities corresponding to many of the intellectually stimulating exposure variables associated with CR (e.g. "gone to classes of any kind", "club or center activities", etc.). Prior investigation into this scale demonstrated both a severe decrease in ceiling effects (from 67% to 3%) over the traditional IADL, and good predictive utility (Fieo et al, 2013). The sum of all items representing risk factors (e.g. endorsing difficulties performing chores without assistance) were subtracted from the sum of all items representing stimulating leisure activities (e.g. going to classes of some kind). All scores were individually standardized within age group (since older adults displayed higher levels in all of these measures as a function of more life experience), before being averaged for each participant. This measure was also used to screen out subjects at a high risk for cognitive impairment.

LEQ Items. Educational and Occupational sections of the Lifetime Experiences Questionnaire (LEQ) were administered (Valenzuela & Sachdev, 2006). This questionnaire has been shown to produce a dominant factor with high loadings on education, occupation, and leisure activity, has strong correlations with tests of cognitive ability, and good predictive validity towards cognitive decline when applied to a sample of 18-month longitudinal data (ibid). It additionally allows the measurement of education and occupation values over a respondent's lifetime rather than statically.

Using the LEQ items, we computed a composite CR-proxy score designed to quantify the level of complex cognition participants had engaged in throughout their lifetimes. For education, total years endorsed of primary and secondary education were totaled, and to this we added the higher-weighted years of post-secondary education endorsed, weighted by 1) percent of coursework completed and 2) enrollment status (half-time was weighted by .5). To this was added the years of professional or graduate education, weighted by a factor of 1.5 to reflect the additional cognitive demands of this track, and additionally weighted by percent completed and enrollment status. For occupation, each subject's reported occupations were matched to the closest available occupation in the Dictionary of Occupational Titles (DOT), which provides complexity scores along the dimensions of work with data (ranging from "comparing" to "synthesizing"), people (ranging from "taking instructions" to "mentoring"), and things (ranging from "handling" to "setting up"). We combined these scores to avoid multicollinearity among the predictors.

Revised STRATA

Based in part on our findings from Study 1, the STRATA task was redesigned to more precisely match the structure of a traditional task-switching paradigm. This version of the task was programmed using the jsPsych Javascript library, which allowed it to be directly accessed via a web browser just as the PHP version had been. Most importantly, stimuli in the mixed blocks were reordered to follow an alternating-runs structure, wherein a task(strategy)-switch is induced after every two successive trials; i.e. trial order became: A,A,B,B,A,A,B…(Monsell, 2003).The purpose of this was twofold: first, it allowed us to maximize the distribution of the two strategy types, and second, it provided participants with a more predictable pattern of strategy switches. The task was also altered to allow for a completely randomized presentation of block orders, so that participants would not be affected by exposure to one type of strategy before the other. Finally, the pure blocks were redesigned to include two distinct rule sets within each particular strategy which also corresponded to an alternating-runs structure. Thus the pure logico-analytic block, although still containing only trials falling under the logico-analytic strategy, now consisted of both "OR" and "XOR" problems presented in a predictable sequence [A,A,B,B…]. Likewise, the visuospatial block was revised to present alternating runs of "shape + orientation" rule trials, along with "size + number" rule trials. The mixed blocks meanwhile used only one rule type of each strategy ("OR" trials for logico-analytic, and "shape + orientation" trials for visuospatial). This was done to test for the effects of a smaller, rule-specific switch cost, which we hypothesized would be present in both pure strategy blocks but that would be lesser in magnitude than the between-strategy switch costs. As before, both RT and accuracy were recorded for each item.

Procedures

Participants were recruited via Mechanical Turk and linked directly to the URL for the study, taking them to a Qualtrics survey where they viewed an informed consent form before completing the Cognitive Status questionnaire. Following the questionnaire, participants were given a link to the task. As in Study 1, they first saw the instructions for the task. They were instructed to complete each item by finding the pattern in the matrix, and to select the answer choice that best completes the pattern. They were then given 4 sample trials, with two examples of each item type (visuospatial and logico-analytic). These practice trials were further divided into two varieties of each strategy–an example “OR” problem and an example “XOR” problem for the logico-analytic trials, and an example “shape + orientation transformation” problem and an example “size + number transformation” problem for the visuospatial trials. For each self-paced practice item, detailed feedback about the series of necessary steps to solve the item were shown to participants after each response, regardless of their initial accuracy. This was done to ensure that participants had explicit knowledge of the types of strategies available to them. Following the practice, participants were instructed that they would have a 30-second time limit on each problem so they had to work as quickly as possible on each problem without sacrificing accuracy. They indicated when they were ready to commence the task with a press of the space-bar. Between each trial, a fixation cross was displayed in the center of the screen for the 500ms ITI. Between each block, participants were presented a screen telling them that they had 1 minute to rest, and told that they could rest for this time, or press a key to continue with the next block. This was done to reduce cognitive fatigue as the test progressed. Following the last block, participants were thanked for their participation and their study compensation was automatically released.

Data Analysis

As in study 1, behavioral analyses were conducted on global, local, and mixing costs in both RT and ACC. Additionally, analyses were conducted on bin scores representing both accuracy and reaction time. This was done both to obviate the possibility of a speed-accuracy trade-off, and to ensure the expected psychometric associations with covariates would not be confounded by individual differences in one of the dependent measures. It also allowed us to retain the full set of responses in our analyses.

Binning Procedure. In addition to investigating reaction times and error rates, we also computed a single combined measure designed to incorporate both of these components. Such metrics are useful whenever a task may be subject to a speed accuracy trade-off, and though study one did not indicate that this was the case on a group-wide level there remains the possibility that some individual subjects may prioritize responding accurately while others may prioritize responding quickly. The binning procedure was conducted as follows: first, mean RTs from accurate non-switch trials were calculated and subtracted from each subject's RT for accurate switch trials. This results in a score for each accurate switch trial that represents the speed in that trial relative to the subject's average non-switch speed. These scores are combined for all subjects, then rank ordered into deciles; the fastest decile is assigned a score of one, the second fastest is assigned a score of two, and so on. Thus each accurate switch trial receives a bin score ranging from 1 (fastest) to 10 (slowest), and then inaccurate switch trials were given a bin score of 20. This value is arbitrary, but it's key property is that it is sufficiently larger than the slowest accurate responses to penalize inaccurate responses; moreover variations in this number have been shown to only very mildly change the strength of resulting correlations. Finally, individual bin scores for each participant were summed to obtain a single bin score.

To analyze the effects of covariates on the dependent measures, we constructed two separate general linear models (GLM) that were analyzed in stages (heterogeneous slopes)

(Kumar et al., 2008; Siegel, 1956). In each model we introduced the covariates of interest, in this case gender and a composite CR score based on the Cognitive Status questionnaire. The initial, heterogeneous-slopes model used a repeated measures analysis of variance (ANOVA) design with strategy type as a within-subjects factor (visuospatial versus logico-analytic problems), and added the main effects of the covariates in addition to their respective interaction with age group. The latter effects tested the assumption of the analysis of covariance that the effects of the covariates are equivalent at each level of the model's fixed effects, and only these equivalent effects were further inspected. In the first stage, we constructed a full model with the following predictors: age group, CR, gender, and the interaction of each covariate by age group. After performing this full model, retaining strategy-type as a within-subjects factor, we constructed a reduced model which retained only the covariate main effects, as well as any interaction terms which were statistically significant in the full model ($p < .05$).

The switch cost models, in contrast, used a 2 x 2 repeated-measures ANOVA, with one within-subjects factor (either block-type or switch-type) and one between-subjects factor (age group). The level of significance was set at .05.

A separate set of 2 (age group) x 2 (strat type) x 2 (rule type) mixed ANOVA models were conducted to analyze the effect of rule switching within any particular strategy, but these failed to reach significance and are not reported here.

Finally, to investigate the effects of CR proxies on the dependent measures within the context of repeated measures, a series of mixed effect linear models were constructed. These models allow the specification of both random slopes (e.g. random effects from subjects or across items) and fixed slopes (e.g. stable effects based on the condition of interest, in this case strategy type and age group.) These models were constructed using CR and other covariates and without them then compared to determine if CR is useful in accounting for variance in the dependent measure.

For bin scores, the initial model constructed consisted of fixed effects, plus random slopes accounting for the possible random differences between subjects and individual items in their susceptibility to strategy types:

> Model 1 = lmer(Dependent Measure ~ strategy-type + age-group + (1 + strategy-type|subject) + (1 + strategy-type|item)

Successive models first added CR as a covariate, then as an interaction term with age group, then as both a covariate and interaction term:

> Model 2 = lmer(Dependent Measure ~ strategy-type + age-group + CR + (1 + strategy-type|subject) + (1 + strategy-type|item) Model 3 = lmer(Dependent Measure ~ strategy-type + (age-group * CR) + (1 + strategy-type|subject) + (1 + strategy-type|item) Model 4 = lmer(Dependent Measure ~ strategy-type + age-group + CR + (age-group * CR) + (1 + strategy-type|subject) + (1 + strategy-type|item)

These models were compared using AIC and BIC scores to determine if the addition of CR to the fixed or random effects contributed to the explanatory power of the model in a parsimonious way.

RT models were constructed in the same manner, except (i) all trials were included (whereas bin scores are only computed for switch trials), and as a result (ii) CR level–low versus high along a median split of the data–was used rather than CR and was used as a random effect per subject, since the degrees of freedom in these models was greater. All other modeling steps and comparison metrics were conducted the same.

Results

Demographic information for participants are shown in table 14.

Table 14

Study 2 Participant Characteristics

	Young	Old
Age	34.26(7.80)	61.5(3.12)
N	45	33
% Female	29.18	52.87
Education	14 (5.6)	18.4 (6.32)

Note. Education is the simple summation of reported years with standard deviations.

Global (Block-Level) Switch Costs

To analyze the overall impact of age and strategy type on switch costs, we replicated the analytic strategy of study 1 by conducting two mixed design ANOVAs with age group as the between-subjects factor and block type as the within-subjects factor.

RT. Analysis of global differences to reaction time across all blocks as a function of age group and type of block resulted in a significant main effect of block type, $F = 1186.551$, $p < .05$. Age group also displayed a significant main effect ($F = 99.837$, $p < .05$). Additionally, the interaction of age group with block type was significant in this model ($F = 14.684$, $p < .05$). Results of this ANOVA are reported in table 16.

