UNIVERSITY OF CALIFORNIA,
IRVINE

Adaptive Learning in Later Life

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Cognitive Sciences

by

Priyam Das

Dissertation Committee:
Professor Mark Steyvers, Chair
Professor Joachim Vandekerckhove
Assistant Professor Aaron M. Bornstein

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I would like to thank the members of my doctoral committee. Thank you to my advisor, Mark Steyvers, for giving me the freedom to pursue my intellectual interests and for treating me as an expert on cognitive aging long before I saw myself as one. Thank you to Joachim Vandekerckhove for welcoming me into the department and for his unwavering support throughout my time in the program. Thank you to Aaron Bornstein for always seriously listening to my ideas and providing mentorship. I thank all three for reviewing my dissertation and for their helpful comments. I'd also like to thank Susanne Jaeggi and Susan Charles for serving on my advancement committee and for their helpful feedback on my dissertation proposal. Additionally, I thank Clara Schultheiss who has helped me as much as any professor by patiently answering my numerous questions about how the department and university function.

I must also thank my first research mentor, Falk Lieder, for being so patient with me and teaching me all the necessary steps for conducting a research project. Chapter 1 would not exist without his guidance. Thank you to Fred Callaway for patiently answering my numerous programming questions, for helping me work out research ideas, and for your friendship. Chapters 1 and 3 benefitted from your helpful feedback. I also have immense gratitude for Tom Griffths and all the graduate students who were part of the Computational Cognitive Science Lab during its final years at UC Berkeley. Thank you for always welcoming me to your lab meetings, for providing guidance while I applied for Ph.D. programs, for being a friendly face at conferences, and for being role model researchers.

I truly could not have made it this far without the support of so many friends. I considered quitting multiple times and every time, there was someone there who listened to my frustrations and helped me get back on my feet. Thank you to the members of my cohort, in particular Xiaojue Zhou, Paulina Silva, Aakriti Kumar, Nidhi Banavar, Kexin Chen, Jinwei Xing, Jeff Coon, and Bobby Thomas, for always encouraging and inspiring me. Thank you to Nora Harhen, Lauren Montgomery, Manuel Villareal Ulloa, Brandon Hackney, Nora Bradford, Adriana Felisa Chávez De la Peña, Jaime Islas Farias, Heliodoro Tejeda Lemus, Alexander Bower, Arseny Moskvichev, Alexander Etz, Mac Strelioff, Emily Sumner, Hirak Jyoti Kashyap, Tulika Kakati, and Jessica Maya Hernandez for creating a supportive learning environment and for helping me leave UC Irvine with good memories.

A very special thanks goes to Nora Harhen, who has been encouraging me in my research efforts ever since I was struggling with my undergraduate thesis. I am so thankful we ended up in the same graduate program, though I'm sure you would have continued to provide support from afar much like you have during my final months in the program. Thank you for spending too many hours baking with me, for spending 6+ hours in a car with me, for all the cups of tea, and for always hearing me out. I am so lucky to have a friend like you.

I am also grateful for the friends outside of academia who rarely understood what I was going through but showered me with love and support nonetheless. Thank you to Prerana Sarma, Shubha Chakravarty, Shreya Bhattacharyya, and Runa Sarma for all the beach days. Thank you to Guy Geva, Kay Toma, Kaat Ceder, Christine Okazaki, Elizabeth Yen, Yves Menshikova, Shorwei Gong and Jasmine Balsara for always reminding me that I am so much more than just a scientist and that the world contains so much more than academia. A special thanks goes to Shorwei, whose work experiences inspired the ideas behind chapters 3 and 4.

I am fortunate to have a family who strongly believes in education and has supported my studies every step of the way, even if it was in a field they'd never heard of. Thank you Ma, for caring about the minutiae of my Ph.D. program and for maintaining a warm home for me to return to and recharge whenever I needed. Thank you Deuta, for helping me move to Irvine and for caring about my mental health in your own way. Thank you to my brother Pritish, for helping me stay up to date with the slang used by undergraduates and for always making me laugh. Thank you to my extended family who regularly inquired about my studies and cheered me on from afar.

I simply cannot thank Michael Oshiro enough for fiercely believing in my abilities despite all the evidence I gathered to try and prove him wrong. I do not have enough space to thank him for everything he's done to help me get this far, but I'll list a few. Thank you for being so willing to set aside your own research to help me with mine. Thank you for indulging me every time I got on my soapbox about the way science should be done. Thank you for traveling great distances to be by my side during program milestones. Thank you for your unconditional support. Let's continue to share a brain cell and eat tasty food together.

Finally, I would not have been able to continue with my research during the coronavirus pandemic nor subsequently produce this dissertation without the guidance of my therapist, Kai Chiu LMFT. If any struggling graduate students are reading this, I highly recommend seeking out a therapist and I hope you find someone as wonderful and effective as Kai.

VITA

Priyam Das

EDUCATION

Doctor of Philosophy in Cognitive Sciences **2024**
University of California, Irvine *Irvine, CA*

Master of Science in Cognitive Sciences **2021**
University of California, Irvine *Irvine, CA*

Bachelor of Arts in Cognitive Science **2018**
University of California, Berkeley *Berkeley, CA*

RESEARCH EXPERIENCE

Graduate Student Researcher **2018–2023**
University of California, Irvine *Irvine, California*

Undergraduate Research Assistant **2015–2018**
University of California, Berkeley *Berkeley, California*

TEACHING EXPERIENCE

Teaching Assistant **2018–2022**
University of California, Irvine *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

Older Adults Catch Up to Younger Adults on Cognitive Tasks After Extended Training **2023**
Collabra: Psychology

Authors: **Das, P.** & Steyvers, M.

Leveraging Artificial Intelligence to Improve People's Planning Strategies **2022**
Proceedings of the National Academy of Sciences

Authors: Callaway, F., Jain, Y.R., van Opheusden, B., **Das, P.**, Iwama, G., Gul, S., Krueger, P.M., Becker, F., Griffiths, T.L., & Lieder, F.

Rational use of cognitive resources in human planning **2022**
Nature Human Behavior

Authors: Callaway, F., van Opheusden, B., Gul, S., **Das, P.**, Krueger, P.M., Lieder, F., & Griffiths, T.L.

REFEREED CONFERENCE PUBLICATIONS

Remediating Cognitive Decline with Cognitive Tutors **July 2019**
The Fourth Multidisciplinary Conference on Reinforcement Learning and Decision Making

Authors: **Das, P.**, Callaway, F., Griffiths, T.L., & Lieder, F.

A cognitive tutor for helping people overcome present bias **July 2019**
The Fourth Multidisciplinary Conference on Reinforcement Learning and Decision Making

Authors: Lieder, F., Callaway, F., Jain, Y.R., Krueger, P.M., **Das, P.**, Gul, S., & Griffiths, T.L.

A resource-rational analysis of human planning **July 2018**
Proceedings of the 40th Annual Meeting of the Cognitive Science Society

Authors: Callaway, F., Lieder, F., **Das, P.**, Gul, S., Krueger, P.M., & Griffiths, T.L.

AWARDS & HONORS

Social Science Communication Fellowship **2021**
University of California, Irvine

NSF Graduate Research Fellowship Program Honorable Mention **2019**
National Science Foundation

Undergraduate Thesis Highest Honors **2018**
University of California, Berkeley

SERVICE

Cognitive Sciences Graduate Student Liaison **2020-2022**
University of California, Irvine

Cognitive Sciences Colloquium Committee Member **2020-2021**
University of California, Irvine

California Cognitive Science Conference Organizer **2015-2018**
University of California, Berkeley

ABSTRACT OF THE DISSERTATION

Adaptive Learning in Later Life

By

Priyam Das

Doctor of Philosophy in Cognitive Sciences

University of California, Irvine, 2024

Professor Mark Steyvers, Chair

As people age, they can expect some of their memory to fail or to become slower while doing mental calculations. However, it is an oversimplification to describe cognitive aging as only cognitive deterioration. This dissertation explores various ways that adult learners can adapt to cognitive changes brought by aging. In chapter 1, we find that older adults are worse than younger adults at a planning task because they use suboptimal planning strategies. Once we teach older adults the optimal strategy, there are no longer differences in performance between the two age groups. In chapter 2, we show how extra practice can help older adults increase their ability to match or exceed the performance of a younger adult on various cognitive tasks. In chapters 3 and 4, we investigate how older adults allocate their effort towards learning new technologies depending on ease of use and the amount of time needed to learn. We found that older adults can learn a complex virtual machine as well as younger adults and that older adults are as willing as younger adults to expend effort towards learning difficult machines. However, we also found that older adults will conserve their effort and choose easier machines when we manipulated the amount of time they have to learn. Altogether, these results support the notion of older adults as adaptive learners and help provide a more nuanced view of cognitive aging, one where older adults are sometimes capable of excelling on cognitive tasks.

INTRODUCTION

Humans have an incredible capacity to learn and they engage in learning all throughout their lives. From the baby learning about herself and her environment, to the child learning his letters and numbers, to the young adult learning to participate in society, to the older adult learning to adapt to a changed environment – people are continuously learning. At different points in life, people can choose to learn different things, and they learn them in different ways. In particular, adults who are 60 years and older learn differently from adults who are in their 20s and 30s. These differences partly attributed to changes in cognitive processing, changes which naturally accompany healthy aging.

Aging Effects on Learning & Decision Making

Learning is supported by various cognitive processes, many of which decline with age. As people grow older, the speed at which they are able to process information becomes slower and their ability to keep multiple pieces of information in their working memory starts to fail, which in turn affects problem solving and decision making (Salthouse, 1996, 2010; Murman, 2015). Cognitive control, which refers to the ability to focus on relevant information during learning while ignoring irrelevant or outdated information, has also been well documented to decline with age (Kramer, Hahn, & Gopher, 1999; Kray & Lindenberger, 2000; Kray & Ferdinand, 2014). Deficits in cognitive control are apparent when older adults are slower to switch between different tasks compared to younger adults or make more errors on a new task due to applying a strategy which worked well on a previous task.

Another notable age-related change in learning behaviors is the preference for using prior knowledge to solve problems over learning about new information or options. This is often referred to as the explore-exploit tradeoff, where exploring refers to investigating new, potentially risky, options while exploiting refers to choosing what you already know as a safe bet. Older adults are more likely to exploit known options compared to younger adults (Queen, Hess, Ennis, Dowd, & Grühn, 2012; Mata, Wilke, & Czienskowski, 2013; Wiegand, Seidel, & Wolfe, 2019; Spreng & Turner, 2021). Declines in cognitive control may be driving this shift from exploration to exploitation since cognitive control is required to switch from the current, known, option to a new option (Mata & von Helversen, 2015; Spreng & Turner, 2021). However, another explanation for this shift may depend on perceptions of time. Socioemotional selectivity theory (SST) states that older adults have a bias for maintaining positive experiences in the present moment due to considerations of their lifespan (Carstensen, 2006; Löckenhoff, 2011; Carstensen, 2021). Though usually used to explain social situations such as an older adult preferring to spend time with close friends over making new friends, SST can be applied more broadly to older adults' non-exploratory behavior (Spreng & Turner, 2021). Staying with and exploiting known options is more likely to bring a positive outcome compared to risky new options, especially if the known option has produced positive outcomes before. Thus, time perception is important to consider along with declines in cognitive control when trying to understand older adults' reluctance to learn new information.

Learning efficient decision making strategies is critical for handling all of life's decisions both big and small. When deciding between the organic bananas or the regular bananas at the supermarket, you might only focus on one feature, the price, in order to make the decision. However, when deciding between different mortgage loans, you might compare them based on multiple features such as the monthly payment, the interest rate, and the length of the loan, before making your decision. These are just a couple examples of different decision making strategies. In studies of decision making, older adults show a strong preference for simpler strategies even when those strategies are not best suited for the task (Mata, Schooler, & Rieskamp, 2007; Mata, von Helversen, & Rieskamp, 2010; Lemaire, 2010). This preference for simpler strategies leads to poorer performance on some decision making tasks compared to young adults who engage in more cognitively effortful strategies (Lemaire, 2010; Worthy & Maddox, 2012; Hinault & Lemaire, 2020). Rather than laziness, the reason for preferring cognitively easier strategies appears related to declines in cognitive control which promote a reliance on habitual, tried-and-true strategies (Blanco et al., 2016; Bolenz, Kool, Reiter, & Eppinger, 2019).

An Optimistic Outlook for the Aging Mind

Although results such as cognitive deficits and poor decision making outcomes dominate much of the aging literature and conjure a bleak outlook for cognitive abilities during old age, there is a growing body of results which paint older adults as adaptive learners. While it appears that older adults perform poorly compared to younger adults, these age differences can disappear with minor task modifications or consideration of older adults' cognitive resources. In this next section, I discuss how additional practice and resource-rational approaches shine a different light on older adults' behaviors during learning.

When older adults can do a task as well as younger adults, this is referred to as age equivalence. Simply practicing more on a task is one way that older adults can reach age equivalent performance with younger adults. For example, in task switching experiments, older adults show larger switch costs, or have more difficulties switching to a new task, compared to younger adults. However, these switch costs are reduced and can become age equivalent after a few additional sessions of practice (Kramer et al., 1999; Buchler, Hoyer, & Cerella, 2008; Karbach & Kray, 2009). Some other tasks in which older adults can reach age equivalence after extra practice are collaborative, visual discrimination, and implicit learning tasks, to name a few (Derksen et al., 2015; Ratcliff, Thapar, & McKoon, 2006; Myers & Conner, 1992). These studies show that for some tasks, older adults learn slower and thus require more time than younger adults to become proficient.

When older adults are unable to match the performance of younger adults, their behavior might be explained by resource-rationality (Griffiths, Lieder, & Goodman, 2015; Lieder & Griffiths, 2020). Under resource-rationality, older adults act in a way that is optimal for them given their limited cognitive resources. This explanation accounts for situations in which older adults demonstrate that they are capable of using the same cognitively effortful decision making strategies as younger adults, even if they use such strategies less frequently (e.g. Worthy & Maddox, 2012; Blanco et al., 2016; Devine et al., 2021). Because older adults have less cognitive resources on average compared to younger adults, they are more selective in expending their cognitive effort. For tasks requiring cognitive control, one hypothesis is that all individuals have a "sweet spot" in which they are engaging in the right amount of control given the constraints of the task and their own cognitive limitations (Ruel, Devine, & Eppinger, 2021). When more cognitive control is required to perform better at the task, if the individual does not consider the increase in performance to be worth the increase in control, then they will maintain their current effort levels and appear to be performing suboptimally.