Table 16

ANOVA results: Study 2 global switch costs to reaction time

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	76	35840414176.85	306934674.41	8874.43	.000	.99
age_group	1	76	403204794.15	306934674.41	99.84	.000	.55
block_type	1	76	279799785.15	17921506.64	1186.55	.000	.46
age_group x block_type	1	76	3462747.99	17921506.64	14.68	.000	.01

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for Global RT Switch Costs. At the block-level, post-hoc analyses revealed significantly higher RTs for older adults relative to younger adults, regardless of condition. The lone exception to this was the comparison of young RTs in mixed blocks with old RTs in pure blocks, where no significant difference was observed. Additionally, both young and olds in the mixed blocks displayed higher RTs than their corresponding RTs in the pure blocks. See table 17 for these results.

Table 17

Post Hoc Test: Global RT Differences (Mixed Blocks Minus Pure Blocks)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	2952.8	332.1	8.890	<2e-16 ***
Young, Mixed Blocks - Young, Pure Blocks	2407.3	102.4	23.512	<2e-16 ***
Old, Mixed Blocks - Young, Pure Blocks	5964.0	333.5	17.884	<2e-16 ***
Young, Mixed Blocks - Old, Pure Blocks	-545.5	333.1	-1.637	0.609
Old, Mixed Blocks - Old, Pure Blocks	3011.2	119.9	25.115	<2e-16 ***
Old, Mixed Blocks - Young, Mixed Blocks	3556.7	334.5	10.633	<2e-16 ***

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using Bonferroni method.

ACC. Analysis of global differences to mean proportion based on age group and type of block were less successful than the RT analyses; none of the main effects (age group or block type), nor the interaction term showed significance. An ANOVA table of these results can be seen in table 18.

Table 18

ANOVA results: Study 2 global switch costs to accuracy

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	76	103.89	0.45	17739.65	.000	.99
age_group	1	76	0.01	0.45	2.55	.115	.02
block_type	1	76	0.00	0.29	0.65	.424	.00
age_group x block_type	1	76	0.01	0.29	2.06	.156	.01

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for Global ACC Switch Costs. None of the group-level comparisons were significant. See table 19 for these results.

Table 19

Post Hoc Test: Global ACC Differences (Mixed Blocks Minus Pure Blocks)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	-0.034225	0.013782	-2.483	0.0781 .
Young, Mixed Blocks - Young, Pure Blocks	-0.022538	0.013018	-1.731	0.5004
Old, Mixed Blocks - Young, Pure Blocks	-0.028506	0.015973	-1.785	0.4459
Young, Mixed Blocks - Old, Pure Blocks	0.011687	0.015380	0.760	1.0000
Old, Mixed Blocks - Old, Pure Blocks	0.005719	0.015255	0.375	1.0000
Old, Mixed Blocks - Young, Mixed Blocks	-0.005968	0.017370	-0.344	1.0000

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Local Switch-Cost Analyses

Mixed ANOVAs were conducted controlling for the impact of block type on both reaction time and accuracy to analyze local switch costs. Again, a 2 (age group: young or old) by 2 (trial type: switch or nonswitch) design was utilized exclusively within the mixed blocks.

Local Switch Costs: RT. Local switch-costs to RT within mixed blocks were significantly associated with the main effect of trial type (switch or nonswitch), the main effect of age group, as well as the interaction of age group with trial type. Results of this ANOVA are given in table 20.

Table 20

ANOVA results: Study 2 local switch costs to reaction time

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	76	40465730554.00	379352559.36	8106.96	.000	.99
age_group	1	76	486346982.90	379352559.36	97.44	.000	.55
switch_type	1	76	555823707.10	26426019.19	1598.52	.000	.58
age_group x switch_type	1	76	5637706.76	26426019.19	16.21	.000	.01

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for Local RT Switch Costs. Older adults displayed higher RTs in switch trials relative to nonswitch trials, as did young adults. Additionally, older adults displayed higher RTs than younger adults, regardless of trial type combinations with one exception; again young adults in switch trials only slightly outperformed older adults in nonswitch trials, but this did not reach significance. See table 20 for these results.

Table 21

Post Hoc Test: Local RT Differences (Switch Trials Minus Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	3183.75	372.64	8.544	<2e-16 ***
Young, Mixed Blocks - Young, Pure Blocks	3430.71	42.11	81.469	<2e-16 ***
Old, Mixed Blocks - Young, Pure Blocks	7375.07	372.03	19.824	<2e-16 ***
Young, Mixed Blocks - Old, Pure Blocks	246.96	372.08	0.664	1
Old, Mixed Blocks - Old, Pure Blocks	4191.32	49.19	85.213	<2e-16 ***
Old, Mixed Blocks - Young, Mixed Blocks	3944.35	371.47	10.618	<2e-16 ***

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Local Switch Costs: ACC. Local switch-costs to ACC within mixed blocks were again assessed using mixed design ANOVA, and as with the analysis of global switch costs did not result in any significant main effects or interaction effects. Post-hoc analyses for local ACC switch costs confirmed that none of the group-level comparisons were significant.

Mixing Cost Analyses

Mixing costs were assessed using a 2 (age group: young or old) by 2 (block type: mixed or pure) mixed design conducted exclusively among nonswitch trials to control for the trial type factor.

Mixing Costs: RT. Results of the ANOVA of RT values of nonswitch trials revealed a significant main effect of age group (F = 103.072, p < .05), as well as block type (F = 54.893, p <.05), in addition to the interaction term (age group x block type), F = 4.685, p <.05. Detailed results are reported in table 22.

Table 22

ANOVA results: Mixing costs to reaction time

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	76	30655365327.99	264754834.28	8799.87	.000	.99
age_group	1	76	359064684.95	264754834.28	103.07	.000	.57
block_type	1	76	6241535.57	8641416.21	54.89	.000	.02
age_group x block_type	1	76	532736.85	8641416.21	4.69	.034	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for Mixing Costs in RT. In all group by group comparisons, older adults displayed higher mixing costs than young adults, and within-group RTs were always higher in mixed blocks than pure blocks. See table 23 for these results.

Table 23

Post Hoc Test: Mixed RT Differences (Pure Block Nonswitch Trials Minus Mixed Block Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	3322.91	405.87	8.187	1.33e-15 ***
Young, Mixed Blocks - Young, Pure Blocks	265.26	30.28	8.760	< 2e-16 ***
Old, Mixed Blocks - Young, Pure Blocks	3851.26	405.59	9.495	< 2e-16 ***
Young, Mixed Blocks - Old, Pure Blocks	-3057.65	405.67	-7.537	2.88e-13 ***
Old, Mixed Blocks - Old, Pure Blocks	528.35	34.94	15.123	< 2e-16 ***
Old, Mixed Blocks - Young, Mixed Blocks	3586.00	405.40	8.846	< 2e-16 ***

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Mixing Costs: ACC. ANOVA results for accuracy mixing costs conformed to those of the local switch costs: no significant main effects or interactions were observed. See table 24 for these results. Post-hoc analyses for mixing costs in ACC confirmed that none of the group-level comparisons were significant. See table 25 for these results.

Table 24

ANOVA results: Mixing costs in accuracy

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	76	106.68	0.66	12333.47	.000	.99
age_group	1	76	0.00	0.66	0.01	.921	.00
block_type	1	76	0.01	0.52	1.09	.300	.01
age_group x block_type	1	76	0.04	0.52	5.99	.017	.03

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Table 25

Post Hoc Test: Mixed ACC Differences (Pure Block Nonswitch Trials Minus Mixed Block Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	-0.035061	0.014033	-2.499	0.0748 .
Young, Mixed Blocks - Young, Pure Blocks	-0.022433	0.013200	-1.699	0.5354
Old, Mixed Blocks - Young, Pure Blocks	-0.028224	0.016060	-1.757	0.4731
Young, Mixed Blocks - Old, Pure Blocks	0.012628	0.015507	0.814	1.0000
Old, Mixed Blocks - Old, Pure Blocks	0.006837	0.015465	0.442	1.0000
Old, Mixed Blocks - Young, Mixed Blocks	-0.005792	0.017363	-0.334	1.0000

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Figures 6-9 display the mean RT differences based on age-group, block type, rule type, and strategy type.

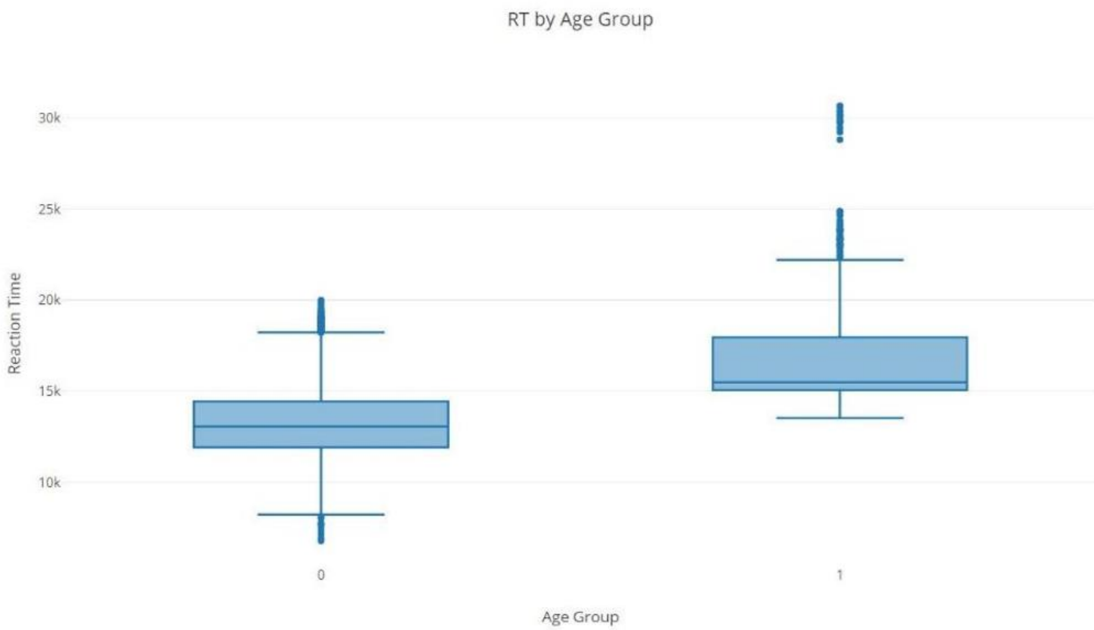

Figure 6. RT by Age Group. Mean RT in milliseconds, 0 = youngs, 1 =olds.