Dissertation Overview

My dissertation work adopts this optimistic outlook on learning in later life. In chapter 1, I investigate whether older adults struggle with planning due to their choice of planning strategies, much like they do in decision making tasks. I use recently developed methods for visualizing and categorizing people's planning strategies and find that older adults use the optimal strategy much less frequently than younger adults. Since older adults seem unaware of the optimal strategy, I introduce a cognitive tutor to teach older adults the optimal planning strategy and this enables them to reach age equivalence with younger adults on the planning task. This chapter is based off of the conference paper "Remediating Cognitive Decline with Cognitive Tutors" which is included in the non-archival proceedings for the Fourth Multidisciplinary Conference on Reinforcement Learning and Decision Making. The work in this chapter builds off of ideas from my co-authored journal publications, "Rational use of cognitive resources in human planning" and "Leveraging Artificial Intelligence to Improve People's Planning Strategies", but is a completely separate project from those two publications since this work focuses on the effects of age.

In chapter 2, I present evidence that additional practice helps older adults get closer to reaching age equivalence with younger adults on tasks that engage a variety of cognitive processes. I come to this conclusion from a secondary analysis of a large scale data set of thousands of users of a popular online brain training website. Additionally, I present a new method of quantifying the benefits of additional practice at an individual, rather than a group, level. This chapter appears in this dissertation exactly as it was published, with the same title, in the journal *Collabra: Psychology.*

Chapters 3 and 4 dig deeper into older adults' preferences for exploiting known options and conserving cognitive effort during learning. In chapter 3, I introduce a new task that lets us observe how people invest effort during learning. This task asks people to learn how to use a complex machine and gives people the option to choose between different models of the machine which vary in their ease of use. I find that both older adults and younger adults are not only equally capable of learning to use the hardest version of the machine but that both groups prefer this machine over easier options. In chapter 4, I investigate on how people's time perception affects their preference for using the different types of machines and whether older adults optimally conserve their effort to learn a new machine when time is short. I find that younger adults are more strongly affected by the experimental manipulation of time and that older adults seem to conserve their efforts regardless of how much time they actually have to learn.

Finally, I close with a discussion of what my findings mean for the optimistic outlook on cognitive aging and some suggestions for future research.

Chapter 1

Cognitive Tutors Help Older Adults Plan Effectively

1.1 Introduction

Planning is an essential skill for achieving goals. However, planning can be very difficult for people to do well, particularly when there are many possible paths towards the goal. Then planning becomes more like a decision making problem as people evaluate the pros and cons of each path before making a choice about how to proceed. Unfortunately, both planning and decision making get harder for people as they get older.

Previous studies have found that older adults have trouble formulating plans and in some cases, executing them (Allain et al., 2005; Sorel & Pennequin, 2008; Sanders & Schmitter-Edgecombe, 2012). This may be linked to other studies showing that older adults tend to perform poorly on decision making tasks compared to younger adults (Frank & Kong, 2008; Mell et al., 2005; Mata et al., 2007; Wiegand et al., 2019). In some cases, this is because older adults use different decision making strategies compared to younger adults and end up using a less efficient strategy (Mata, Josef, & Lemaire, 2015; Blanco et al., 2016; Lemaire, 2010). For example, older adults tend to rely on heuristics or easier decision making strategies (Mata et al., 2007; Lemaire, 2010; Worthy, Cooper, Byrne, Gorlick, & Maddox, 2014). However, older adults are also able to adapt their decision making strategies in certain task environments and in these cases, they are able to adapt as well as younger adults (Mata et al., 2007; Queen et al., 2012; Worthy & Maddox, 2012).

Despite multiple studies investigating how age affects decision making and the strategies that older people use to make decisions, there has been little work studying the problem of planning. This is because planning is generally hard to observe and depending on the task, researchers struggle to disentangle the formulation phase of planning from the execution phase (Sanders & Schmitter-Edgecombe, 2012). We can overcome these difficulties with a new process-tracing task called Mouselab-MDP (Callaway, van Opheusden, et al., 2022; Callaway, Jain, et al., 2022). Mouselab-MDP externalizes the planning process by having participants click to gather information about various paths to a goal before they can take any actions to follow a particular path. By observing participants' clicks, we can observe which action on which path they are considering at any time. Furthermore, computational models of various planning strategies can be fit to participants' data on the Mouselab-MDP task, allowing us identify the different strategies that people use while planning (Callaway, van Opheusden, et al., 2022).

Although we know that using different strategies from younger adults leads to poorer outcomes for older adults, there is a dearth of solutions to help older adults choose better strategies. As with other aging-related cognitive issues such as failing memory or increased inattentiveness, cognitive training programs targeting decision making skills have gained popularity as a potential solution, but their effectiveness remains to be seen (Consensus on Brain Training, 2014; Simons et al., 2016; Kable et al., 2017). Recently, we introduced a new training method to teach people better planning strategies (Callaway, Jain, et al., 2022).

Our cognitive tutor differs from other cognitive training programs by giving feedback about how someone chose an action, rather than giving feedback about the action itself. For example, on the Mouselab-MDP task, there are multiple paths to consider which have hidden rewards and punishments along the way. Feedback based on action alone would describe when a person chooses a path to go down and experiences all the rewards and punishments on that path. However, feedback given on the decision making process would be based on how the person gathered information about each of the potential paths before making their choice of path. Rather than giving feedback such as pointing out a better path, our cognitive tutor gives feedback by pointing out whether someone spent too much time or too little time considering their options. This is a kind of metacognitive feedback which has shown to improve people's performance in different reward environments on the Mouselab-MDP task (Callaway, Jain, et al., 2022).

In the present chapter, we use Mouselab-MDP and the method of modeling planning strategies to study how planning strategies differ by age in Experiment 1. We will also examine how well older adults perform on the task compared to younger adults. In line with the results from previous research, we expect that older adults' will use strategies which are different from younger adults and that this will lead to poorer performance on the task. Thus, in Experiment 2 we will use our cognitive tutor to teach both younger and older adults better planning strategies and examine whether both age groups benefit equally. We expect that the cognitive tutor will help minimize differences in task performance between younger and older adults. Even if we do not find age differences in task performance in Experiment 1, we can evaluate how much both age groups benefit from using the tutor compared to a control group.

1.2 Experiment 1

Methods

We recruited participants younger than 25 years old to form our younger adults group ($19-24$ y.o., median $= 23$, $n = 49$) and adults older than 47 years old to form our older adults group ($48 - 70$ y.o., median $= 52$, $n = 29$). The experiment was conducted online via Amazon Mechanical Turk. In the experiment, participants completed 30 trials of the Mouselab-MDP paradigm (Callaway, Lieder, Krueger, & Griffiths, 2017) with a three-step route planning task. On each trial, participants were shown a map of gray circles (Figure 1.1) and instructed to move the spider in the middle to one of the outermost nodes, picking up the rewards hidden along the way. For each trial, rewards are independently drawn from discrete uniform distributions; in the first step the possible values were $\{-4, -2, +2, +4\}$; in the second step the possible values were $\{-8, -4, +4, +8\}$; and in the third step the possible values were $\{-48, -24, +24, +48\}$. Participants could uncover rewards beforehand by clicking on the gray circles and paying a cost of -1 for each reveal. Participants were instructed to maximize their rewards and were incentivized with a monetary bonus based on their in-game score.

We use the clicks our participants made to infer which kind of planning strategy they used. We considered six different planning strategies: depth-first search, breadth-first search, best-first search, progressive deepening, the optimal planning strategy, and an impulsive strategy that chooses randomly. Depth-first search explores a single path at a time – from its beginning to its end. Once it reaches the end of this path, it returns to the most recent unexplored fork in that path and continues exploring until all nodes have been inspected. Breadth-first search explores the first nodes of all possible paths, then the second nodes, and so on until all paths have been explored. Best-first search explores paths in the order of highest expected sum of rewards. Progressive deepening is a strategy proposed by Newell and Simon (1972) and is similar to depth-first search. The main difference is that after exploring a path

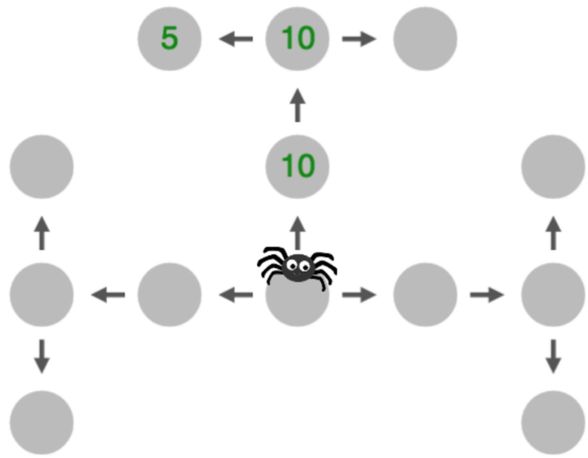

Figure 1.1: A typical Mouselab-MDP trial used in Experiment 1 and the control condition of Experiment 2. Some of the rewards have already been revealed by the participant.

in its entirety, progressive deepening skips back to the starting node, treating branches as part of another path for later exploration. Callaway et al. (2018) found that the optimal strategy for the task environment used in this experiment is to first set a goal by evaluating potential final destinations. As soon as inspecting a potential final destination uncovers the highest possible reward (+48), the optimal strategy selects the path that leads to it and terminates planning. If multiple potential final destinations are good (i.e., +\$24) then the optimal strategy collects additional information about the paths leading to those promising potential final destinations starting with the nodes right before a potential final destination.

Modeling Strategies

We modeled participants' click sequences as a combination of following one of the six strategies described above and some random moves. Formally, the probability of making click c when following strategy k is defined as

$$(1-\epsilon)\cdot\sigma(c;V_{b,M_k},\tau)+\epsilon\cdot\mathrm{Uniform}(c;C_b) \tag{1.1}$$

where the first term, $\sigma(c; V_{b,M_k}, \tau)$, is a softmax over the possible clicks c in state b when following strategy k and τ is the temperature parameter. The second term, $\text{Uniform}(c; C_b)$, can account for actions that are inconsistent with strategy k; the probability of such "random clicks" is modeled by a uniform distribution over all possible clicks and the action of stopping planning. Finally, ϵ is the probability that a random click will be made.

The random strategy can therefore be modeled by the second term alone. The systematic behavior of the other strategies was modeled in terms of the values $V_{b,M}(c)$ they assign to different clicks c and the decision to terminate planning. For example, in the depth-first search model, the preference function $V_{b,DFS}(c)$ is the depth of the node inspected by click c if that node lies on a partially explored path and a large negative value otherwise. As a result, deeper nodes are prioritized and partially explored paths will be explored to the end before others are considered. In the optimal strategy model, the value assigned to $V_{b,O}$ is given by the optimal solution to the problem of deciding how to plan. In previous work, we formalized this problem as a meta-level Markov Decision Process and computed its solution for the environment used in this study using backwards induction (Callaway et al., 2018). Aside from the random and optimal strategy models, all of our strategy models also capture previous findings that people often act as soon as they have identified an alternative they deem good enough (i.e., satisficing; Simon, 1956) and tend to stop considering a course of action when they realize it would entail a large loss at one point or another (i.e., pruning; Huys et al., 2012). To model satisficing and pruning, our models include two free parameters for the participant's aspiration level and pruning threshold respectively.

When the expected reward for terminating in belief state b exceeds the aspiration level, then our models assign a very large value to the terminate planning action. Conversely, if the expected sum of rewards for any path falls below the pruning threshold, then clicks on the remaining unobserved nodes on that path are assigned a large negative value such that the strategy was discouraged from continuing to explore that unprofitable path. We fit all models to each trial for each participant using maximum likelihood estimation for all model parameters (i.e., τ, ϵ, and the thresholds for satisficing and pruning). We then performed model comparisons using the Bayesian Information Criterion (Schwarz, 1978) to determine which strategy each participant is most likely to have used on each trial.

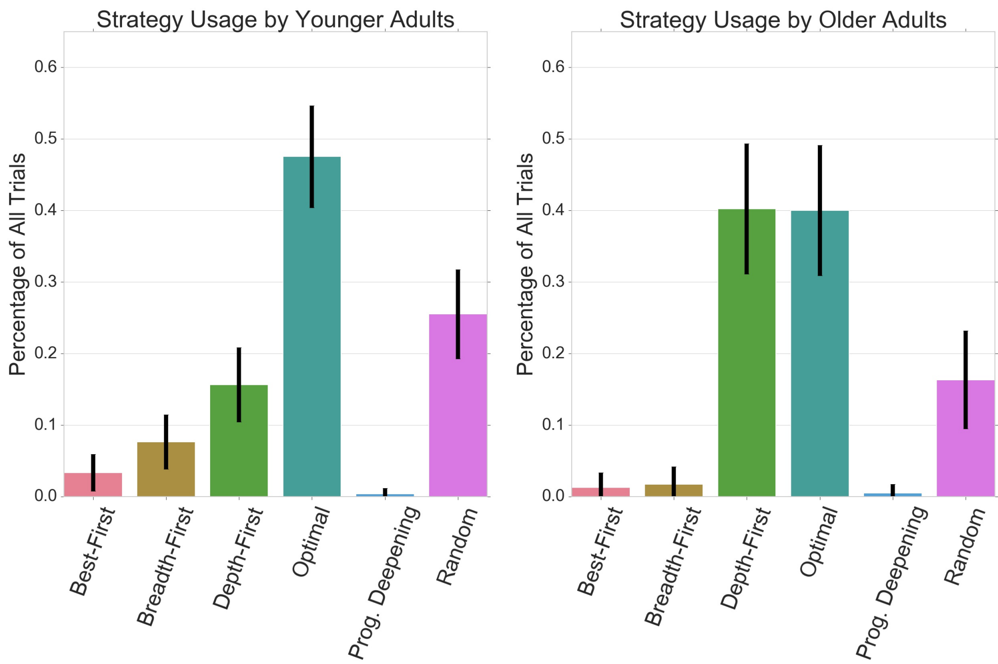

Figure 1.2: Strategy usage frequencies for younger adults versus older adults over all trials. The strategies we modeled are (from left to right): Best-first search, breadth-first search, depth-first search, optimal, progressive deepening, and random.

Results

In Experiment 1, we found that older adults differed significantly from younger adults in how often they used each of the six planning strategies introduced above ($\chi^2(5) = 205.43, p < .001$). While both age groups used the optimal strategy the most, older adults also favored the depth-first search strategy, using it almost as much as the optimal strategy (Figure 1.2). Taking a look at how participants' strategy usage evolved over time indicates that older adults were adopting the optimal strategy later in the experiment compared to younger adults (Figure 1.3). However, by the end of the experiment, the older adults were still not using the

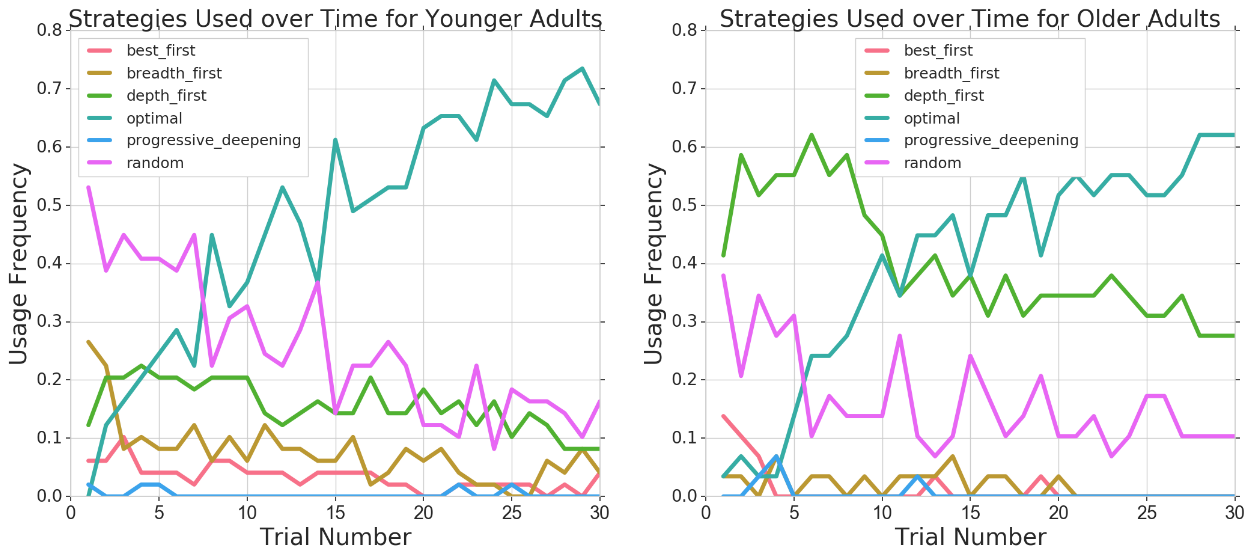

Figure 1.3: The frequencies of strategy usage for every trial in the experiment for younger adults (left) and older adults (right). The strategies shown are best-first search (red), breadth-first search (yellow), depth-first search (green), optimal (teal), progressive deepening (sky blue), and random (magenta).

optimal strategy as frequently as the younger adults (avg. frequency in the last five trials: 69.0% vs. 58.6%, $\chi^2(1) = 3.86, p < 0.05$). We also found that older adults were performing worse on the task compared to younger adults, even after discovering the optimal strategy on their own (avg. score in the last five trials: 24.7 vs. 37.4, $t(76) = -2.87, p < 0.01$). This is consistent with our expectation that using the optimal strategy less than another group will lead to lower scores.

If this difference in strategy usage is due to older participants being unaware of the existence of the optimal strategy, then it should be possible to remedy their deficits by teaching them the optimal strategy using our cognitive tutor. We tested this hypothesis in Experiment 2.

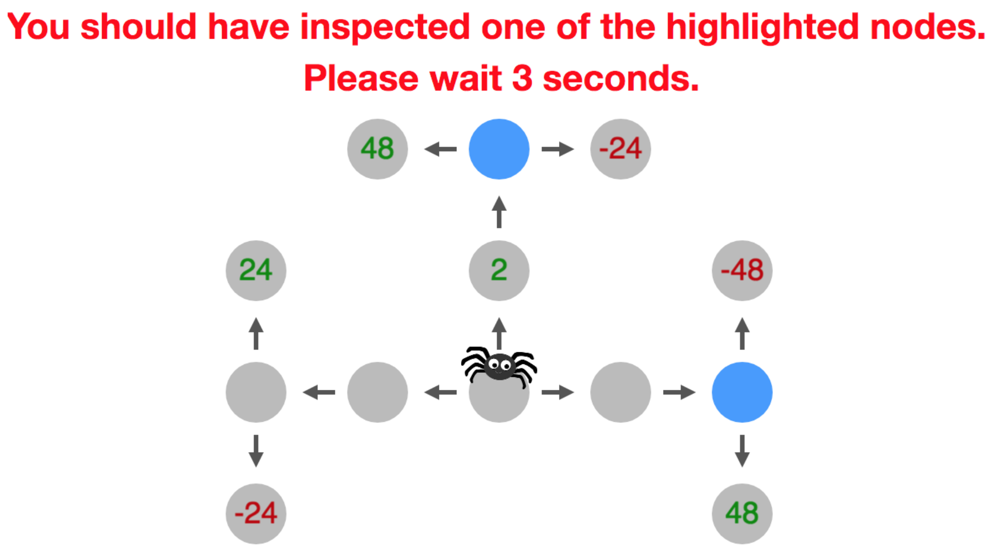

Figure 1.4: Example feedback from the cognitive tutor in the training phase of Experiment 2.

1.3 Experiment 2

Methods

Participants & Procedure

For Experiment 2, we recruited and sorted participants into two groups: younger than 25 (median = 22 years, $n = 41$) and older than 47 (median = 53 years, $n = 37$). We conducted the experiment via Amazon Mechanical Turk. Participants were randomly assigned to either train with the cognitive tutor (feedback condition: $18 - 69$ y.o, $n_{young} = 24, n_{old} = 23$) or to practice the task on their own (control condition: $18 - 68$ y.o, $n_{young} = 17, n_{old} = 14$). Participants in the control condition performed 30 trials of the Mouselab-MDP task described in the Methods section of Experiment 1. Participants in the feedback condition were first given 15 trials where they practiced the Mouselab-MDP task while receiving our cognitive tutor's optimal metacognitive feedback (Lieder et al., 2018; Lieder, Krueger, Callaway, & Griffiths, 2017). As illustrated in Figure 1.4, the tutor's feedback comprised i) a delay penalty whose duration was proportional to how suboptimal the participant's planning operation was, and ii) a visual demonstration of what the optimal planning strategy described above would have done differently. The feedback thereby supported both reinforcement learning and learning from demonstrations. Participants were then given 15 test trials of Mouselab-MDP without any feedback, identical to the trials given to the control group.

Analysis Plan

We conducted four t-tests to compare age groups and conditions on participants' score, or the average amount of reward participants earned per trial. We compared older adults' scores to that of younger adults' in the training phase, or first 15 trials, in both conditions in order to verify the age differences in performance which were found in Experiment 1. Then, we compared the scores of both age groups during the test phase, or last 15 trials, of the task. We also compared the scores of the feedback group and the control condition during the test phase. Finally, we compared the scores of older adults in the feedback condition to the scores of younger adults in the control condition during the test phase. Additionally, we fit a three-way ANOVA model to examine the effects of age, condition, and trial number on participants' scores.

Results

During the training phase of the experiment, there was no difference in the average scores of the younger adults (19.45, SEM = 4.55) and the older adults (14.27, SEM = 3.76) in the control condition ($t(29) = -0.85, p = 0.40$). However, in the feedback condition there was a difference in the average scores between younger adults (36.4, SEM = 0.85) and older adults (28.69, SEM = 1.89; $t(45) = -3.79, p < 0.001$). Moving on to the test phase of the experiment, there was no difference in the average scores of younger and older adults

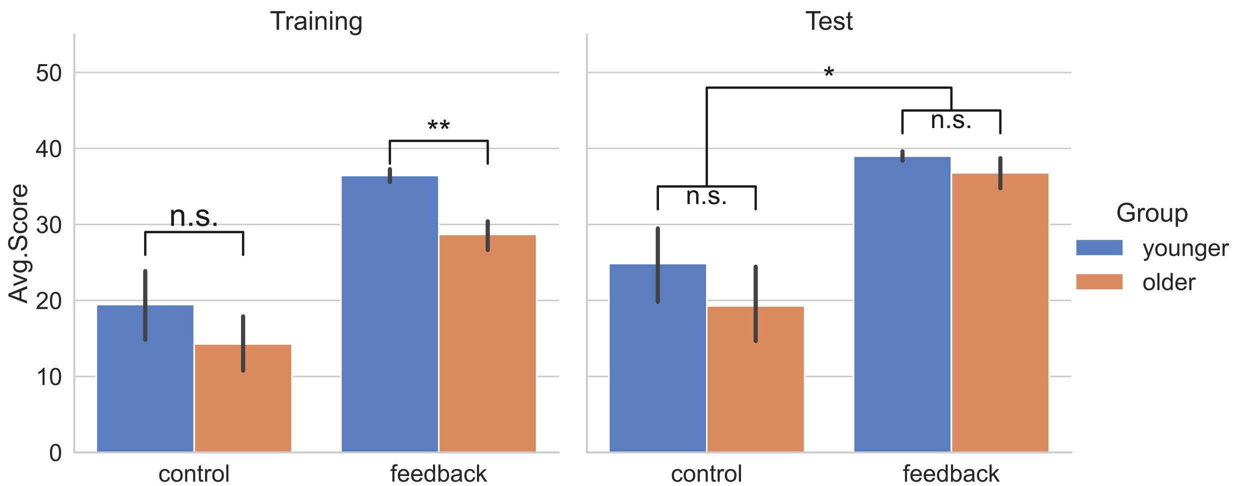

Figure 1.5: The average scores on the task for older and younger adults in both the control and feedback conditions. Participants in the feedback condition trained with our cognitive tutor during the first 15 trials of the experiment. The last 15 trials consisted of the test phase where there was no feedback given in either condition. * means $p < 0.01$ while ** means $p < 0.001$

in both the control condition ($t(29) = -0.76, p = 0.45$) and in the feedback condition ($t(45) = -1.02, p = 0.31$). Thus, we only found evidence of an age effect on performance in the feedback condition, but this went away by the end of the task, indicating that older adults learned and improved throughout the task.

However, we did see a difference in average scores between the feedback condition and the control condition in both younger adults ($t(39) = 3.31, p < 0.01$) and older adults ($t(35) = 3.56, p < 0.01$). On average, older adults scored 36.79 (SEM = 2.10) in the feedback condition compared to older adults in the control condition who scored 19.28 (SEM = 5.30), while younger adults in the feedback condition scored 38.99 (SEM = 0.63) compared to younger adults in the control condition who scored 24.85 (SEM = 5.02). These results indicate that practicing with our cognitive tutor was effective at improving decision-making skills regardless of age (Figure 1.5).

Encouragingly, for older people the benefit of training with our cognitive tutor was so large that they were able to outperform younger people in the control condition ($t(38) = 2.41, p <$

0.05). Furthermore, it appears that older adults benefited more from the cognitive tutor than younger adults. According to the results of the ANOVA, the advantage of young people gradually vanished over time in both conditions ($\beta_{\text{trial}\times\text{young}} = -0.13, F(1, 2332) = 5.21, p < 0.05$). Taken together, these results suggest that feedback from our cognitive tutor helps close the gap in performance between older and younger adults on the task.

1.4 Discussion

Older adults are known to struggle with planning but until recently, researchers had very limited ways of observing how people conducted the planning process. In Experiment 1, we used the Mouselab-MDP task and the associated planning strategy models developed by Callaway, van Opheusden, et al. (2022) to reveal the kinds of planning strategies that older adults and younger adults use when solving a route-planning task. We found that older adults and younger adults differed in how often they used particular planning strategies, such as the optimal strategy which resembles planning backward planning from the goal. Younger adults used the optimal strategy for most of their trials whereas older adults favored both the optimal strategy and the depth-first, or forward planning, strategy equally. While some older adults did discover and favor the optimal strategy by the end of the task, they still performed worse on the task compared to the younger adults who overwhelmingly used the optimal strategy by the end of the task. This finding supports the results of previous studies which found that older adults use decision making strategies at different rates compared to younger adults (Lemaire, 2010; Hinault & Lemaire, 2020).

We thought that the reason younger adults outperformed the older adults on the task was because not enough older adults had discovered the optimal strategy on their own. Therefore, in Experiment 2, we introduced a cognitive tutor which gave people feedback on their planning process, effectively teaching people the optimal, backwards planning strategy. Both older and younger adults trained with the tutor for half of the experiment and then were left to do the task on their own for the latter half. In the group that received training, there was a difference in performance between the older and younger adults in the beginning of the experiment but after training, there was no difference between the age groups by the end of the experiment. Interestingly, in the control group, there was no difference between age groups towards the beginning of the experiment nor towards the end of the experiment, which contradicts our results from Experiment 1. This may have been because of the relatively small sample size we had in the control group in Experiment 2, or perhaps more older adults in this sample were able to discover the optimal strategy earlier on. Despite this, the group who received feedback from the cognitive tutor outperformed participants in the control group regardless of age. This means that everyone was able to benefit from instruction in the optimal strategy, though older adults seemed to benefit even more. Although older adults in the control group performed about the same as younger adults, the older adults who received feedback from the tutor were able to significantly outperform the untrained younger adults.