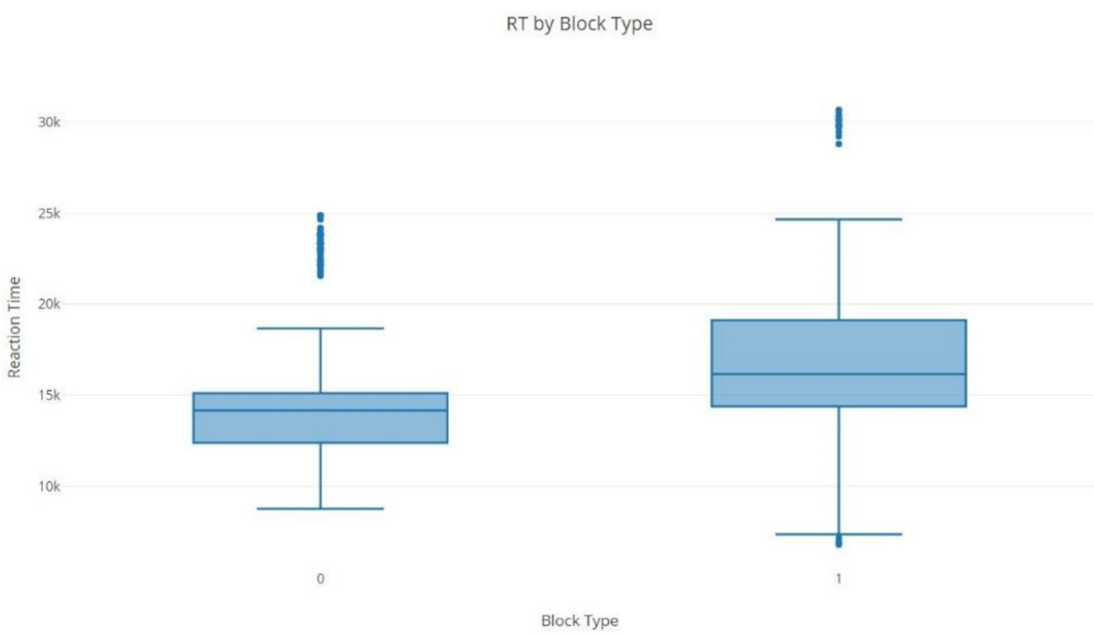

Figure 7. RT by Block Type. Mean RT in milliseconds. 0 = pure blocks, 1 = mixed blocks.

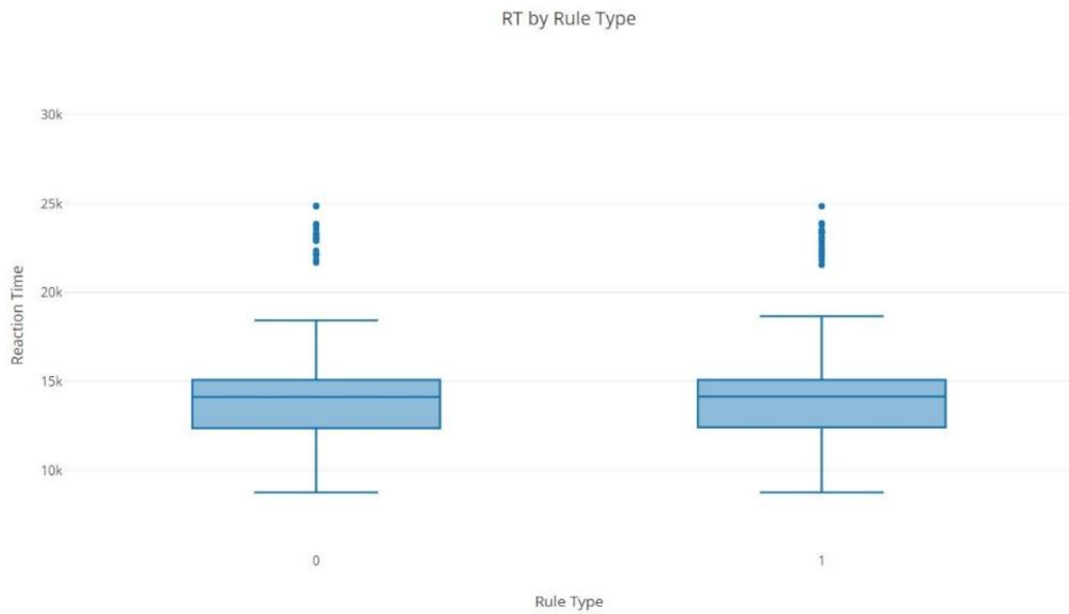

Figure 8. RT by Rule Type. Mean RT in milliseconds, majority rule type is labeled as â�œ0â�� for both strategy conditions, alternative rule type labeled as â�œ1â�� across both conditions. Note these are collapsed within-strategy as the interaction of strategy type with rule type had no effect on RT.

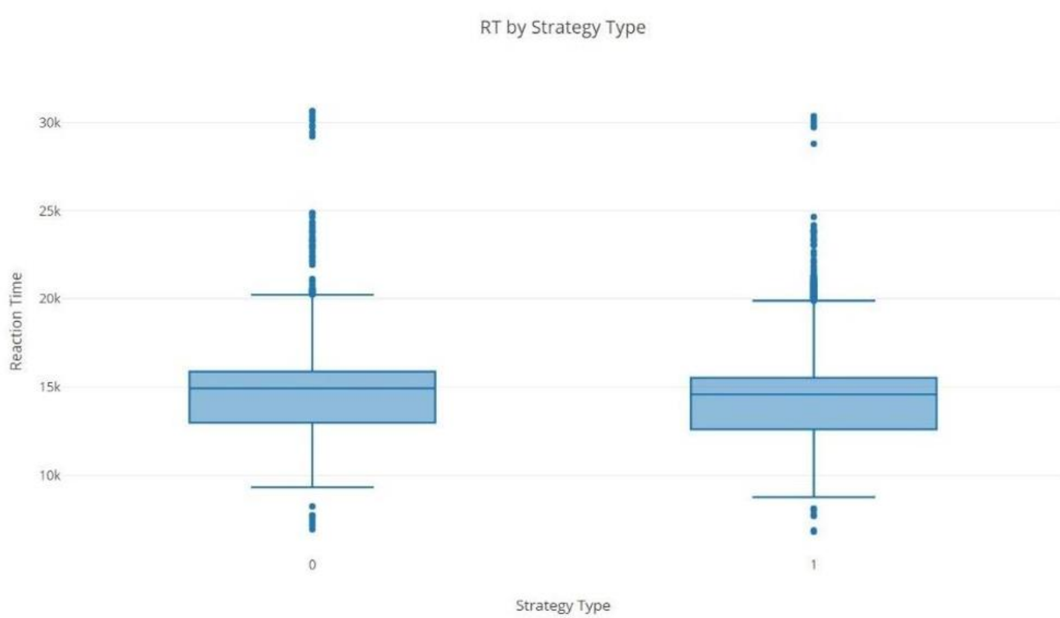

Figure 9. RT by Strategy Type. Mean RT in milliseconds, 0 = logico-analytic strategy, 1 = visuospatial strategy.

Bin Score Analysis. Bin scores for each subject were highly correlated with the CR measure in the old group (-.619) but not in the young group (-.231). This suggests that CR is protecting the strategy-switching flexibility in older adults but is not associated with it in younger adults, consistent with prior studies showing that CR's protective role increases in importance over the lifetime.

Mixed Effects Linear Models

Bin Scores. Model comparisons amongst the four models for bin scores showed that CR did not contribute to improved model fit, as the AIC and BIC values remained lowest for the first model. A liklihood ratio test confirmed that the 'p-value' for these latter models did not reach an estimated significance of .05. These results are reported in tables 26 and 27.

Table 27.

Df	AIC	BIC	loglik	deviance	Chisq	Df	Pr(>Chisq)
mod1	179.77	232.21	-79.886	159.77			
mod2	181.59	239.28	-79.796	159.59	0.1787	1	0.6725
mod3	183.40	246.33	-79.697	159.40	0.1981	1	0.6562
mod4	183.40	246.33	-79.697	159.40	0.0000	0	1.0000

RT. Model comparison results showed that the addition of CR as a fixed effect did decrease the AIC and BIC values, suggesting a contribution to the explanatory power of the model to RT. Moreover, further adjusting the model to include the interaction of CR level with age group further reduced AIC and BIC, albeit slightly, suggesting that treating CR as a moderator of a fixed effect (as it is theorized to be) lends the highest explanatory power. These results could not be confirmed using a likelihood ratio test, but such tests are only rough approximations of p-values for mixed models and so the model comparison approach gives a more reliable interpretation.

	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
lme1	9	-19761	-19700	9889.3	-19779			
lme2	10	-19759	-19692	9889.5	-19779	0.3664	1	0.5450
lme3	11	-19757	-19683	9889.5	-19779	0.0076	1	0.9304

RT. Model comparison results showed that the addition of CR as a fixed effect did decrease the AIC and BIC values, suggesting a contribution to the explanatory power of the model to RT. Moreover, further adjusting the model to include the interaction of CR level with age group further reduced AIC and BIC, albeit slightly, suggesting that treating CR as a moderator of a fixed effect (as it is theorized to be) lends the highest explanatory power. These results could not be confirmed using a likelihood ratio test, but such tests are only rough approximations of p-values for mixed models and so the model comparison approach gives a more reliable interpretation. See table 26.

Discussion

The second study largely replicated the effects of the first, and extended the method of analyzing performance data by combining RT and accuracy into a combined metric. Age was shown to interact with block type (mixed versus pure) such that older adults performed worse in mixed blocks relative to younger adults. Also as expected, our CR measure correlated highly with overall STRATA performance in the old group but not in the young group, consistent with other proxies of CR. An unexpected result was that within-strategy rule shifts did not elicit statistically significant switch costs, but this result provides further evidence that strategy-shifting is a distinct phenomenon with potentially greater magnitudes of influence on overall performance.

The results from the mixed models conducted on bin scores and RT surprisingly yielded no evidence that CR may be contributing to the variation in bin scores, but does suggest it is interacting with age group to modulate RT. One potential explanation for this may be the sheer number of trials analyzed by these models; since bin scores were only conducted for switch trials, each subject contributed only 18 unique trials to these analyses as opposed to the potential 80 in the RT models. Moreover the effect of switch type itself is effectively parsed out in such an analysis, and this may be what is primarily interacting with CR.

While we chose to redesign the STRATA to more closely reflect the alternating runs paradigm within task-switching, the options left unchosen are an area rife for exploration with this task and others like it. An alternative to the alternating runs paradigm is the task-queuing paradigm, wherein tasks are presented in a randomized way but the appropriate

S-R mappings are given in an explicit cue that either precedes or accompanies each trial. Studies utilizing this paradigm have often shown worst performance in the form of RT or error rates in switch trials than a non-switch trials, and sometimes have demonstrated declining RTs over several trials of the same task (Monsell, 2003).