While our sample sizes are small and our results are constrained to the particular Mouselab-MDP task, this work is an important start towards understanding the ways people plan and how age affects this particular type of decision making. The Mouselab-MDP task can be customized in order to study many different reward environments, each with their own unique optimal planning strategies. For example, the depth-first strategy that many older adults preferred in Experiment 1 would work well on an environment where rewards are increasing. In that case, there may not be any age differences in planning strategies or older adults may even outperform the younger adults. Future work can use Mouselab-MDP to understand which types of planning problems older adults struggle with and whether this is due to being unaware of the optimal strategy.

Furthermore, while it is difficult to generalize the effectiveness of our cognitive tutor beyond the Mouselab-MDP task, we believe that the overall method of giving feedback on someone's problem solving process can prove beneficial in a variety of other tasks. The metacognitive skill of how to evaluate options before making a decision can be flexibly applied to a variety of decision making problems compared to knowing how to solve a particular type of problem, but much further research is needed to know whether people can be taught such flexible metacognitive skills via digital cognitive tutors. In any case, the present tutor was able to help close the performance gap between older and younger adults so we will optimistically look towards a future where a variety of cognitive tutors can help give older adults a more even footing against their younger counterparts on a variety of problem-solving tasks.

Chapter 2

Older Adults Catch Up to Younger Adults on Cognitive Tasks After Extended Training

2.1 Introduction

"Practice makes perfect" is an old adage which applies to people of all ages. Whether someone is learning something for the first time, or rehearsing a skill once learned but unused for some time, practice or training can help people improve their current ability.

Older adults, in particular, have shown improvement and maintenance of new skills after training (Baltes, Dittmann-Kohli, & Kliegl, 1986; Dahlin, Nyberg, Bäckman, & Neely, 2008). In some cases, such as on collaborative, visual discrimination, or implicit learning tasks, older adults and younger adults perform equally well after training (Derksen et al., 2015; Ratcliff et al., 2006; Myers & Conner, 1992). However, for other tasks, such as those that require memory, response inhibition, or task-switching, a noticeable performance gap exists. Despite potentially starting out at the same performance level, younger adults often outperform older adults over time (Kliegl, Smith, & Baltes, 1989; Davidson, Zacks, & Williams, 2003; Dahlin et al., 2008; Karbach & Kray, 2009).

Can training for a longer period of time help an older person improve enough to close such performance gaps? Baltes and Kliegl (1992) had younger and older adults train for 38 sessions over a year, 18 more than their previous study (Kliegl et al., 1989), on a free-recall memory task and found that younger adults continued to outperform older adults. They also found that the older group never reached the same level of performance as that of the younger group near the beginning of training. Similarly, Noack, Lövdén, Schmiedek, and Lindenberger (2013) found that younger adults outperformed older adults on a spatial and temporal memory task after training for 100 daily sessions. These results suggest that it is unlikely for an average individual to reach the same level of performance as someone several years younger. However, most lab studies of training rarely last longer than 12 weeks, mainly due to resource limitations (Lampit, Hallock, & Valenzuela, 2014; Nguyen, Murphy, & Andrews, 2019). Perhaps this is not enough time for older adults to close the performance gap between themselves and younger adults.

We can circumvent the resource issues of traditional lab studies by using naturally occurring data that was collected online (Goldstone & Lupyan, 2016; Griffiths, 2015). In particular, we can use data from online cognitive training platforms to investigate the effect of extended training. One platform which provides this kind of data is Lumosity. Lumosity has a collection of more than 50 engaging games, some of which are based on common tasks used in lab studies, that target various cognitive processes. Each game is categorized by the cognitive domain being trained, such as attention or memory, and each gameplay lasts a few minutes. While Lumosity has yet to prove that training improvements transfer to other tasks outside of Lumosity (Simons et al., 2016), the detailed data collected on the platform is helpful for understanding how people learn and improve on these games.

Previously, Steyvers, Hawkins, Karayanidis, and Brown (2019) used data from Lumosity to examine the effects of extensive practice on task switching. They found that a small sample of older adults who played a task switching game over 1000 times were able to match or exceed the performance of younger adults who played up to 60 times. Although this study demonstrates that older adults can bridge the performance gap with their younger counterparts, the extent of training needed and the degree of benefit gained still remain unresolved questions.

In this paper, we investigate how much training older adults need to catch up to younger adults on a variety of cognitive tasks. We define "catch up" as an older adult matching or exceeding the score of a younger adult on one of the Lumosity games. We leverage the Lumosity data set used in Steyvers and Schafer (2020) which contains data from 9,923 users between 18 and 90 years old on 57 different games to examine catch up in different training situations.

One way catch up could occur is if younger adults' performance reaches an asymptote much earlier in training compared to that of older adults', but the older adults' performance reaches the same asymptote after extended training. Thus, much like previous studies, we will look at the scenario where older adults and younger adults train for the same amount of time. However, our primary focus is to what extent catch up occurs when older adults train longer than younger adults. Many older adults in our sample have performance levels which lag behind those of younger adults at the beginning of training, but by the time 200 games have been played, some of the older adults' performance levels surpass those of younger adults' at an earlier training point. This scenario is another way that catch up can occur.

The data set used in our analyses is well suited to address catch up by older adults for two main reasons: it contains data from thousands of users, with 55% of them over the age of 60, and the training data spans several years, with many users training on the same game well over a hundred times. However, as is the case with using any naturally occurring data, many users also drop out before training for very long. Since dropout is related to performance (Steyvers & Benjamin, 2019), we address this issue and its impact on the generalizability of our findings later in the paper. Despite this limitation, the data set enables us to accurately assess the degree of benefit that additional training imparts to various age groups among those who have trained for an extended duration.

2.2 Methods

Participants

The data used for analysis is the same as that which was analyzed in Steyvers and Schafer (2020), which contained 36,297 English-speaking Lumosity users located in either the United States, Canada, or Australia who primarily used the web version (as opposed to the mobile app). These users signed up between August 1st, 2013 and December 31st, 2016 and the data was collected between August 1st, 2013 and June 30th, 2019 (see Appendix A for further information on this data set). A subset of 9,923 Lumosity users between the ages of 18 and 90 at signup (under 40: 360 males, 239 females, 55 gender unavailable; 40-59: 1210 males, 1504 females, 276 gender unavailable; 60-79: 1989 males, 3056 females, 632 gender unavailable; 80 and over: 216 males, 311 females, 75 gender unavailable) were included in our analyses. No racial data was available. 15% of users had a high school diploma or completed some high school, 19% had completed some college, 25% had a bachelor's degree, 3% had an associate's degree, 25% had a postgraduate degree, and the rest declined to specify their education level. Users were included in the subset if they had played any of Lumosity's games at least 100 times. We chose 100 gameplays to ensure that users had trained for an extended period of time and to avoid noise in the data caused by dropout, or users who play for a bit and then stop playing completely (Steyvers & Benjamin, 2019). The users in our sample played a median number of 2,284 games total. Thus, our sample size of nearly 10,000 people is sufficient for investigating catch up abilities on various tasks across the lifespan.

Games

There are 57 games in the data set. The original data set labeled each Lumosity game by the cognitive domain that the game was targeting and we kept these labels to observe trends within domains (Steyvers & Schafer, 2020). The six domains are attention (12 games), flexibility (6 games), memory (21 games), reasoning (7 games), language (6 games), & math (5 games). Previous results have shown that the domains of math and reasoning show some internal consistency such that games within these domains show more correlated scores within the domain than across domains (see Figure 3 of Steyvers and Schafer (2020)). However, we should note that the domain labels used by Lumosity platform do not uniquely describe the cognitive processes involved in each game as most games involve multiple types of cognitive processes.

Preprocessing

Lumosity games each have a unique scoring system which is generally based on the user's speed and accuracy but also involves game-specific factors, leading to game scores on different scales. In order to compare the performances across games with these different scoring systems, we first normalized the game scores using a min-max transformation (Han, Pei, & Kamber, 2011). Under this normalization, scores closest to 1 are the best of the whole sample while scores close to 0 are the worst performers. This involved first setting outliers greater than 3 standard deviations above the mean equal to 3 standard deviations above the mean. We can do this because we care about the relative ranking of the scores rather than the actual value of the score. This value is now the maximum score achieved by users, so after normalization, anyone who got this score (or higher) would have their score represented as a 1. Normalization follows the formula $(score - scores_{min})/(scores_{max} - scores_{min})$. For example, if Mark plays Ebb and Flow and scores 21,400 and the best score achieved by someone in our sample in that game is 35,000 and the lowest score is 1,000, then Mark's score of 21,400 would be normalized to 0.6, which means that his score is a bit better than half of those in the sample.

After applying the transformation to all the game scores in the data set, we smoothed the learning curves for each user in each game so that user scores would more accurately reflect the current performance level and our results would be less susceptible to temporary score fluctuations. Finally, a user's age at the time of gameplay (extrapolated from time passed since signup) was added to each of their gameplay records. Further details are in Appendix A.

Data Analysis Approach

We conducted two different analyses which together help us answer our questions related to catch up by older adults. The first is a group-level analysis which looks at the learning curves for each age group. The second analysis calculates catch up probabilities at an individual level. At the outset, we should note that none of the analyses involve curve fitting or computational modeling. Previous research has investigated the learning trajectories of individual users on the Lumosity platform using exponential and power law functions that only consider the amount of practice (Steyvers & Benjamin, 2019; Donner & Hardy, 2015) as well as more complex computational models that also take into account the effect of spacing and retention (Kumar, Benjamin, Heathcote, & Steyvers, 2022). Given the aims of the current data analysis, we are not focused on explaining the underlying functional form of the learning process and instead use a much simpler approach of comparing performance levels at different levels of training.

Group performance analysis

In order to clearly visualize the learning curves of each age group for different levels of training, we grouped users into age bins that were mostly five years apart. For example, one bin would contain users from 40-44 years old while the next bin contained users from 45-49 years old. However, this only applied to the ages between 40 and 89. A few age bins were merged at the extreme ends of the age range due to insufficient data in the five year bins. In the end there was also a 18-29 bin, 30-39 bin, and 90-95 bin. This binning procedure applies only to the data presented in Figure 2.1, which visualizes the mean performance for each of these age bins. The rest of the results presented in the text follow the analysis procedure explained in the next section.

Catch up analysis

Our main analysis focused on catch up, or the idea of whether an older adult who has trained for a while can match or exceed the performance of a younger adult. For each game, we computed the catch up probabilities across pairs of age groups. To calculate this probability, we looked at all individual pairwise comparisons between the two age groups and calculated the proportion of older adults that had a higher score than a particular younger adult, for every adult in the younger group. Depending on the game and age groups, pairwise comparisons numbered anywhere from 160 to over 4.6 million. The resulting probability is akin to the probability of randomly sampling an individual from one age group and an individual in another age group and observing that the older individual has a higher score.

We chose to look at catch up in this manner as opposed to comparing group performance means because there is a lot of individual variability among people in the same age group.

We started calculating catch up probabilities from 20 gameplays, when users have reasonably learned how to play the game, and continued in 20 gameplay increments until 100 gameplays, beyond which the analysis would suffer from insufficient data. We did so in order to calculate how the amount of training relates to catch up probability.

Catch up probabilities were calculated separately for each Lumosity game. Not all Lumosity games are equally popular, so games that had less than ten users from an age group of interest were excluded from that age group's catch up analysis. Thus, instead of all 57 games in the original data set, the number of games included in our catch up analysis ranged between 32-36 games, depending on the age comparison (10 attention, 4-5 flexibility, 8-10 memory, 4 reasoning, 2-4 language, and 3 math games).

In order to have enough data to directly compare two age groups, we grouped users into larger age bins of ten years. Thus, we had the following bins: 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80-89. The exceptions are at the extreme ends of the age range: 18 and 19 year old users are included with the 20-29 year old users and the catch up results for those 90-95 are not reported since there was only one game which had at least ten users from this age group.

2.3 Results

We present our results according to the two scenarios of catch up discussed in the introduction: the scenario where older and younger adults train an equal amount, and one where the amount of training is unequal, with older adults training longer.

To answer our first question of whether older adults can catch up with the same amount of training, we calculated the catch up probability of the older group when both the older group and the younger group have trained for 100 gameplays. For our second question concerning how different amounts of training affects catch up, we compared older adults who trained for 20, 40, 60, 80, & 100 gameplays to younger adults who only trained for 20 gameplays. After calculating catch up probabilities for each game, we looked for age equivalence, which we define as a 50% or greater chance for a randomly sampled older adult to score better than a randomly sampled younger adult.

Throughout the rest of this paper, we will use the decade marker as a shorthand for each age bin. For example, ”60s” refers to those users between 60 and 69 years of age while ”70s” includes those between 70 and 79 years. The singular exception is the ”20s” group which also includes 18 and 19 year old users along with those between 20 and 29 years.

Additionally, we report Bayes factors (BFs) for our analyses, which were computed using Pingouin (Vallat, 2018), since they are easier to interpret over p values (Kass & Raftery, 1995). Following the notation for the alternative hypothesis (1) against the null (0), $BF_{10} > 1$ indicates evidence for the alternative hypothesis while $BF_{10} < 1$ indicates evidence for the null hypothesis. The value of the Bayes factor increases with the likelihood of the alternative hypothesis. For example, $BF_{10} = 10$ means that the data are 10 times more likely under the alternative hypothesis compared to the null hypothesis. Generally, BFs between 3 and 10 indicate moderate evidence against the null hypothesis, and BFs greater than 10 indicate strong evidence against the null (Kass & Raftery, 1995).

Performance gaps persist after equal training

First, we looked at the scenario where older and younger adults both train for an extended amount, in case older adults simply need more time in order to catch up to younger adults.

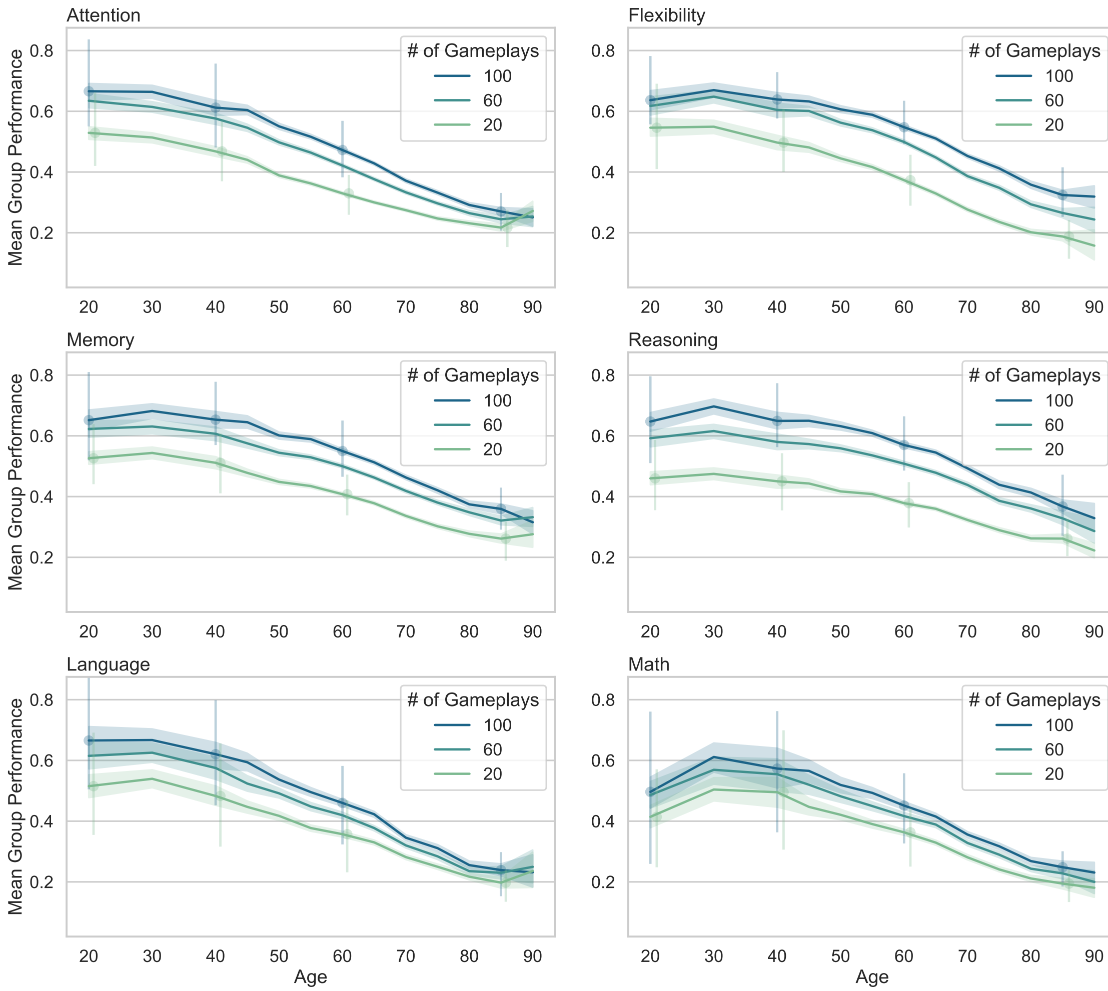

Figure 2.1: Mean performance for each age group on Lumosity games grouped by cognitive domains, plotted for three different levels of training (20, 60, and 100 gameplays). Game scores were normalized between 0 and 1 such that performance values closer to 1 indicate the best game performance. Error bands represent the 95% confidence interval while error bars describe the 25% and 75% percentiles of the data to show the range of scores.

When older adults and younger adults train up to 100 gameplays, the mean score for nearly all age groups improves regardless of cognitive domain (BF ¿ 10 on paired t-tests; Figure 2.1). The exceptions were people over 90 on attention, memory, language and math games and people in their 20s and 40s on math games (BF ¡ 10 on paired t-tests; see Appendix A for training improvements). However, despite training for 100 gameplays, most older age groups continued to score lower on average compared to younger groups (Figure 2.1).

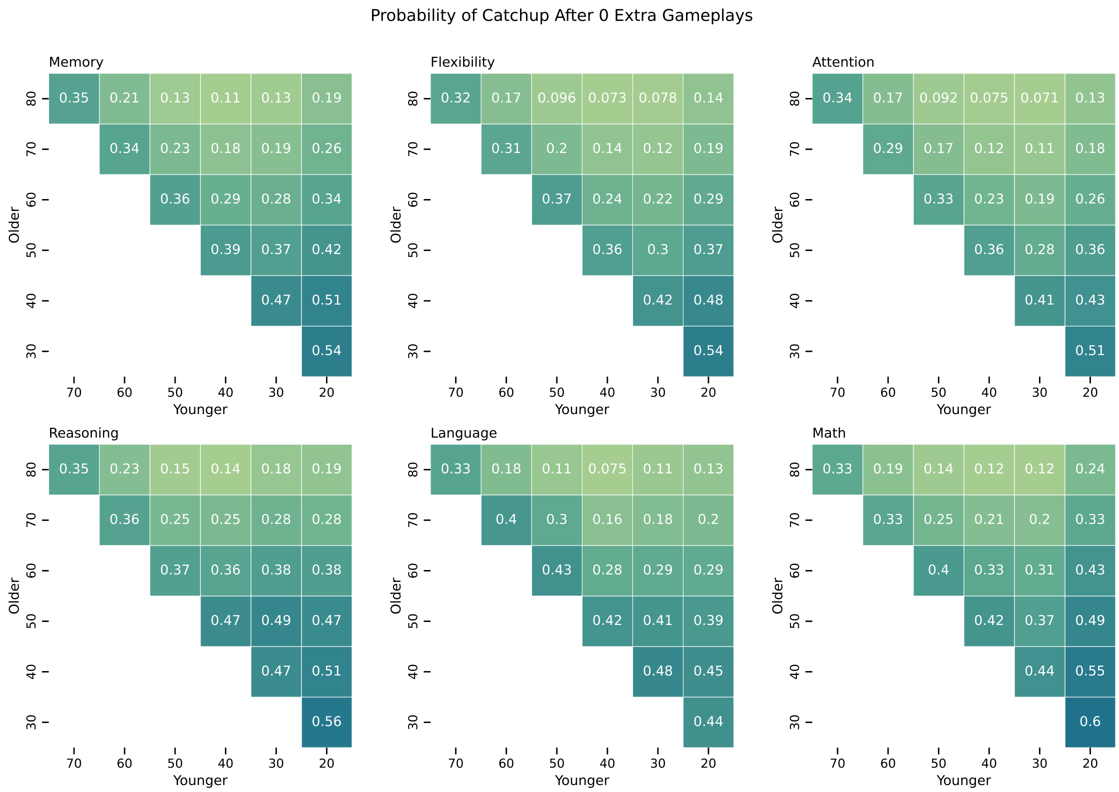

Figure 2.2: The probabilities of an older adult catching up to a younger adult after both have played a game 100 times. A value of 0.50 or greater indicates age equivalence. Age bins are ten years wide such that "30" represents those users 30-39, "40" means 40-49, etc. "20" also includes those who are 18 or 19. The diagonal represents the smallest age difference (≤ 10 years) amongst the age comparisons, while the upper right corner represents the greatest age difference (80s vs 20s).

When examining the catch up probabilities, age equivalence was observed for one game, a vocabulary game called Taking Root, when comparing the 70s group to the 60s group. We found no age equivalence when we compared older adults in their 70s and 80s to individuals 20 years their junior at the same training level of 100 gameplays (Figure 2.2). However, the probability of catch up for these older adults was significantly greater than zero on attention, flexibility, memory, and reasoning games ($BF_{10} > 10$ on one-sample Bayesian t-tests) suggesting that some older individuals have the ability to catch up to adults who are nearly twenty years younger.

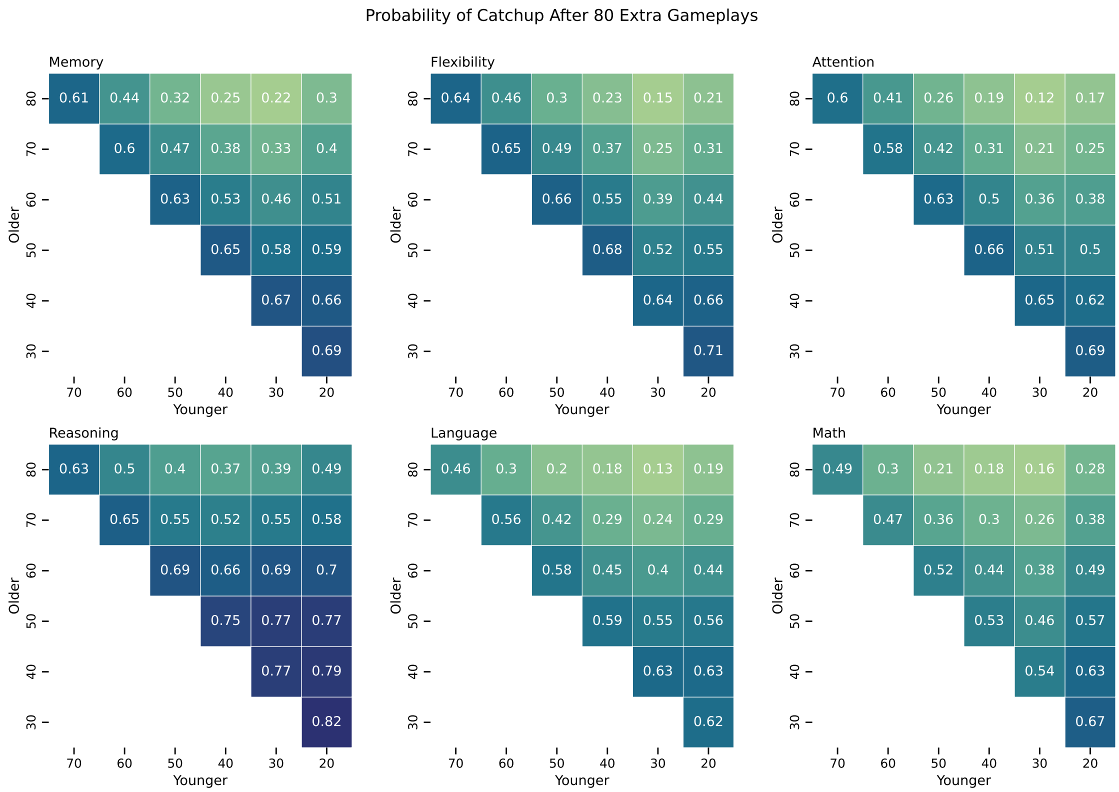

Figure 2.3: The probabilities of an older adult catching up to a younger adult on a game after the younger adult has played 20 times and the older adult has played 100 times. A value of 0.50 or greater indicates age equivalence. Age bins are ten years wide such that "30" represents those users 30-39, "40" means 40-49, etc. "20" also includes those who are 18 or 19. The diagonal represents the smallest age difference ($\leq$ 10 years) amongst the age comparisons, while the upper right corner represents the greatest age difference (80s vs 20s).

Unequal training promotes catch up

Next, we shift our focus to the primary question concerning the impact of unequal training on the catch-up potential for older adults. In the analysis, we assessed the performance of two groups of adults, a younger group after 20 gameplays and an older group after 20, 40, 60, 80, and 100 gameplays (corresponding to 0, 20, 40, 60, 80 extra gameplays respectively). Figure 2.3 shows the catch up probabilities for a subset of comparisons: older adults who trained for 80 extra gameplays relative to younger adults (more detailed results are shown in Appendix A). These results generally show that additional training makes it possible for the older group to catch up to the performance of slightly younger groups, although the potential for catch up is limited for the largest age differences.

When the amount of additional training increases from 0 to 80 extra gameplays, the probability of an older adult in the 60s, 70s, and 80s groups catching up to someone in a 20 years younger group increases. Across all games, the increase in catch up probability for these groups is substantially different from 0 ($BF_{10} > 10$; one-sample t-test to quantify the evidence for catch up). When comparing different domains, there is substantial evidence for catch up on attention, flexibility, and memory games ($BF_{10} > 10$) as opposed to moderate evidence for math ($BF_{10} > 5$) and reasoning ($BF_{10} > 3$ except for the 80s group). Evidence for catch up on language games was very weak ($BF_{10} < 3$), though this may be due in part to the small number of language games included in the analysis.

For the 80s vs 60s, 70s vs 50s, and 60s vs 40s comparisons, the average increase in catch up probability ranges from 0.22 to 0.32 on attention, flexibility, memory, and reasoning games. The increase for language and math games is between 0.09 and 0.15. The catch up gains are greater when comparing adults to those in a 10 years younger group. For this comparison, the average improvement in catch up was 0.30 for attention, flexibility, memory, and reasoning, compared to 0.16 for language and math ($BF_{10} > 50$ for attention, flexibility, memory; $BF_{10} > 5$ for reasoning, language, math).

The prevalence of age equivalence also increases when we allow the older adults to train more than younger adults. When we compared the 60s, 70s, and 80s groups to groups that were 20 years younger, then equivalence was possible after 100 gameplays in 3 games for the 80s vs 60s comparison. These games included a Stroop task (Color Match), a face-name recall task (Familiar Faces), and a planning task (Pet Detective). Equivalence was also observed in 9 games for the 70s vs 50s comparison and in 19 games for 60s vs 40s (see Appendix A).

2.4 Discussion

In this paper we used a data set with cognitive training scores from almost 10,000 people ages 18-90 to investigate whether a longer period of training helps to close the commonly observed performance gap between age groups. We were interested in how much extra training would help older adults catch up to younger adults on these cognitive training tasks. When older adults train more than younger adults, up to 100 training sessions, we found evidence of the performance gap diminishing as the older adults catch up to the younger adults. In some cases, with unequal training, the performance gap completely disappears.

Additionally, we looked at whether catch up would occur after equal amounts of extended training for both older and younger groups. Like Baltes and Kliegl (1992), our results also showed that age differences continued to persist even after both age groups undertake extended practice on the task. Unlike this previous study, which compared people in their 70s to people in their 20s on a memory task, we found that when people in their 70s trained up to 100 sessions, the mean group score overlapped with the mean score of people in their 20s who had only trained for 20 sessions on a memory game (see Figure 2.1). These discrepancies may be due to the greater number of extra practice sessions (80 compared to 18) and the larger sample size in our study. Additionally, in the unequal training case, we were able to find instances where little to no age differences were observed between adults who were closer in age, consistent with Steyvers et al. (2019).