Another potential task switching paradigm is the intermittent instructions cueing, wherein participants complete some series of trials utilizing bivalent stimuli, randomly preceded by explicit instructions cueing the next task to be performed. Not only are local switch-costs present in such a paradigm, but in addition so-called restart costs can also be inferred; these are costs associated with non-switch trials that are explicitly cued versus non-switch trials that were not (Gopher, Armony, Greenshpan 2000). Altmann and Gray (2008) further showed the presence of increasing RTs and error rates as block length increased, a further source of costs they termed within-run slowing.

A more recent task switching paradigm developed is voluntary task selection (Arrington and Logan 2004a, 2005) which relies on subjects using internally generated task switches as opposed to externally cued ones. Here using bivalent stimuli, the participants can choose which set of S-R mappings to apply to any given trial, and even in the absence of exogenously enforced switching switch costs can still be seen.

The STRATA can be modified to suit any of these existing paradigms, and doing so may result in greater switch costs, the emergence of local switch costs, and potentially other avenues to study more strategy dimensions (such as strategy distribution).

CHAPTER 5

STUDY THREE

Introduction

In the final experiment, we aimed to replicate the effects found in Studies 1 and 2 in a sample of participants who had previously been investigated by our lab. The Reference Ability Neural Network (RANN) study was designed to study potential neural underpinnings of four "reference abilities" previously proposed by Salthouse (2009) as accounting for the vast majority of age-related variance in cognitive performance: episodic memory, perceptual speed, fluid reasoning, and vocabulary. Using a battery of 12 tasks (3 from each reference ability) adapted for fMRI imaging, this study's early conclusions suggest that a) these reference abilities can be used to define largely task-invariant activation patterns, b) these activation patterns differ for each reference ability, and c) age changes the typical pattern of expression of such networks (Stern et al., 2014).

While the findings of this study are less immediately relevant than its procedures and participant characteristics, it does serve as an example of just where cognitive strategic decomposition may further come into play; the fluid reasoning reference ability network, for instance, of which an adapted version of the RPM is a measure, may easily show differential expression patterns as a function of strategy use in young versus old, among olds alone, or among youngs.

Some previous research has been conducted examining the relationship between strategy use and neurofunctional patterns. Miller et al. (2012) note that topographical patterns of brain activity during fMRI memory retrieval tasks show vast inter-individual differences (Miller, 2009) yet can be consistent over time. Such differences are often masked in neuroimaging literature by group analyses, which can show brain regions as equally active that may in fact be differentially preferred by subgroups of participants. These authors attempted to induce such differential activation patterns through the use of differential cognitive strategies within a recognition memory task. Previous research (Kirchoff & Buckner, 2006) identified a set of alternative strategies subjects could adopt during the encoding phase of a memory task, finding that strategies such as verbal elaboration were associated with greater prefrontal activity whereas strategies relying more on visual inspection were associated with extrastriatal cortex activation.

These authors also note the potential importance of overall cognitive styles to interpreting group-level differences in activation, with styles being associated with a more general set of strategic preferences over time (for instance, the preference to rely on verbal strategies in general over and above visual strategies). Using a memory retrieval task, these researchers speculated that individual differences in strategy and/or cognitive style would account for activation differences inter-individually beyond anatomical differences or performance differences. Cognitive styles were assessed using a test battery classifying subjects along the visualizer-verbalizer dimension using principal components analysis, while encoding strategy was measured using a strategy questionnaire adopted from Kirchoff and Buckner (2006) as well as explicitly probing subjects to report their chosen strategy. In a set of 50 participants ranging from 18 to 55, cognitive style and strategy was associated with whole brain differences in activation controlling for presentation order, demographics, anatomy, and performance. Interestingly, with advancing age the variability in activation patterns increased. FA maps derived from diffusion tensor imaging were particularly impactful. The greater the magnitude of difference in strategy use and cognitive style was also predictive of the magnitude of differences in activation patterns. FA in the temporal cortex was the strongest anatomical difference in its relation to activation patterns. Overall this study confirmed the expected activation patterns, with the majority of significant differences being found in frontal and parietal regions.

Evidence more directly related to figural reasoning of the type induced by RPM-type items also exists to show differential underpinnings. Prior research within neuroimaging shows reliable prefrontal and posterior parietal cortical activation during fluid intelligence tasks (Masunaga et al., 2008). In a study of the RPM comparing analytic and figural/visuospatial reasoning items, Prabhajaran (1997) found that whereas the former correlated with the right frontal and parietal activations, left hemisphere linguistic and object working memory regions, and left hemispheric regions associated with induction of visuospatial relations, as well as frontal regions link to goal management and executive processes. Moreover these authors note that cortical structures related to G are not invariant with age, but differ in somewhat predictable ways (e.g. demonstrating more diffuse patterns of activation.) They also review research into beginner through advanced expert “go” players, and find that while age continues to negatively correlate with each subgroup of players, the more expert a subgroup is these correlations decrease such that beginner go players show a correlation between topology scores of -.5 whereas professional go players show only a correlation of -.07 (Horn & Masunaga. 2006), indicating that advanced levels of performance and expertise in the game can mitigate age-related deficits. These authors found activation patterns in the frontal and parietal lobes associated with performance of the topology test when compared to control tests. Heyer, Siegel, and Tang (1992) show that after practicing the computer game Tetris, individuals with high Gf demonstrated a reduction in glucose metabolic rates in the brain overall as well as in neural regions associated with Tetris performance, suggesting that the learning associated with expertise can increase neural efficiency in corresponding regions.

In order to begin to probe how strategic flexibility can be used to further explore relationships among age, brain structure, and other metrics of cognitive performance, this study will utilize mediation, moderation, and moderated-mediation models. Previously, Steffener et al. (2013) used moderated-mediation analyses to explore how CR (as measured with verbal IQ and years of education) may modulate the effect of age on MRI-derived brain structure variables, and/or how it may modulate the effect of those variables on cognitive performance in the form of performance scores taken from a lengthy neuropsychological battery. Testing the impact of age on cognitive performance in the domains of perceptual speed, episodic memory, and fluid intelligence/reasoning, as mediated by brain-wide cortical thickness and volume measures, these authors applied models designed to test: a) that CR is reducing the effect of age on brain structure, b) that CR is reducing the effect of brain structure on cognitive performance, c) that CR is reducing the effect of age on brain structure, and the effect of brain structure on cognitive performance, and d) that CR is not moderating any of these relationships.

Results indicated that in one cognitive domain (fluid intelligence/reasoning) and under one of the above models (b above), CR was indeed reducing the effect of age-related cortical thickness and volume declines on cognition in at least 6 brain regions: bilateral putamen, left accumbens, the bank of the right superior sulcus, right middle frontal gyrus, right posterior central gyrus, and right superior temporal gyrus mean thickness. While these regions have some reported relationships to cognitive performance (e.g., Macdonald et al., 2013; Samanez-Larkin et al., 2010; Alexander et al., 2008), what is more significant to the current study's motivation was the potential to investigate how exposure variables like CR can interact with neurostructural variables in the context of such models to infer causal mechanisms. With this in mind, we collected data using the STRATA task from the previous studies in a sample of subjects who have already had extensive neuropsychological and neuroimaging testing performed.

Using similar moderated-mediation models, we hypothesized that if strategic flexibility can be taken as a byproduct of CR, it may function in a similar way to it; i.e., we expected that the STRATA score can be used to successfully moderate the age-related effect of brain structure on cognitive performance when the same brain regions and performance measures are tested. We further expected this relationship because of the closer association between the nature of the two performance measures: one of the fluid reasoning components is a variant of the RPM task, and so we would expect that any metric based on a similar task may be an even stronger moderator.

These predictions derive from the role that cognitive strategies have played as neurofunctional reinforcers within neuroimaging contexts. As reviewed in the introduction, cognitive strategy (and/or cognitive style) might be conceptualized as a multifaceted factor reflecting the repertoire, distribution, selection, execution, and flexibility of specific sets of mental operations that have been acquired either through explicit processes of instruction (education) or implicit sources of learning (occupational complexity, leisure activities).

While each of these subcomponents of cognitive strategy may themselves interact in some way with other interindividual differences (e.g., brain structure, executive functions, motivational factors), they may also collectively serve as a mediator between exposure variables like CR and activation differences during task performance.

Methods

Participants

Participants who had already been studied within the context of the Reference Ability Neural Network (RANN) study were recruited. Forty-four young (20-40) and forty-two old (60-80) participants took part in the study. These participants were initially recruited using market-mailing procedures designed to equalize the recruitment approaches of the two groups. Participants who responded to the mailing were screened over the phone to ensure that they were right-handed, English was their first language, they had no recent history of neurological disorders, were not currently taking psychotropic medications, and their vision was normal or corrected-to-normal. Participants meeting these basic inclusion criteria were then tested in person with an extensive neuropsychological battery. Global cognitive functioning was assessed with the Mattis Dementia Rating Scale, on which a score of at least 130 was required for inclusion in the remaining study procedures (Mattis, 1988). This study was approved by the Internal Review Board of the College of Physicians and Surgeons of Columbia University. Written informed consent was obtained from all participants prior to study participation, and participants were compensated $20 for their participation in the STRATA study.

Measures

Previous factor analyses from our laboratory identified neuropsychological and behavioral measures underlying the construct of CR and three cognitive domains: memory, processing speed/attention, and fluid ability (Siedlecki et al., 2009). Composite scores for each of these measures were created using the mean of the z-transformed measurements. Missing values from any of the measurements were imputed using a simplified version of multivariate imputation based on Principal Components Analysis (PCA) without regard for assumptions such as robustness or the randomness of the missing values. (Missing values are not believed to be due to a participant's unwillingness or inability to complete the test, but rather to time constraints during administration and/or experimenter error.) We estimated the factor scores from the test values present for each participant separately for young and old. Based on this process, we believe that the PC structure is the same for subjects with complete data as those with incomplete data.

Fluid ability. Fluid ability was defined as the composite score comprising the Letter-Number Sequencing, Matrix Reasoning, and Block Design subtests of the WAIS-3 (Wechsler, 1997). Fluid ability refers to the capacity to solve novel problems in tests of abstract reasoning; as discussed above, the Raven;s matrix reasoning tests tend to have the highest loadings on this construct (Raven, 1962). The Letter-Number Sequencing test requires participants to repeat verbally-presented lists of intermixed letters and numbers in alphabetical and numerical order, with list lengths increasing on each subsequent trial. The Matrix Reasoning subtest requires participants to find a pattern in a set of eight possible patterns that best completes the missing cell of a matrix. The Block Design task measures visuospatial ability and requires participants to construct a series of increasingly complex geometrical shapes presented to them in a booklet using 4 or 9 identical blocks that are colored half-red and half-white on either side of their diagonals.