One reason we found robust increases in catch up probability for the attention, memory, and flexibility domains and not in others is that the domain categories used here may not accurately represent the underlying cognitive processes that these games try to target. For example, many of the language and math games test vocabulary knowledge and simple arithmetic under time pressure and so performance on these games may be greatly influenced by response speed rather than domain knowledge. In addition, even though some of the games in the attention, memory, and flexibility categories are directly modeled after classic lab tasks used to assess these cognitive abilities, only the math and reasoning games were previously found to have high internal consistency (Steyvers & Schafer, 2020). Therefore, care must be taken when interpreting these results at the domain level and may be more informative at the single game level.

One important limitation of this work is that it is not clear to what degree our results will generalize to the full population. To address potential concerns arising from comparing groups with unequal total training amounts, as noted by Steyvers and Benjamin (2019), we confined our sample to players who had completed a minimum of 100 gameplays. However, this subset is not fully representative of the full population of Lumosity players as the older players tend to persist longer. Consequently, our sample exhibits a bias towards older players: for example, our sample's average age is 64.8 years, compared to the 61-year average age among players with a minimum of 20 game sessions. When working with longitudinal data that spans a few years, dropout is inevitable and limits the generalizability of the results. Thus, while we have been careful to control for the amount of practice in the sample, our results only hold for people who persist to 100 gameplays.

Furthermore, while we have used number of gameplays as a measure of training, it is difficult to directly compare the amount of training done on the online training platform to that conducted in labs during cognitive training studies. Each gameplay of a particular game only lasts a couple of minutes and the player is free to play them whenever they'd like, often taking a month just to play 20 times, whereas participants in lab studies come in for training sessions on a rigid schedule with each session lasting anywhere from 15 minutes to 2 hours (Lampit et al., 2014; Reijnders, van Heugten, & van Boxtel, 2013). However, our results are compatible with lab-based studies which find that participants improve on the trained task over time (Anguera et al., 2013; Verhaeghen, Marcoen, & Goossens, 1992; Baltes & Kliegl, 1992; Kliegl et al., 1989).

While our analysis has demonstrated that it is possible for older adults to match younger adults on task performance, this is only a start. One extension of this work would be to use naturally occurring data for other tasks to build a model which could predict a person's performance in the future based on their current learning trajectory, similar to work done by Steyvers and Schafer (2020). Such a model could help inform older adults how much more practice they require to reach their target performance level (which might be expressed in terms of a younger age group's performance). In addition, future studies can analyze other factors which might affect an individual's catch up rate, such as the frequency of their practice sessions (Kumar et al., 2022) or particular game features. Future lab based training studies can use our catch up probabilities to inform study design and estimate the magnitude of the expected results.

In conclusion, some older adults who persist in extended training have the potential to match younger adults on a subset of short cognitive tasks even when younger adults outperform them initially. The key seems to be for older adults to train for much longer than the younger adults. Additionally, we quantified the degree of benefit gained from different amounts of training. These results, along with future studies, can help us form a more complete picture of how age differences can be overcome with additional practice.

Chapter 3

Older Adults Invest Effort Similarly to Younger Adults When Learning

3.1 Introduction

When people make decisions to learn something new, they often engage in a cost-benefit analysis where they consider whether the benefit of the new skill or information is worth the cost of the mental effort and time required to learn it (on top of any other ancillary costs). For example, a person may be introduced to a new technology that will help them work more efficiently. However, the person must learn how to use this technology well in order to reap the benefits. If they decide to try it out while working on an important task, the process of learning will slow down this person's current pace of work. Thus, a decision must be made as to whether investing the time and effort into learning this new technology now is worth the future benefits of the task being completed more efficiently.

People perform a cost-benefit analysis in these kinds of situations because mental effort is unpleasant, but they are willing to expend that effort when the rewards, or benefits, are sufficiently high (Kool & Botvinick, 2018). Though mental effort is difficult to measure directly, some methods for estimating relative amounts of people's mental effort include analyzing their response time differences in task switching experiments or classifying the strategies they use in decision making experiments, with some strategies requiring less mental effort than others.

Previous work has shown that the consideration of costs and benefits differs between older and younger adults. For example, older adults are more sensitive to costs compared to younger adults on tasks requiring mental effort (Westbrook, Kester, & Braver, 2013; Hess, Smith, & Sharifian, 2016). Older adults are also less likely than younger adults to use cognitively effortful decision making strategies (Mata et al., 2007; Worthy & Maddox, 2012; Hinault & Lemaire, 2020). These behaviors are not due to a lack of motivation, since older adults' use of model based strategies isn't affected by high rewards (Bolenz et al., 2019). Furthermore, older adults will expend more effort than younger adults when rewards are easier to obtain (Devine et al., 2021). One explanation is that older adults have limited cognitive resources and that they are exhibiting resource-rational behavior.

The research presented in this chapter will continue to probe the relationship between these costs and rewards over the lifespan, but on a new type of task. Previous mental effort studies use tasks that either do not require much effort to learn (such as task switching experiments involving perceptual decisions) or do not give participants much control over what to learn. There aren't many experimental tasks that approximate learning a complex tool and then using that tool to achieve a goal, as in the example of learning a new technology to help with one's work. Thus, we have designed such a task, called the IntelliBaker task, to fulfill this need.

(a) An example of the feedback given after a suboptimal guess.

(b) An example of the feedback given once the best settings are found.

Figure 3.1: Some examples of the feedback screens in the IntelliBaker task featuring an IntelliBaker with three knobs.

The IntelliBaker

The main task consists of participants learning how to manipulate knobs on a futuristic kitchen appliance called the IntelliBaker. The IntelliBaker produces cakes of varying quality based on the settings of these knobs. Participants are given no instructions on how to set the knobs to produce the best results, but are encouraged to find this setting while baking cakes with the IntelliBaker. Feedback is given through a star rating system, with better cakes earning a higher number of stars (Figure 3.1). Participants who can skillfully manipulate the machine and produce high quality cakes can progress faster through the experiment compared to less skillful participants.

The IntelliBaker works similarly to the code breaking game Mastermind where the goal is to find the correct combination of some numbers. In this case, rather than numbers, participants must find the correct combination of knob settings on the machine. One particular setting of these knobs will yield the best cake possible that the machine can produce. Other settings result in cakes of differing qualities, and the amount of stars awarded is determined by the distance between the setting the participant tried and the actual best setting for the machine.

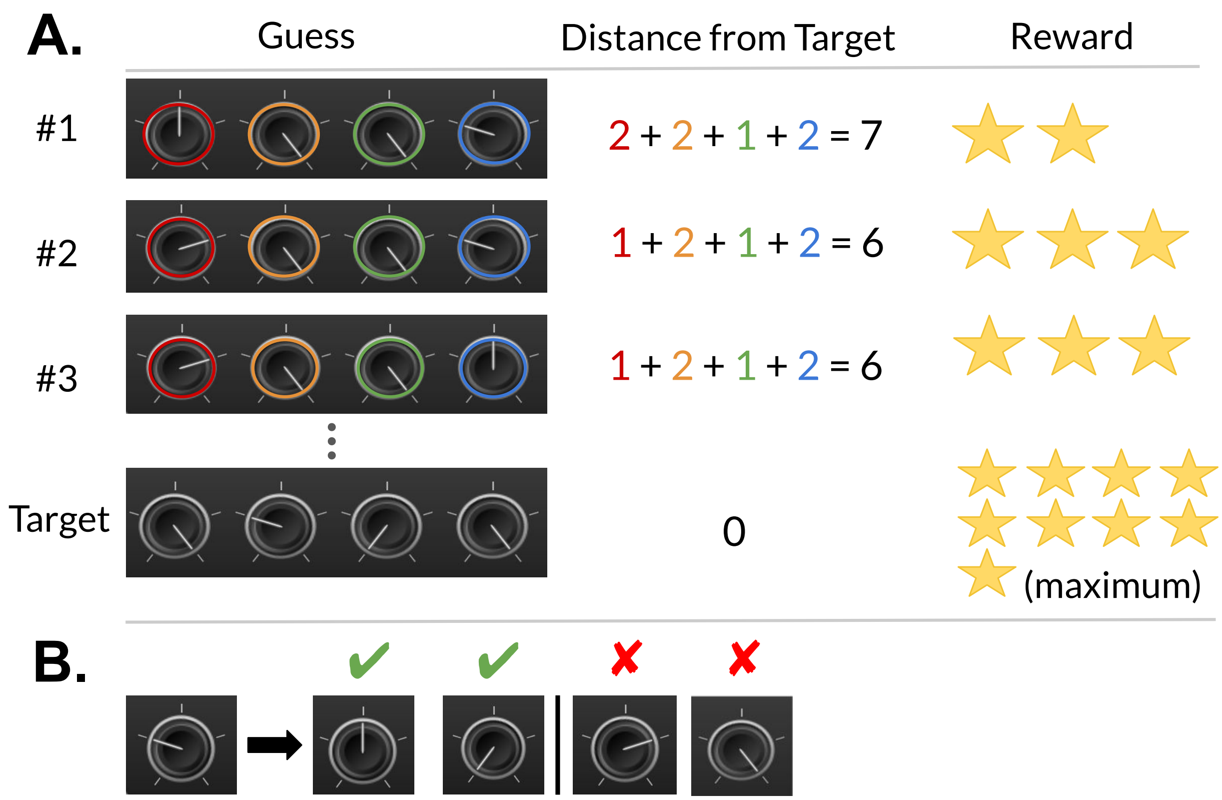

Figure 3.2: (A) An example showing three guesses tried in succession on an IntelliBaker with four knobs. Rewards are earned based on the distance of the guess from the target settings. This distance is a linear sum of the number of turns each knob needs in order to reach the correct position in the target setting. The rewards in this example follow the mapping in Figure 3.3. (B) Each knob can only be turned one setting to the left or right from its position during the previous guess.

The smaller the distance, the better the cake, and the more reward the participant gets per trial. Figure 3.2A shows a detailed example of how a participant's attempts translate into distances and rewards.

The mechanics of the IntelliBaker

In experiments, the number of knobs on the IntelliBaker may vary, but the number of settings per knob is 5. 2 settings away is the maximum distance someone can be from the correct knob setting per knob. Depending on the number of knobs on the machine, the maximum distance for the machine is 2 * number of knobs + 1. The initial knob settings are randomized such that the distance away from the correct setting lies uniformly between the max distance and the max distance - 3. This is to ensure that participants start at a relatively low reward of 1 to 3 stars should they decide to try the initial setting. Furthermore, to prevent participants from getting lucky and earning high rewards from random guesses, the movement of the knobs is restricted such that the knobs will only turn one notch forward or backward per trial. This is to encourage a more incremental and systematic approach towards earning higher rewards and getting the correct combination of settings (Figure 3.2B).

The Experiments

We designed two experiments using the IntelliBaker to answer the question of how people allocate their mental effort towards learning new tools and to investigate how expected rewards influence this effort allocation.

In order to induce different amounts of mental effort in our experiment, we created three different types of IntelliBakers which varied by the number of knobs each machine had. The more knobs on an IntelliBaker, the more combinations of settings there are to try and the more difficult it is to get high rewards from it. However, we also ensured that more difficult IntelliBakers could reward participants with more stars than easier IntelliBakers once the participant invested effort to learn it. If people are very sensitive to the balance of effort and rewards, then we expect that in order to conserve mental effort, people will try to avoid the hardest IntelliBaker, but will be enticed enough by higher reward to also resist using the easiest one and settle for learning the medium difficulty level IntelliBaker. In Experiment 1A, we test how good people are at learning to use these three different IntelliBakers and that the difference in difficulty between the machines is salient.

We are also interested in the question of whether older adults follow a resource-rational approach to this task and expend most of their mental effort on learning easier tools compared to younger adults. While we can observe how well older adults learn to use the IntelliBaker in Experiment 1A, we directly tackle this question in Experiment 1B by explicitly giving participants the choice of which type of IntelliBaker to use.

3.2 Experiment 1A

Experiment 1A is for determining how long it takes people to learn to use the IntelliBaker at varying levels of complexity. How long it takes for participants to learn the IntelliBaker is measured by the number of attempts, or trials, a participant uses to find the correct settings for the IntelliBaker. There are three different types of IntelliBaker machines which we have named machine A, B, & C for the purposes of reporting. We expect that it will take participants the fewest number of attempts to solve machine A, which should be the easiest machine to solve. We also expect that it will take participants the most number of attempts to solve machine C, with the number of attempts needed for machine B falling somewhere in between the number needed for A and C. Furthermore, we hypothesize that older adults will need more attempts on average to solve the machines compared to younger adults.

1A: Methods

Participants

Participants were recruited from the online platform Prolific. A person could take part in the experiment if they lived in the United States of America, were fluent in English, and had an approval rate of at least 90% on the platform. If the person was between the ages of 18 and 40 years old, they were assigned to the younger adult group and if the person was 60 years or older, then they were assigned to the older adult group. Due to participant feedback during the piloting phase, the two groups were given a different estimate of the time needed to complete the experiment. We made this change so that all participants felt they were being fairly paid at a rate of at least $12/hr. As a result, the younger group were given a time estimate of 15 minutes while the older group were given an estimate of 30 minutes. In actuality, participants were free to spend as much time as they wanted on the task.

49 participants took part in experiment 1A. 1 participant was excluded for not understanding the task (their data showed that they tried the same settings repeatedly for the majority of trials), leaving 48 participants. There were 24 participants in both the younger and older adult groups. The age range of the younger adult group was 21-40 years (mean 29.33 years) while the range of the older adult group was 60-76 years (mean 65.08 years).

Procedure

Experiment 1A asked participants to solve three different types of IntelliBaker machines. These types varied in the number of knobs and difficulty. Machine A had three knobs, gave a max reward of 7 stars, and rewards followed a 1-to-1 linear function where a distance of 0 (the correct settings) gave 7 stars, a distance of 1 gave 6 stars, and so on, with 1 star being the lowest reward one could earn. Machine B had four knobs, gave a max reward of 9 stars, and rewards followed the same structure as Machine A, with 9 stars given for a distance of 0 and 1 star given for a distance of 8. Machine C had five knobs, gave a max reward of 11 stars, but the reward structure differed from the other two machines. Any guesses that were a distance of 7 to 10 away gave 0 stars as a reward, while distances smaller than 7 paid out linearly such that a distance of 6 resulted in 5 stars, a distance of 5 resulted in 6 stars, up until 11 stars for a distance of 0 (see Figure 3.3). This reward structure was implemented to adhere to our experimental need for a machine that is unrewarding until effort is invested into learning it, at which point the rewards make up for the effort.

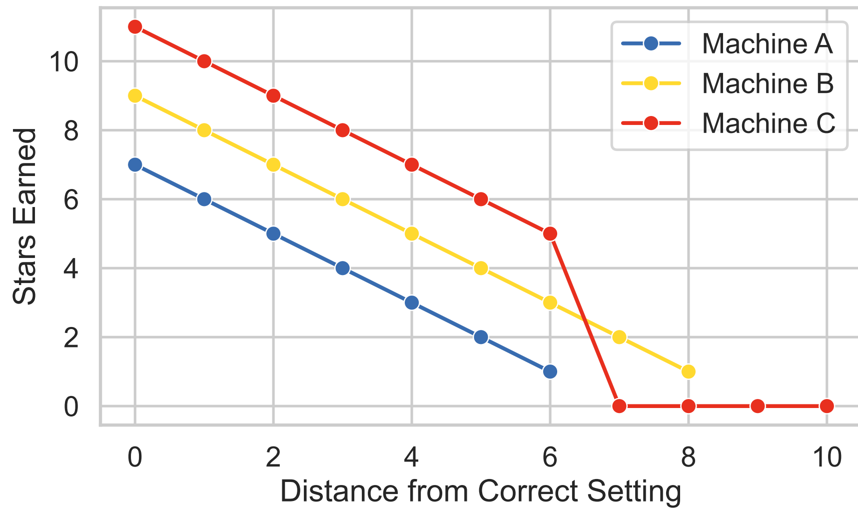

Figure 3.3: The reward structures for the different machines. Each machine awarded a number of stars relative to how close a participant's guess was to the correct setting. A distance of 0 represents guessing correctly and awards the highest possible amount of stars for that particular machine.

This experiment had 9 blocks total, with each block featuring one type of machine. The three different machines were shown three times each. The blocks were presented in sets of three, where each set had the order of the machines randomized. The participant was not aware of the differences between sets and viewed the experiment as a progression of 9 blocks. Within each block, participants were given 30 trials, or attempts, to solve each type of machine. If participants were able to find the correct settings in under 30 attempts, then they were able to move on to the next block immediately. Otherwise, they moved on to the next block after 30 attempts.

Analysis Plan

The number of attempts each participant needed to solve a particular type of machine (A, B, or C) was averaged across the three blocks in which it appeared. If the participant did not solve the machine, this block was omitted from the average. To account for blocks when the participant did not solve the machine, their probability of success in solving for each machine was calculated as well. We used a Bayesian A/B test to compare younger adults' and older adults' probabilities of success on machines A, B, & C (Hoffmann, Hofman, & Wagenmakers, 2022). We used a Bayesian paired samples Wilcoxon signed-rank test to compare the two age groups on their average number of attempts needed to solve for these machines (van Doorn, Ly, Marsman, & Wagenmakers, 2020). These analyses were done using JASP (JASP Team, 2023).

In the results section, we will report Bayes factors (BFs) for our analyses, since they are easier to interpret over p values (Kass & Raftery, 1995). Following the notation for the alternative hypothesis (1) against the null (0), $BF_{10} > 1$ indicates evidence for the alternative hypothesis while $BF_{10} < 1$ indicates evidence for the null hypothesis. If the alternative hypothesis is one sided, then the notation changes to BF_{+0} or BF_{-0}, depending on if the alternative hypothesis is that effect is greater or less for one group over the other. The value of the Bayes factor increases with the likelihood of the alternative hypothesis. For example, $BF_{10} = 10$ means that the data are 10 times more likely under the alternative hypothesis compared to the null hypothesis. Generally, BFs between 3 and 10 indicate moderate evidence against the null hypothesis, and BFs greater than 10 indicate strong evidence against the null (Kass & Raftery, 1995).

1A: Results

One of the main goals of experiment 1A was to ensure that participants could learn to solve the three types of IntelliBaker machines and that the three types were significantly different in terms of how difficult they were to solve. Once we verified this, the other goal of the experiment was to check for age differences in the capability to solve these machines.

Of the 48 participants, 2 in the younger group were unable to solve any of the machines even once. 2 participants in the older group were also unable to solve the machine in 8 out of 9 blocks, suggesting that these participants did not learn how to solve the IntelliBaker either. The remaining 44 participants were able to solve machines A and B at least once and 35 of these 44 participants were successful in solving machine C. Of the 9 participants who struggled with machine C, 2 were younger adults and 7 were older adults.

Since some participants were unable to solve a machine more than once, we calculated everyone's probability of success to get a better idea of their mastery of the IntelliBakers. The probability of success on machine A was 83.3% for both younger and older adults ($BF_{10} = 0.40$). For machine B, the probability of success was 70.8% for younger adults and 75% for older adults ($BF_{10} = 0.40$). Finally, the probability of success for machine C was 50% for younger adults and 37.5% for older adults ($BF_{10} = 0.90$). Thus, there was no statistical difference between the age groups in their ability to solve the machines.

On average, it took participants 12.53 attempts (std 5.8) to solve machine A, 16.18 attempts (std 5.46) to solve machine B, and 19.15 attempts (std 5.82) to solve machine C. The differences between the number of attempts were statistically significant within subjects on a paired samples Wilcoxon test ($BF_{10} > 30$), which means the machines were qualitatively different. However, there were no statistically significant differences between the two age groups in the number of attempts needed to solve any of the machines. This indicates that on average, older adults can use the machines as well as younger adults.

1A: Discussion

The IntelliBaker task is a brand new task created to assess how people learn to use new tools when given a choice between various tools. Thus, we first ran experiment 1A to test the validity of using the IntelliBaker in future experiments where participants of various ages would be asked to choose between the different machines. The experiment had two goals: the first was to validate that both young and older adults are able to show mastery of the

IntelliBaker and that the three different versions were different in difficulty, while the second goal was to test for age differences in mastery of the different versions.

We found that a majority of participants, both young and old, were able to solve for the correct settings on three different versions of the IntelliBaker. Furthermore, each version of the machine differed from the others in terms of difficulty, with machine A taking the fewest attempts to solve, machine C taking the most attempts to solve, and machine B was somewhere in between. Taken together, these results validate the use of the Intellibaker task in future experiments testing how the difficulty of learning a new tool affects people's decisions on whether to learn said tool.

Additionally, we were unable to find much evidence for age differences in the number of attempts used to solve a machine nor the probability of success of solving a particular version of the machine. We interpret these results to mean that both older and younger adults are equally capable, on average, of solving the different types of IntelliBakers.

3.3 Experiment 1B

After we validated the use of IntelliBakers in experiment 1A, it was time to conduct an experiment where people could freely choose which IntelliBaker they wanted to learn to use and then use the IntelliBaker to work towards a goal, much like learning a new tool in order to complete a specific task. We accomplished this by presenting machines A, B, and C all at once to the participant and instructing them to use any of the machines to earn a total of 150 stars. Crucially, participants were only given limited exposure to the three types of IntelliBakers before being asked to use them in the main task, but participants could switch which machine they used at any time. This was to ensure that the majority of learning took place at the same time as participants were making decisions about which machine to use.

Furthermore, the main task didn't have a set number of trials, ending only when participants reached a set number of stars. In this way, experiment 1B was designed to have participants assess how much time to spend using a machine in order to efficiently complete the task and to understand their preferences regarding the different IntelliBakers.

While the current experiment bears similarities to task switching experiments, there are also key differences from a classic task switching experiment (e.g. Kray & Lindenberger, 2000). One important difference is that the participant, rather than the experimenter, is in control of when the switch happens, if at all. Switch costs are generally reduced when subjects are aware of an upcoming switch and can make mental preparations for the task switch (Kramer et al., 1999; Kray & Ferdinand, 2014). Another key difference is that the different tasks, or IntelliBakers, take a nontrivial amount of time and effort to learn. Unlike classic task switching experiments where the goal is to study switch costs, the design of the current experiment promotes learning of a particular IntelliBaker and discourages frequent switching.

We hypothesized that older adults would switch machines less times compared to younger adults. Additionally, we expected that older adults would spend more time on easier machines compared to younger adults. For example, older adults might use machines A and B more often than younger adults who spend most of their time using machine C.

1B: Methods

Participants

Participants were recruited from the online platform Prolific. A person could take part in the experiment if they lived in the United States of America, were fluent in English, and had an approval rate of at least 90% on the platform. If the person was between the ages of

18 and 40 years old, they were assigned to the younger adult group and if the person was 60 years or older, then they were assigned to the older adult group. Both groups were told the estimated time for the task was 15 minutes and were paid at a rate of roughly $12/hr. Participants were free to spend as much time as they wanted on the task.

60 participants took part in experiment 1B, with 30 participants each in the older adult group and the younger adult group. The age range of the younger adult group was 20 - 38 (mean 30.27 years) while the age range of the older adult group was 60 - 79 (mean 65.27 years).

Procedure

This experiment made use of the same three Intellibaker machine types (A, B, & C) as described in the Methods section of experiment 1A. There were three practice blocks followed by the main task. The practice blocks had 10 trials each and were arranged in the order of machine A first, machine B second, and machine C third, to acquaint participants with the increasing difficulty. As in experiment 1A, participants moved on to the next block as soon as they found the optimal setting for the current machine or after 10 attempts, whichever came first.

The main task presented all three types of machines at once, arranged on the screen in a pyramid pattern (see Figure 3.4). The position of the machines on the screen was randomized between participants and participants were informed that they were seeing new machines with different combinations of settings than those they had interacted with during the practice block. Participants were also informed that they were free to use any of the machines at any time in order to earn a total of 150 stars, after which the task would end. The goal of 150 stars was chosen based on twice the median number of trials participants needed to solve machine B in experiment 1A (15) multiplied by the average reward earned on those

Figure 3.4: A screenshot of the main task in experiment 1B. Participants could use any of the machines to try to reach the goal shown in the upper left.

trials (5 stars). This was to allow participants enough time to explore the different machines and also have time left to find the best settings for either machine A or B.

Analysis

To answer the question of whether older adults switched less compared to younger adults, we compared the average number of switches between the older and younger age groups using a Bayesian Mann-Whitney U test (van Doorn et al., 2020). For the Mann-Whitney U test, the effect size δ was assigned a Cauchy prior distribution with $r = 1/\sqrt{2}$, truncated to only allow negative values due to our one-sided hypothesis. For our second hypothesis regarding older adults' preference for easier machines, we measured preference for each machine as the number of trials a participant spent using a particular machine out of the total number of trials participants did to finish the main task. We then used a Bayesian A/B test to compare the group preferences for each machine (Hoffmann et al., 2022). For machines A and B, we assigned a standard normal prior to the log odds ratio ψ and compared the null hypothesis of $\psi = 0$ to the one sided alternative hypothesis that older adults use machines A and B more than younger adults, or $\psi > 0$. For machine C, we also assigned a standard normal prior to ψ, but this time we compared the null hypothesis $\psi = 0$ to the one sided alternative hypothesis $\psi < 0$, or that older adults use machine C less than younger adults.

We also looked post-hoc at the correlation between the number of switches and the total trials needed to finish the experiment, as well as the correlation between the number of switches and average reward earned for all the participants. For these post-hoc tests, we assigned an uninformed stretched beta prior with a width of 1 to Kendall's τ_B (van Doorn, Ly, Marsman, & Wagenmakers, 2018). These statistical analyses were done using JASP (JASP Team, 2023). Just like we did for experiment 1A, we will report Bayes factors (BFs) for our analyses in the results section.

1B: Results

After being allowed to use any of three different IntelliBakers freely to earn 150 stars, both young adults and older adults switched between using the various machines multiple times. Older adults switched 9.87 times on average (std 7.54 times) while younger adults switched machines 9.17 times on average (std 8.23 times). The Mann-Whitney U test supported the null hypothesis that there is no difference between age groups in terms of the number of switches ($BF_{0-} = 5.49$; Figure 3.5).