Memory. Memory was defined as the composite score of three sub-scores of the Selective Reminding Task (SRT)–total, delayed recall, and delayed recognition (Buschke, 1974). Participants in this task are instructed to read a list of 12 words and then asked to recall the words over the course of six trials. Following a recall attempt, they are reminded of the words they failed to recall. SRT-total is the total number of recalled words for all trials and has a maximum score of 72. SRT-delayed is the number of correctly recalled words after a 15-minute delay. SRT-delayed recognition is the number of correctly recognized words when each of the 12 words is visually presented alongside three distractor words.

Processing speed/attention. Processing speed/attention is defined as the composite score of performance on the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1997) Digit Symbol subtest, the Trail Making Test (Reitan, 1993), and the Stroop test. The Digit Symbol test presents participants with a key at the top of the form wherein each digit from 0 to 9 is paired with an arbitrary symbol, then requires participants to write down the symbol corresponding to each single-digit in an array of digits as quickly as possible. We used the time to complete the Trails A (numbers only) from the Trail Making Test, which requires connecting in sequential order a series of randomly dispersed numerals on a sheet of paper. Time taken to complete the Stroop Color test, wherein subjects must name the color of ink used to spell an incongruent words (e.g. the word “blue” written in red ink) as quickly as possible, was also used.

Cognitive Reserve. Cognitive reserve was defined as the composite score comprising years of education and scores on two IQ indices: the NART (Nelson, 1978) and WAIS-R vocabulary score (Wechsler, 1997). Previous work from our laboratory has demonstrated the validity of this construct using these cognitive tests (Siedlecki et al., 2009).

Procedures

Image Acquisition Procedure

MRI images were acquired in a 3.0T Philips Achieva Magnet using a standard quadrature head coil. A T1-weighted scout image was acquired to determine subject position. One hundred sixty-five contiguous 1 mm coronal T1-weighted images of the whole brain were acquired for each subject with an MPRAGE sequence using the following parameters: TR 6.5 ms, TE 3 ms; flip angle 8°, acquisition matrix 256x256 and 240 mm field of view. A neuroradiologist reviewed anatomical scans and any with potentially clinically significant findings, such as abnormal neural structure were removed from the sample prior to the current analysis.

Freesurfer Methods

Participants' structural T1 scans were reconstructed using FreeSurfer (Fischl, 2012) (http://surfer.nmr.mgh.harvard.edu/). Subcortical segmentation and cortical parcellation (Fischl et l., 2002; Fischl et al., 2004) accuracy using FreeSurfer has been reported to be comparable to manual labeling. Each subject's white and gray matter boundaries, and gray matter and cerebrospinal fluid boundaries were visually inspected slice by slice by an experienced user, manual control points were added in the case of any visible discrepancy, and reconstruction was repeated until we reached satisfactory results within every subject. The subcortical structure borders were plotted by Freeview visualization tools and compared against the actual brain regions. In case of discrepancy, they were corrected manually. The regions of interest used in this analysis are listed above.

STRATA Administration

Participants took the STRATA task online as detailed in Study 2. After providing informed consent via Qualtrics, they were directed to the task directly in their web browser. They were instructed to take the task on a laptop or desktop computer with a full screen, and to ensure that they could work continuously on the task in a distraction-free environment for at least one hour.

Data Analysis

Behavioral Analysis

Switch Cost Models. Switch-cost models were conducted identically to studies 1 and 2, using a 2 x 2 repeated-measures ANOVA, with one within-subjects factor (either block type or switch type) and one between-subjects factor (age group); these were run for both RT and ACC.

Behavioral analyses were conducted on the bin scores as in Study 2. To analyze the effects of covariates on the dependent measures, we constructed two separate general linear models (GLM) that were analyzed in stages (heterogeneous slopes) (Kumar et al., 2008; Siegel, 1956). In each model we introduced the covariates of interest, including gender, CR composite scores (CR), executive composite scores (EF), processing speed composite scores (PS), fluid reasoning composite scores (FR), and memory composite scores (M). The initial, heterogeneous-slopes model used a repeated measures analysis of variance (ANOVA) design with strategy type as a within-subjects factor (visuospatial versus logico-analytic problems), and added the main effects of the three covariates in addition to their respective interactions with age group. The latter effects tested the assumption of the analysis of covariance that the effects of the covariates are equivalent at each level of the model's fixed effects, and only these equivalent effects were further inspected. In the first stage, we constructed a full model with the following predictors: age group, CR, EF, FR, PS, M, and gender. We also added interaction terms by multiplying the group predictor by each of the covariates, allowing us to test for the presence of group differences in slopes describing the relationship between the covariates and dependent measures. After performing this full model, retaining strategy-type as a within-subjects factor, we constructed a reduced model which retained only the covariate main effects, as well as any interaction terms which were statistically significant in the full model ($p < .05$).

Brain Covariate Analyses

Moderated-Mediation Models. Using statistical path modeling we tested the hypothesis that age-related neural differences affected fluid reasoning ability in a selection of ROIs, and that CR would decrease the effect of age-related declines in these areas on performance. Statistical path model tested each of the four hypotheses of this study. All brain measures of thickness were corrected for mean cortical thickness and brain measures of volume were corrected for normalized brain volume, and corrected for sex. Two regression models were constructed, the first examining the effect of age group on the mediator (brain measures), and second one looking at the effect of age group, bin score, and age group X bin score on fluid reasoning performance. Next, we examined the effect of bin score (the moderator) on the mediation path of brain measure to fluid reasoning score.

This was accomplished by testing the bin score at both low (1 S.D. below the mean) and high (1 S.D. above the mean) values.

Results

Participant characteristics are given in Table 28.

Table 28

Study 3 Participant Characteristics.

	Young	Old
Age	30.23 (5.63)	66.7(4.26)
N	40	40
% Female	42.5	60
Mean CR	-.092(.63)	-5.03(.87)
Mean bin score	.026(.99)	1.19(1)

Note. CR is based on the composite of NART IQ, WAIS Vocab score, plus years of education

Global (Block-Level) Switch Costs

To analyze the overall impact of age and strategy type on switch costs, we replicated the analytic strategy of studies 1 and 2, using two mixed-design ANOVAs with age group as the between-subjects factor and block type as the within-subjects factor.

RT. Analysis of global differences to reaction time across all blocks as a function of age-group and type of block resulted in a significant main effect of block-type, and a significant interaction effect of block-type with age-group. Table 29 displays results from this ANOVA.

Table 29

ANOVA results

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	79	25567129801.21	2713380005.77	744.39	.000	.90
age_group	1	79	57827391.81	2713380005.77	1.68	.198	.02
block_type	1	79	6191802.49	79717530.12	6.14	.015	.00
age_group x block_type	1	79	5705837.32	79717530.12	5.65	.020	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for Global RT Switch Costs. The only significant group difference was found in the RT difference between older adults in the pure blocks and older adults in the mixed blocks, with the latter displaying higher RTs. See table 30 for these results.

Table 30

Post Hoc Test: Global RT Differences (Mixed Blocks Minus Pure Blocks)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	833.996	935.087	0.892	1.00000
Young, Mixed Blocks - Young, Pure Blocks	3.988	233.493	0.017	1.00000
Old, Mixed Blocks - Young, Pure Blocks	1568.475	934.344	1.679	0.55927
Young, Mixed Blocks - Old, Pure Blocks	-830.009	935.208	-0.888	1.00000
Old, Mixed Blocks - Old, Pure Blocks	734.479	231.699	3.170	0.00915 **
Old, Mixed Blocks - Young, Mixed Blocks	1564.488	934.465	1.674	0.56454

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

ACC. Analysis of global differences to mean proportion accuracy did not reveal any main effects of block type or age group. And while the interaction between age group and block type also failed to reach significance, it was trending in that direction (.078). An ANOVA table of these results can be seen in table 31. Post-hoc analyses for global ACC switch costs confirmed that none of the group-level comparisons were significant. See table 32 for these results.

Table 31

ANOVA results

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	79	102.43	2.58	3137.53	.000	.97
age_group	1	79	0.00	2.58	0.13	.715	.00
block_type	1	79	0.01	0.45	2.12	.150	.00
age_group x block_type	1	79	0.02	0.45	3.18	.078	.01

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Table 32

Post Hoc Test: Global ACC Differences (Mixed Blocks Minus Pure Blocks)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	-0.031458	0.030792	-1.022	1.0000
Young, Mixed Blocks - Young, Pure Blocks	-0.003532	0.016808	-0.210	1.0000
Old, Mixed Blocks - Young, Pure Blocks	0.007424	0.030783	0.241	1.0000
Young, Mixed Blocks - Old, Pure Blocks	0.027926	0.030795	0.907	1.0000
Old, Mixed Blocks - Old, Pure Blocks	0.038882	0.016608	2.341	0.115
Old, Mixed Blocks - Young, Mixed Blocks	0.010956	0.030786	0.356	1.0000

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Local Switch Costs

Mixed ANOVAs were conducted controlling for the impact of block-type on both reaction time and accuracy to analyze local switch costs. Again, a 2 (age group: young or old) by 2 (trial type: switch or nonswitch) design was utilized exclusively within the mixed blocks.

Local Switch Costs: RT. Local switch-costs to RT within mixed blocks were significantly associated with the main effect of trial-type (switch or nonswitch), as well as the interaction of age-group with trial-type. Results of this ANOVA are given in table 33.

Table 33

ANOVA results

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	79	26630972320.98	2772563319.57	758.81	.000	.90
age_group	1	79	111888507.54	2772563319.57	3.19	.078	.04
switch_type	1	79	107551946.27	173291729.45	49.03	.000	.04
age_group x switch_type	1	79	24305622.65	173291729.45	11.08	.001	.01

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for Local RT Switch Costs. Older adults displayed significantly higher RTs in switch trials relative to their own performance on nonswitch trials, and relative to young adults' performance on nonswitch trials. Younger adults showed a significant difference in RTs between switch and nonswitch trials as well. See table 34 for these results.

Table 34

Post Hoc Test: Local RT Differences (Switch Trials Minus Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	854.22	924.58	0.924	1.00000
Young, Mixed Blocks - Young, Pure Blocks	764.92	259.92	2.943	0.01951 *
Old, Mixed Blocks - Young, Pure Blocks	3125.31	943.30	3.313	0.00554 **
Young, Mixed Blocks - Old, Pure Blocks	-89.31	944.61	-0.095	1.00000
Old, Mixed Blocks - Old, Pure Blocks	2271.09	254.92	8.909	<2e-16 ***
Old, Mixed Blocks - Young, Mixed Blocks	2360.40	962.95	2.451	0.08542 .

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Local Switch Costs: ACC. Local switch-costs to ACC within mixed blocks were again assessed using mixed design ANOVA, and as with the analysis of global switch costs did not result in any significant main effects or interaction effects. See table 35 for the ANOVA results. Post-hoc analyses showed again that none of the group-level comparisons were significant–see table 36 for these results.

Table 35

ANOVA results

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	79	104.49	2.93	2817.61	.000	.97
age_group	1	79	0.01	2.93	0.14	.705	.00
switch_type	1	79	0.00	0.44	0.55	.461	.00
age_group x switch_type	1	79	0.00	0.44	0.36	.551	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Table 36

Post Hoc Test: Local ACC Differences (Switch Trials Minus Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	-0.0170465	0.0289691	-0.588	1
Young, Mixed Blocks - Young, Pure Blocks	-0.0138757	0.0173981	-0.798	1
Old, Mixed Blocks - Young, Pure Blocks	-0.0007385	0.0316388	-0.023	1
Young, Mixed Blocks - Old, Pure Blocks	0.0031708	0.0316999	0.100	1
Old, Mixed Blocks - Old, Pure Blocks	0.0163080	0.0171841	0.949	1
Old, Mixed Blocks - Young, Mixed Blocks	0.0131372	0.0341568	0.385	1

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Mixing Costs

Mixing costs were assessed using a 2 (age-group: young or old) by 2 (block-type: mixed or pure) mixed design conducted exclusively among nonswitch trials to control for the trial-type factor.

Mixing Costs: RT. Results of the ANOVA of RT values of nonswitch trials revealed a significant main effect of block-type. Age-group, and its interaction with block-type, did not reach significance. Detailed results are reported in table 37. None of the post-hoc group comparisons yielded a statistically significant difference–see table 38 for values.

Table 37

ANOVA results

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	79	24060380503.13	2719944078.75	698.83	.000	.90
age_group	1	79	29503397.70	2719944078.75	0.86	.357	.01
block_type	1	79	5266224.96	89644577.75	4.64	.034	.00
age_group x block_type	1	79	46638.76	89644577.75	0.04	.840	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Table 38

Post Hoc Test: Mixed RT Differences (Pure Block Nonswitch Trials Minus Mixed Block Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	1225.6	943.5	1.299	1.000
Young, Mixed Blocks - Young, Pure Blocks	-238.7	230.4	-1.036	1.000
Old, Mixed Blocks - Young, Pure Blocks	1377.4	937.0	1.470	0.850
Young, Mixed Blocks - Old, Pure Blocks	-1464.3	937.7	-1.562	0.710
Old, Mixed Blocks - Old, Pure Blocks	151.8	228.8	0.663	1.000
Old, Mixed Blocks - Young, Mixed Blocks	1616.1	931.2	1.735	0.496

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Mixing Costs: ACC. ANOVA results for accuracy mixing costs were again negative: no significant main effects or interactions were observed. See table 39 for these results.

Table 39

ANOVA results

Predictor	df_{Num}	df_{Den}	SS_{Num}	SS_{Num}	F	p	η^2_g
(Intercept)	1	79	102.90	2.61	3111.90	.000	.97
age_group	1	79	0.01	2.61	0.22	.637	.00
block_type	1	79	0.02	0.50	2.81	.098	.01
age_group x block_type	1	79	0.01	0.50	2.07	.154	.00

Note. df_{Num} indicates degrees of freedom numerator. df_{Den} indicates degrees of freedom denominator. SS_{Num} indicates sum of squares numerator. SS_{Den} indicates sum of squares denominator. η^2_g indicates generalized eta-squared.

Post-Hoc Analyses for Mixing Costs in ACC. Despite the nonsignificant findings of the overall ANOVA, post-hoc comparisons displayed statistically significant higher accuracy for young adults in the pure blocks relative to older adults in the pure blocks, young adults in the pure blocks relative to older adults in the mixed blocks, and young adults in the mixed blocks relative to older adults in either block type. Results are given in table 40.

Table 40

Post Hoc Test: Mixed ACC Differences (Pure Block Nonswitch Trials Minus Mixed Block Nonswitch Trials)

Factor Contrasts	*Estimate*	*St. Error*	z value	*Pr(>|z|)*
Old, Pure Blocks - Young, Pure Blocks	-0.06713	0.01544	-4.349	8.20e-05 ***
Young, Mixed Blocks - Young, Pure Blocks	0.03264	0.01642	1.988	0.281
Old, Mixed Blocks - Young, Pure Blocks	-0.09651	0.01807	-5.340	5.58e-07 ***
Young, Mixed Blocks - Old, Pure Blocks	0.09977	0.01768	5.643	1.00e-07 ***
Old, Mixed Blocks - Old, Pure Blocks	-0.02938	0.01793	-1.639	0.608
Old, Mixed Blocks - Young, Mixed Blocks	-0.12915	0.02003	-6.449	6.75e-10 ***

Note. Simultaneous Tests for General Linear Hypotheses. Multiple Comparisons of Means: Tukey Contrasts. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1. Adjusted p values reported using bonferroni method.

Cortical Thickness - Neuropsychological Patterns

Cortical Thickness Expression Scores. Using the MRI-derived cortical thickness of various brain regions, past studies (RANN) in our lab have been capable of finding "expression scores" of these patterns representing the topological network most highly-predictive of performance in a variety of neuropsychological abilities, including fluid reasoning, memory, processing speed, and vocabulary.

To test if these patterns interact with the cognitive flexibility measure (bin score), these patterns for 68 subjects were entered into a moderation model with age as the independent variable, expression score for either memory or fluid reasoning as the outcome variable, and bin score as the moderator. In neither case was a successful interaction observed between age and bin score, possibly reflective of an age-related systematic switch in preference for accuracy over speed as observed in study 1. The t-values for these moderation analyses are reported in table 41.

Moderation and Moderated-Mediation Models

Moderated-Mediation Results. Analyses designed to replicate those of Steffener et al. (2014) using the strategic flexibility measure did not yield any significant findings. Although within this sample two of the regions previously tested were still significant mediators of the relationship between age and fluid reasoning performance, moderation from bin scores was not observed.

Moderation Models. Simple moderation models were conducted to test whether the relationship between brain volumes/thicknesses and fluid reasoning performance could be modified by differing levels of the bin score, which would suggest again that performance on the STRATA behaves in a similar fashion to CR by allowing subjects to take advantage of existing neural networks/volume after the start of age-related decline. To rule out the contradictory impact of age on reasoning performance, these analyses were conducted only in the older group.

The brain measures tested included all seven regions previously identified in being successfully moderated by CR, as well as orbitofrontal cortex volumes due to their behavioral links to task-switching and cognitive strategies, and total gray matter volume. None of these moderation models attained the level of prespecified significance. However the model for total gray matter volume did suggest trends consistent with our hypotheses. See figure 10.

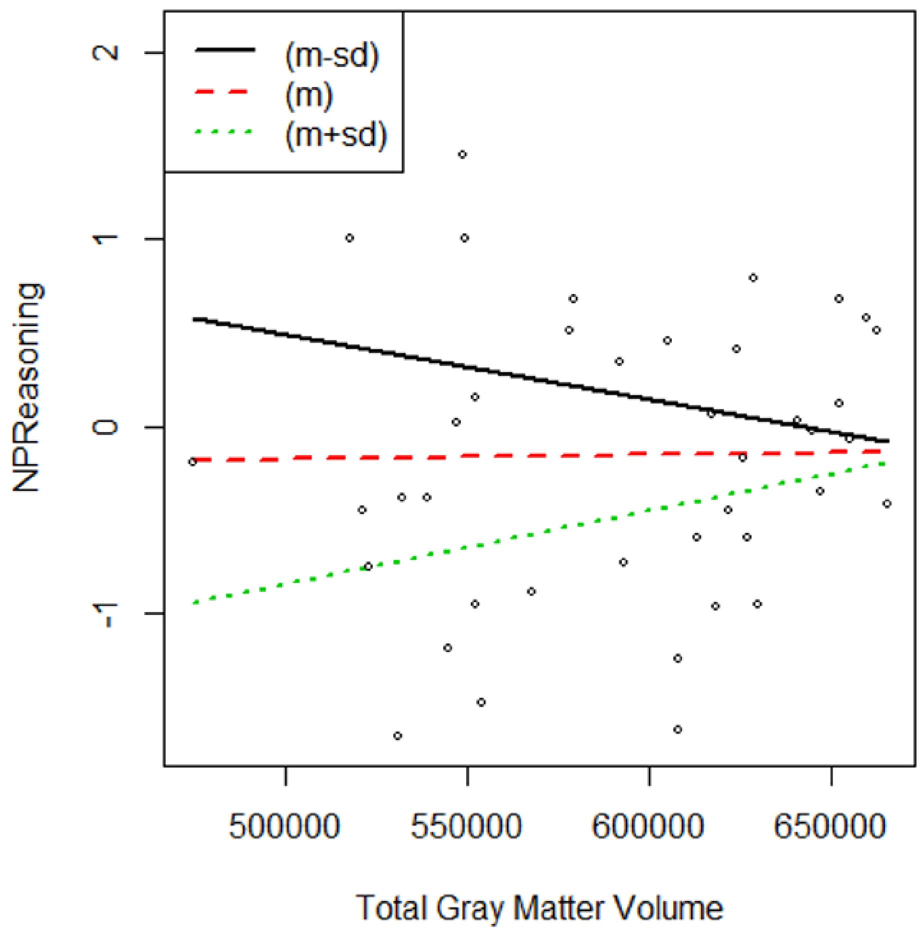

Figure 10. Effect of total gray matter volume on reasoning, moderated by bin scores.

Greater gray matter volume (GMV) was predictive of higher fluid reasoning scores overall, but this trend varied depending on the bin scores of participants; subjects with lower bin scores (meaning better strategic performance) displayed increased performance at lower levels of GMV relative to those with higher bin scores (worse strategic performance). At the highest levels of GMV these trends converge, suggesting that strategic advantage can play a compensatory role up to a point but that it may not confer advantage to those already blessed with high levels of intrinsic brain volume and/or high levels of brain maintenance.

While neither the effect of bin score on fluid reasoning or the interaction with GMV reached significance (p = .07 and .11, respectively) the trend is still suggestive.

To rule out the possibility that strategic flexibility is better captured within the theoretical spectrum of brain maintenance rather than cognitive reserve (e.g., somehow its practice leads to preserved levels of neural integrity/volume/thickness, rather than functioning as a compensatory mechanism once the brain has begun to deteriorate), we repeated these same moderation analyses using age as the independent variable and the regional metrics as dependent variables, with bin scores as moderators. Predictably, none of these models yielded significant or close-to-significant results.

Exploratory Analyses

Regression Analyses. To analyze the effect of covariates on participants' bin scores, we constructed a multiple regression model with age, CR composite, gender, fluid reasoning score, processing speed/attention, memory, executive function composite, as well as total gray matter volume as predictors. This regression model was highly significant ($p<.001$), with CR and fluid reasoning being highly significant predictors accounting for large degrees of variability in the bin score. The coefficients for this model are displayed in table 42.

Table 42

Regression results using bin score as the criterion

Predictor	*b*	*b* 95% CI [LL, UL]	*beta*	*beta* 95% CI [LL, UL]	sr^2	sr^2 95% CI [LL, UL]	*r*	Fit
(Intercept)	-0.24	[-3.98, 3.51]						
CR_Z	0.91**	[0.52, 1.30]	0.61	[0.35, 0.87]	.25	[.08, .42]	.57**	
Age	-0.01	[-0.03, 0.01]	-0.14	[-0.52, 0.25]	.01	[-.02, .03]	-.00	
EXEC_Z	-0.26	[-0.67, 0.14]	-0.16	[-0.41, 0.09]	.02	[-.03, .07]	.02	
NPReasoning	0.20	[-0.23, 0.64]	0.16	[-0.18, 0.51]	.01	[-.03, .05]	.27*	
NPMemory	-0.02	[-0.37, 0.34]	-0.01	[-0.29, 0.26]	.00	[-.00, .00]	.03	
NPSpeed_attention	-0.37	[-0.76, 0.02]	-0.32	[-0.65, 0.02]	.04	[-.04, .12]	-.03	
Gender	0.16	[-0.38, 0.70]	0.08	[-0.18, 0.33]	.00	[-.02, .03]	.17	
TotalGrayVol	0.00	[-0.00, 0.00]	0.03	[-0.31, 0.38]	.00	[-.01, .01]	.14	
								R^2 = .416** 95% CI[NA,NA]

Note. A significant *b*-weight indicates the beta-weight and semi-partial correlation are also significant. *b* represents unstandardized regression weights. *beta* indicates the standardized regression weights. sr^2 represents the semi-partial correlation squared. *r* represents the zero-order correlation. *LL* and *UL* indicate the lower and upper limits of a confidence interval, respectively.
* indicates $p < .05$. ** indicates $p < .01$.

We next computed a partial F-test to compare a full regression model with age, CR, and fluid reasoning as predictors, with a reduced model which included just CR and fluid reasoning. Results showed that age was still not a significant predictor of the bin score (F=1.8282, p = .09284) even after partialling out the effects of CR and speed/attention.

Correlational Analyses. Correlational analyses were conducted between the bin scores and various neuropsychological variables of interest, including all neuropsychological composite scores (reasoning, memory, speed/attention, and vocabulary), as well as the CR and executive function composites. A correlation matrix can be seen in figure 11.

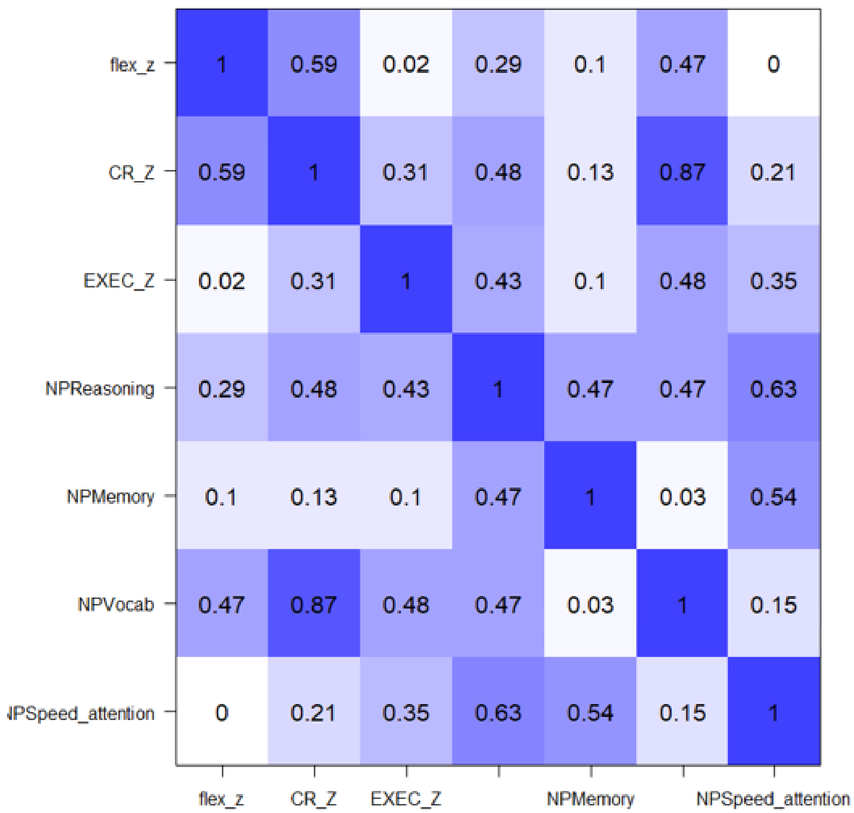

Figure 11. Correlation matrix for bin scores with composite measures (flex_z = bin score, CR_z = standardized CR, EXEC_z = standardized EF.)

In general, the biggest correlations were observed between strategic flexibility and CR measures, including the composite vocabulary score. However the fluid reasoning composite score also correlated well (.29) with the bin measure. Additionally, education individually was highly correlated with this score (.55), and as expected the bin score correlated moderately with matrix reasoning performance from both the WAIS-III (.23) and the total number of correct answers from the RANN matrix reasoning task (.26). The letter-number sequencing task from the WAIS-III also showed moderate correlation (.21), consistent with working memory's role both in task-switching and in matrix reasoning performance.

Mediation Models. To directly compare the effectiveness of CR and the flexibility score as mediators of age-related cognitive performance differences, we entered the simple raw scores from NART IQ and the bin score into a mediation model between age and raw scores from the WAIS-III Matrix Reasoning test. Results from these models are displayed in tables 43 and 44. While both models held significance, NART IQ was still a better mediator of the effect of age on cognitive performance.

Table 43

Mediation Results using NART IQ.

Total Effect Estimates					
	WAIS3_Matrix	SE	t	df	Prob
Age	-0.12	0.03	-4.31	76	4.85e-05
Direct Effect Estimates					
	WAIS3_Matrix	SE	t	df	Prob
Age	-0.15	0.02	-6.21	76	2.61e-08
NARTIQ	0.26	0.05	5.28	76	1.17e-06
R=0.64	R2=0.41	F=26.48	P=1.87e-09		
'a' Effect Estimates					
	NARTIQ	SE	t	df	Prob
Age	0.13	0.05	2.37	77	.0205
'b' Effect Estimates					
	WAIS3_Matrix	SE	t	df	Prob
NARTIQ	0.26	0.05	5.28	76	1.17e-06
'ab' Effect Estimates					
	WAIS3_Matrix	boot	sd	lower	upper
Age	0.03	0.04	0.02	0	0.7

Table 44

Mediation Results using Bin Scores.

Total Effect Estimates					
	WAIS3_Matrix	SE	t	df	Prob
Age	-0.12	0.03	-4.31	76	4.85e-05
Direct Effect Estimates					
	WAIS3_Matrix	SE	t	df	Prob
Age	-0.11	0.03	-3.91	76	0.000198
Flex score	-0.05	0.02	-2.40	76	0.018700
R=0.5	R2=0.25	F=12.75	P=1.68e-05		
'a' Effect Estimates					
	Flex Score	SE	t	df	Prob
Age	0.23	0.14	1.67	77	.0985
'b' Effect Estimates					
	WAIS3_Matrix	SE	t	df	Prob
Flex Score	-0.05	0.02	-2.4	76	0.0187
'ab' Effect Estimates					
	WAIS3_Matrix	boot	sd	lower	upper
Age	-0.01	-0.01	0.01	-0.04	0

Discussion

Results of the third study replicated the findings of studies 1 and 2, with older adults displaying greater impairment to performance in mixed blocks relative to younger adults, and all participants showing greater RTs in mixed than in pure blocks. Additionally, the high correlations between our bin metric and CR, as well as the moderate correlations with other indices of fluid reasoning performance as well as with working memory supported the validity of the STRATA as an effective proxy for both CR and to some extent fluid reasoning itself.

Our lack of significant moderated-mediation findings was surprising given the role that CR had played in relation to the tested set of brain regions, and the theoretically closer relationship between performance on this task and fluid reasoning performance. However, it is possible that strategic flexibility is just one mechanistic component of CR, and hence may play a role in moderating the effect of age-related variance in other brain regions not tested here on cognitive performance. Future studies will need to investigate this possibility, as well as the possibility that this metric would more effectively moderate the relationship of functional rather than structural variance on cognitive performance. Indeed, given the literature reviewed in the general introduction, this relationship would be more expected.

Moderated-mediation is also a stringent analysis that was possibly too ambitious considering the limited sample size. Given this, the moderation results are perhaps a fairer test of the potential for strategic flexibility to serve in a similar role to CR. Even here we failed to observe conclusive evidence that better strategy-shifting performance can differentially affect the relationship between brain volume and reasoning performance in previously identified regions. However there is suggestion that on a global scale brain volume can be utilized differentially depending on one’s strategic performance, exactly as would be predicted if strategic flexibility is a manifestation of reserve.

While this study did not find any conclusive evidence that strategic flexibility could be a mechanism of CR, the interesting pattern of behavioral correlations observed between the STRATA task and various other neuropsychological instruments suggests it may have some utility as not only a new measure of strategy use within fluid reasoning, but also as a non-exposure based proxy variable of CR.

CHAPTER 6
GENERAL DISCUSSION

The studies outlined in this project introduce a new instrument for assessing the degree of flexible engagement in alternative strategies within a fluid reasoning task. Study 1 demonstrated that the STRATA task, and potentially other complex reasoning tasks decomposed into their constituent strategies and presented within a vetted task-switching design, is sensitive to age effects in the form of switch costs. In particular, mixing costs were determined to be significantly impacted by age, which fits well with existing literature on the importance of working memory to both task switching and fluid intelligence. Study 2 replicated this effect, and demonstrated that these forms of tasks need not be confined to analysis with difference scores (and in fact, shouldn't, given the potential to view speed-accuracy trade-offs as a higher-level cognitive strategy in and of itself), and showed a strong correlation between an index of CR and performance on this task in older but not younger participants. Study 3 explored the relationships between this task and standard neuropsychological instruments as well as neurostructural variables. Taken together, these findings suggest a novel method of investigating cognitive strategy use within the context of preexisting tasks and a preexisting and well-understood experimental paradigm.

Despite its findings, some of the major hypotheses of this project were not supported. In particular, the hypothesis that strategic flexibility could serve as a mechanistic proxy for CR in existing moderated-mediation models was invalidated. Likewise, direct comparison of this measure with a traditional proxy variable of CR within mediation models failed to show that it can account for a stronger mediational effect than the existing metric. These results may indicate that strategic flexibility is just one small part of a much larger puzzle, or they may derive from inadequate methods of testing its mechanistic role.

One potential extension of the methods employed here might be to investigate STRATA performance within the context of neuroimaging. As suggested above, decomposition of existing tasks into strategic components has been perhaps most instructive when it has demonstrated that group-wide activation differences may derive from differential strategy use, rather than from cognitive deficits or mostly static individual differences. Moderated-mediation models may even be applied to such activation differences more successfully, since the strategic flexibility metric outlined here is the result of dynamic processing and not necessarily of the features inherent to an individual. It would be particularly interesting to see if task-invariant activation patterns (such as the RANNs described above) can be further decomposed based on domain-wide strategy use rather than their component tasks.

Another limitation of this series is the unclear relationship between the STRATA performance and many of the other dimensions of cognitive strategy use mentioned in the introduction. While performance of this task does seem to measure some aspects of strategy selection and strategy execution in a highly structured way, it leaves untouched the dimensions of strategy repertoire and strategy distribution. It could be the case that a more direct relationship exists between CR on the one hand, and strategy repertoire; as formal education increases, for instance, it's easy to imagine that so does exposure to various novel strategies for performing cognitive tasks both structured and unstructured.

Future work extending this line of research into other strategic dimensions is required to answer this question.

Another potential extension of this line of research lies in the prospect of training reasoning strategies. Evidence already exists that such training can be more successful than more procedural forms of training (such as working memory training). In two separate studies Aries, Groot, & van den Brink (2014) used a cognitive intervention program in attempt to improve reasoning abilities among secondary school students. In the first intervention, the researchers utilized a working memory training intervention resulting in an increase in scores of reasoning ability. While the second study the researchers used independent training of reasoning strategies along with working memory training intervention, resulting in a significant increase in reasoning ability scores based on the reasoning strategies training but not the working memory capacity training. Importantly, the working memory training in these studies was embedded within the context within domain-relevant material (history), suggesting that limitations to transfer effects may be overcome if the context of the cognitive capacity being trained is taken into account by researchers.

REFERENCES

Alexander GE, Chen K, Aschenbrenner M, Merkley TL, Santerre-Lemmon LE, et al. (2008) Age-related regional network of magnetic resonance imaging gray matter in the rhesus macaque. J Neurosci 28: 2710â€“2718. Barulli, D., & Stern, Y. (2013). Efficiency, capacity, compensation, maintenance, plasticity: emerging concepts in cognitive reserve. Trends in cognitive sciences, 17(10), 502-509.

Barulli, D. J., Rakitin, B. C., Lemaire, P., & Stern, Y. (2013). The influence of cognitive reserve on strategy selection in normal aging. Journal of the International Neuropsychological Society, 19(07), 841-844.

Barulli, D., Steffener, J., Lemaire, P., & Stern, Y. (In Revision). Cognitive Strategies : A Mechanism for Cognitive Reserve?

Buckner RL, Head D, Parker J, et al. A unified approach for morphometric and functional dataanalysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume.Neuroimage 2004;23:724-738.

Buschke H, Fuld PA. Evaluating storage, retention, and retrieval in disordered memory and learning. Neurology 1974;24:1019-1025.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. Psychological review, 97(3), 404.

Czernochowski, D., Fabiani, M., & Friedman, D. (2008). Use it or lose it? SES mitigates age related decline in a recency/recognition task. Neurobiology of aging, 29(6), 945-958.

Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 1999;9:179-194.

DeShon, R. P., Chan, D., & Weissbein, D. A. (1995). Verbal overshadowing effects on Raven's Advanced Progressive Matrices: Evidence for multidimensional performance determinants. Intelligence, 21(2), 135-155.

Doran SM, Van Dongen HP, Dinges DF. Sustained attention performance during sleep deprivation: evidence of state instability. Arch.Ital.Biol. 2001;139:253-267.

Fieo, R., Manly, J. J., Schupf, N., & Stern, Y. (2014). Functional Status in the 'Young' Old: Establishing a Working Prototype of an Extended-Instrumental Activities of Daily Living Scale. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 69(6), 766-772.

Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 2002;33:341-355.

Fischl B, van der Kouwe A, Destrieux C, et al. Automatically Parcellating the Human Cerebral Cortex. Cereb. Cortex 2004;14:11-22.

Fischl B, Dale AM. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proceedings of the national Academy of Sciences of the United States of America 2000;97:11050-11055.

Fjell AM, Westlye LT, Amlien I, et al. High Consistency of Regional Cortical Thinning in Aging across Multiple Samples. Cereb. Cortex 2009:bhn232.

Galvin, J. E., Roe, C. M., Coats, M. A., & Morris, J. C. (2007). Patient's rating of cognitive ability: using the AD8, a brief informant interview, as a self-rating tool to detect dementia. Archives of neurology, 64(5), 725-730.

Goodglass H, Kaplan E. The Assessment of Aphasia and Related Disorders. Philadelphia, PA: Lea & Febiger, 1983.

Golden CJ. Stroop Color and Word Test. Chicago, IL: Stolting, 1978. Grant DA, Berg EA. Wisconsin Card Sorting Test. San Antonio: Psychological Corporation,1993.

Hayes, A. F. (2012). PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling.

Hodzik, S., & Lemaire, P. (2011). Inhibition and shifting capacities mediate adults' age-related differences in strategy selection and repertoire. Acta psychologica, 137(3), 335-344.

Huff, C., & Tingley, D. (2015). 'Who are these people?' Evaluating the demographic characteristics and political preferences of MTurk survey respondents. Research & Politics, 2(3), 2053168015604648.

Hunt, E. (1974). Quote the Raven? Nevermore.

Kennedy KM, Erickson KI, Rodrigue KM, et al. Age-related differences in regional brain volumes: a comparison of optimized voxel-based morphometry to manual volumetry. Neurobiology of Aging 2009;30:1657-1676.

Kirby, J. R., & Lawson, M. J. (1983). Effects of strategy training on progressive matrices performance. Contemporary Educational Psychology, 8(2), 127-140.

Kramer, J. H., Mungas, D., Possin, K. L., Rankin, K. P., Boxer, A. L., Rosen, H. J., ... & Widmeyer, M. (2014). NIH EXAMINER: conceptualization and development of an executive function battery. Journal of the International Neuropsychological Society, 20(01), 11-19.

Li, Z., Moore, A. B., Tyner, C., & Hu, X. (2009). Asymmetric connectivity reduction and its relationship to 'HAROLD' in aging brain. Brain research, 1295, 149-158. Macdonald PA, Ganjavi H, Collins DL, Evans AC, Karama S (2013) Investigating the relation between striatal volume and IQ. Brain Imaging Behav. Matzen, L. E., Benz, Z. O., Dixon, K. R., Posey, J., Kroger, J. K., & Speed, A. E. (2010). Recreating Raven's: Software for systematically generating large numbers of Raven-like matrix problems with normed properties. Behavior research methods, 42(2), 525-541.

Mattis S. Dementia Rating Scale (DRS). Odessa, FL: Psychological Assessment Resources, 1988.

Poldrack, R. A. (2015). Is 'efficiency' a useful concept in cognitive neuroscience?. Developmental cognitive neuroscience, 11, 12-17.

Raven, J. (2003). Raven progressive matrices. In Handbook of nonverbal assessment (pp.223-237). Springer US.

Reuter-Lorenz, P. A., & Cappell, K. A. (2008). Neurocognitive aging and the compensation hypothesis. Current directions in psychological science, 17(3), 177-182.

Robertson, I. H. (2013). A noradrenergic theory of cognitive reserve: implications for Alzheimer's disease. Neurobiology of aging, 34(1), 298-308. Samanez-Larkin GR, Kuhnen CM, Yoo DJ, Knutson B (2010) Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking. J Neurosci 30: 1426â“1434. Sandry, J., & Sumowski, J. F. (2014). Working memory mediates the relationship between intellectual enrichment and long-term memory in multiple sclerosis: an exploratory analysis of cognitive reserve. Journal of the International Neuropsychological Society, 20(08), 868-872.

Siedlecki KL, Stern Y, Reuben A, Sacco RL, Elikind MSV, Wright CW. Construct validity of cognitive reserve. Journal of the International Neuropsychological Society 2008.

Steffener, J., Barulli, D., Habeck, C., O'Shea, D., Razlighi, Q., & Stern, Y. (2014). The role of education and verbal abilities in altering the effect of age-related gray matter differences on cognition. PloS one, 9(3), e91196.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve

Stern, Y. (2009). Cognitive reserve. Neuropsychologia, 47(10), 2015-2028.

Stern, Y., Habeck, C., Steffener, J., Barulli, D., Gazes, Y., Razlighi, Q., ... & Salthouse, T. (2014). The Reference Ability Neural Network Study: Motivation, design, and initial feasibility analyses. NeuroImage, 103, 139-151.

United States Employment Service. Dictionary of Occupational Titles, 4th ed, revised Washington, DC: Superintendent of Documents, US Government Printing Office; 1977. Available at: www.occupationalinfo.org. Accessed June 11, 2018.

Uttl, B. (2002). North American Adult Reading Test: age norms, reliability, and validity. Journal of Clinical and Experimental Neuropsychology, 24(8), 1123-1137.

Valenzuela, M. J., & Sachdev, P. (2007). Assessment of complex mental activity across the lifespan: development of the Lifetime of Experiences Questionnaire (LEQ). Psychological medicine, 37(07), 1015-1025.

Walhovd KB, Westlye LT, Amlien I, et al. Consistent neuroanatomical age-related volume differences across multiple samples. Neurobiology of Aging; In Press, Corrected Proof.

Wechsler, D. (1997). WAIS-III administration and scoring manual (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler D. Wechsler Test of Adult Reading. San Antonio, TX: The Psychological Corporation, 2001.

Whalley, L. J., & Smyth, K. A. (2013). Human culture and the future dementia epidemic Crisis or crossroads?. Neurology, 80(20), 1824-1825. concept. Journal of the International Neuropsychological Society,8(03), 448-460.

Woods, S. P., Weber, E., Cameron, M. V., Dawson, M. S., Delano-Wood, L., Bondi, M. W., & Grant, I. (2010). Spontaneous strategy use protects against visual working memory deficits in older adults infected with HIV. Archives of clinical neuropsychology, acq069.