Older adults also had similar preferences for each machine on average compared to younger adults (Figure 3.6). Older adults used machine A for 23.6% of their trials while younger adults used it for 26.1% of trials ($BF_{0+} = 18.12$). For machine B, older adults used it for 35.9% of trials while younger adults used it for 35.5% of trials ($BF_{0+} = 10.54$). Finally, for machine C, older adults used it for 44.7% of trials while younger adults used it for 44.2%

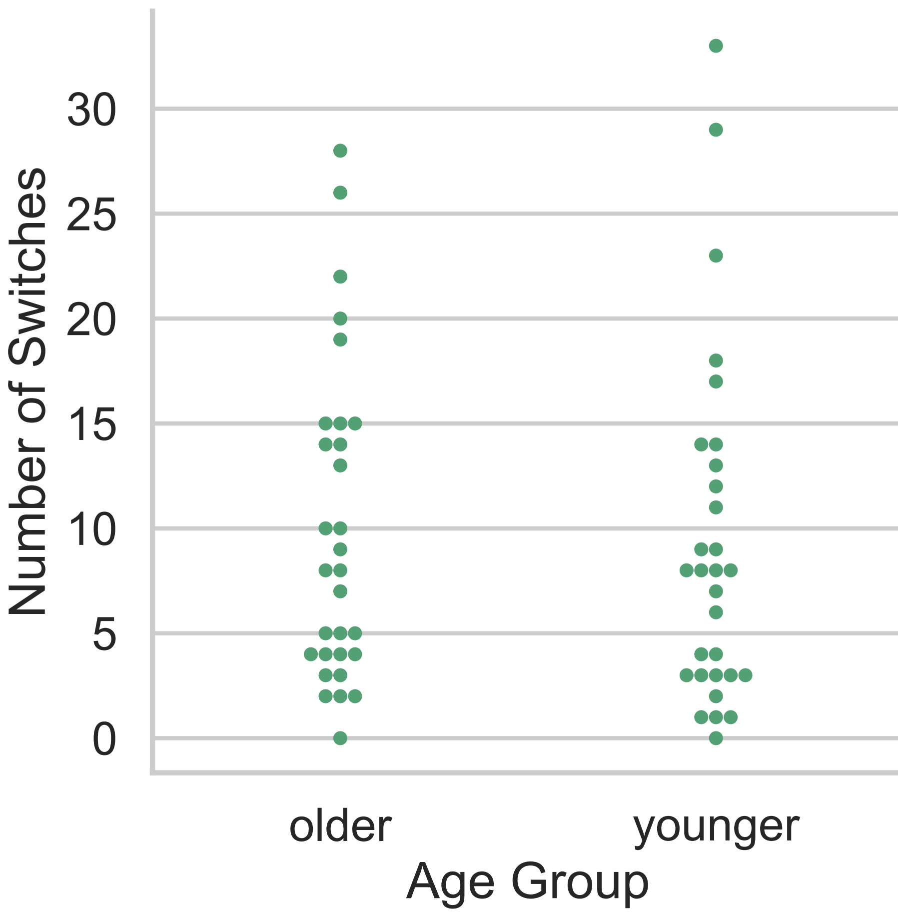

Figure 3.5: A plot of the number of times each participant switched between the three different versions of the IntelliBaker during the experiment. Older adults had similar switching behavior to younger adults.

of trials ($BF_{0-} = 19.52$). The A/B tests for each machine support the null hypothesis that there is no difference in preferences between the two age groups.

Because people switched more often than we expected, we decided to look post-hoc at correlations between the number of switches and how it affected performance on the task across all participants. There was a positive correlation between the number of switches and the total number of trials needed to finish the task ($BF_{10} > 10^6$, Kendall's $\tau_B = 0.501$). There was also a negative correlation between the number of switches and the average reward earned per trial ($BF_{10} > 10^5$, Kendall's $\tau_B = -0.491$). Together, these results indicate that the

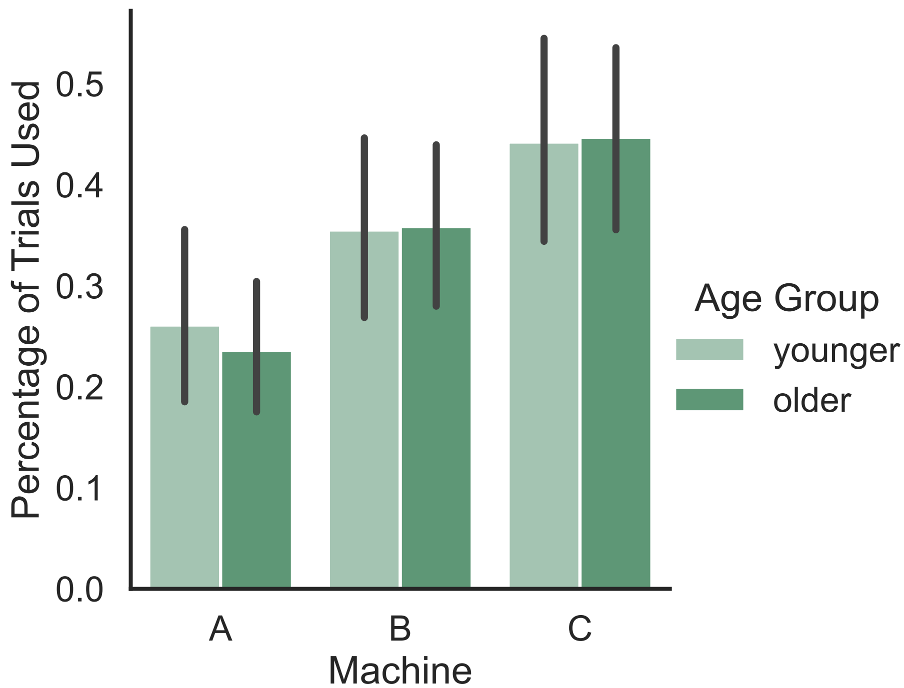

Figure 3.6: A bar graph plotting the percentage of trials that participants used a particular type of IntelliBaker, averaged across the participants in each age group. The error bars represent the 95% confidence interval of the mean. Older adults preferred certain machines as much as younger adults did.

more often that people switched, the less they learned about a particular machine, and their task performance suffered as a result.

1B: Discussion

Experiment 1B was designed with the idea in mind of people needing to learn a new tool to accomplish a goal. We wanted to see if given the choice between using a relatively easy tool, like machine A, and a harder to use yet better performing tool, like machines B and C, whether people would conserve their effort and stick to an easy machine or invest more effort towards learning a harder machine in anticipation of the higher rewards. Furthermore, we anticipated that there would be age differences in people's behavior and preferences when using these machines. Specifically, we hypothesized that older adults wouldn't switch between machines as much as younger adults, since older adults tend to explore less compared to younger adults (Spreng & Turner, 2021). We also hypothesized that older adults would conserve their effort and stick to using easier machines like machine A or B compared to younger adults, who might feel more confident in challenging machine C.

What we found instead was that older adults did not differ from younger adults very much in their switching behavior nor in their preference for a particular machine. For both age groups, the hardest machine, machine C, was the most used machine. This result supports the idea that older adults are willing to invest more mental effort into learning something if they believe the rewards are worth it (Devine et al., 2021).

We also found that some people switched between machines frequently and that this strongly correlated with poorer performance in the task in terms of average reward earned per trial and longer completion times. Perhaps the goal of 150 stars was too low and participants felt that even with increased exploration, they were still making timely progress towards the goal. With a larger goal, participants may realize that a better strategy would be to invest time into getting the maximum reward from one of the machines. Additionally, it's possible that the people who switched the most treated each machine as dispensing a random reward, rather than something which they had control over to produce better rewards. Future iterations of this task can involve different star goals to see how participants' behavior may be affected, as well as include some free response questions at the end of the experiment to probe participants' understanding of the IntelliBaker and the experiment goals.

3.4 General Discussion

We were interested in understanding how age affects people's decisions to invest effort into learning to use new technology. Previous work suggests that older adults will avoid cognitively effortful tasks in order to conserve their cognitive resources, which are thought to be more limited than those of younger adults (Devine et al., 2021; Ruel et al., 2021). With this in mind, we created the IntelliBaker task, a task with varying levels of difficulty that took time to learn. With this new task, we could manipulate the balance of effort and reward and observe how this balance affected people's use of new technologies, represented by the IntelliBakers, while working towards a goal, whether it was finding the best settings or earning a large number of stars.

While we expected to find that older adults would struggle with the most challenging IntelliBaker, machine C, and feel less inclined to use it compared to younger adults, we did not find strong evidence of age differences in how people chose to allocate their mental effort. Both younger and older adults learned how to use machine C and even when they were given the choice to use other machines, both age groups preferred to use machine C.

The simplest possible explanation for this finding is that machine C was not challenging enough for either age group and so older adults did not feel as if it required significant mental effort for them to be able to use it. Although we went through many design iterations of machine C to make it harder to learn, it is difficult to create a task that is both immensely challenging yet can still be learned within a reasonable amount of time. At least, this consideration stopped us from simply putting 10 knobs on machine C, which even younger participants would balk at. In any case, though machine C may not have seemed intimidating to participants, it certainly provided enough of a challenge since fewer older adults were able to find the best setting for machine C in experiment 1A compared to younger adults and only a handful of participants overall were able to earn more reward with machine C than they did with machines A or B in experiment 1B. Thus, it would also not be accurate to say that participants preferred machine C because it was easy for them to use.

Another possible explanation is that older adults were aware of the increased effort required to learn machine C over the other machines, but that they perceived the increased rewards to be worth the effort. Even if they didn't feel confident that they could master machine C, they may have engaged in a satisficing strategy in experiment 1B, where they tried to earn just enough reward per trial in order to make timely progress towards the goal. Although the current chapter does not include any analysis of the different strategies that participants used for the IntelliBaker task, the data is ripe for such analysis and could help us understand how people choose to learn about the IntelliBakers.

Despite the limitations with the current experiments, the data from both experiments clearly show that older adults are as capable as younger adults at learning this complex task. There were no group differences in terms of how long it took participants to find the best setting for each machine nor in terms of how much reward they earned per trial. Though it's unclear here to what extent the older adults were engaging in resource rationality, more generally, we should give older adults more credit for their potential to learn challenging new technologies.

Chapter 4

Older Adults Conserve Effort Regardless of Time Left to Learn

4.1 Introduction

A common stereotype of older adults is that they avoid using new technology such as the latest smartphone or teleconferencing software. However, older adults who resist new technology can have various reasons for doing so. They may worry that the new technology is too complicated for them to use easily or believe that it is unnecessary to learn, especially when there are existing solutions. These are reasonable concerns for anyone to have, so why are they attributed more to older adults? The answer may lie in the perception of the time costs associated with learning to use the new technology.

People seem to consider time costs differently based on their age. A person's perception of how much time they have to achieve their goals is sometimes referred to as their future time perspective. According to socioemotional selectivity theory, someone with a greater future time perspective will take actions to optimize for the future, whereas someone who perceives their time as limited will focus on optimizing for the present (Carstensen, 2006; Lang & Carstensen, 2002; Löckenhoff, 2011). Thus, younger adults can be thought of as having a greater future time perspective than older adults in general. However, future time perspective can be manipulated in experiments - when younger adults are asked to imagine that they will be moving to a new place soon, their responses match those of older adults' asked to imagine the same scenario (Fung & Carstensen, 2004). Similarly, asking older adults to imagine a medical breakthrough that can lengthen their lifespan also reduces any significant response differences with younger adults (Löckenhoff, 2011; Löckenhoff & Rutt, 2015).

Future time perspective may also play a role in older adults expending less effort compared to younger adults on decision making tasks (Mata et al., 2007; Worthy et al., 2014; Hinault & Lemaire, 2020). Older adults may judge that they do not have enough time to benefit from learning a more effortful strategy to make the best decisions, and thus end up using easier strategies. This thought process may also explain why older adults tend to rely heavily on prior knowledge instead of trying to learn new information (Spreng & Turner, 2021; Sherratt & Morand-Ferron, 2018). However, if future time perspective is subjective and can be localized to the time frame of an experiment, then perhaps older adults will choose to expend mental effort when they believe that they have enough time left in the experiment to benefit from the extra effort.

Although a recent meta-analysis (Seaman, Abiodun, Fenn, Samanez-Larkin, & Mata, 2022) found that there was little evidence for a relationship between age and the perception of future rewards, the authors only included studies using monetary rewards in intertemporal choice tasks. Such tasks generally involve asking participants whether they prefer a smaller immediate reward or a larger delayed reward. These tasks, and most tasks studying future time perspective, also rely on asking participants to imagine scenarios. Few studies manipulate participants' future time perspective by varying the length of the experiment to affect how much time participants spend in the experiment. Thus, we designed the present study such that participants' choices directly affect how quickly they can finish the experiment.

The research presented in this chapter makes use of the IntelliBaker task introduced in chapter 3. There are two different versions of the IntelliBaker, type A and type B, with type B being the more difficult of the two to learn. However, type B also gives more rewards than type A when mastered. Participants who can learn to skillfully use either type A or type B can finish the experiment faster than those who cannot. Furthermore, while we have participants use the easier type A for the majority of the experiment, they are given multiple opportunities to switch to type B if they believe that using type B will improve the rate at which they progress through the experiment. By varying the time points in the experiment where we ask participants to make a choice between using type A or type B, we are asking participants to estimate how much time they have left in the experiment and reason about which IntelliBaker type will help them reach the end faster. If someone has a long future time perspective, then they will be inclined to choose the more rewarding type B because they believe they have enough time to master it. On the other hand, if someone doesn't believe that they can master type B in time, they will choose to stick to type A. Considering this, the first hypothesis we plan to test is that older adults will stop switching earlier in the experiment compared to younger adults due to their shorter future time perspectives.

Additionally, we designed the present study such that we can determine whether people are making optimal decisions in their choice of IntelliBaker. If older adults are choosing to stay with type A because they don't have enough time to master type B, we can verify this with simulated data based on participants' task performance. Thus, the second hypothesis we will test is that older adults are acting optimally when they choose to stay with IntelliBaker type A instead of switching.

If we find that older adults perform better on the task when they forgo learning about a new option, then we can conclude that their reluctance to try new technologies is justified.

However, if we find that older adults perform suboptimally when avoiding the new technology, then we will know that their conservation of effort is actually holding them back.

4.2 Methods

Participants

Participants were recruited from the online platform Prolific. A person could take part in the experiment if they lived in the United States of America, were fluent in English, and had an approval rate of at least 90% on the platform. If the person was between the ages of 18 and 40 years old, they were assigned to the younger adult group and if the person was 60 years or older, then they were assigned to the older adult group. Participants were paid $3.00 for completing the study.

60 participants took part in the experiment. 1 participant was excluded for not completing the final block of the experiment for unknown reasons, leaving 59 participants. There were 29 participants in the younger adult group and 30 participants in the older adult group. The age range of the younger adult group was 22-39 years (mean 29.79 years) while the range of the older adult group was 60-77 years (mean 66.30 years).

Procedure

The experiment consisted of 5 blocks total. During each block, participants were tasked with using an IntelliBaker machine to earn 150 stars total (Figure 4.1). The details of the IntelliBaker task are in chapter 3, but in short, participants are asked to find the best settings for the machine and on each attempt, participants can earn between 1 and 7 stars using IntelliBaker type A and between 1 and 9 stars using IntelliBaker type B. However, it

Figure 4.1: An example of the main task screen which features IntelliBaker type A.

takes more attempts on average to earn 9 stars with type B than to earn 7 stars with type A (see experiment 1A in chapter 3). Participants learn to find the best setting based on the number of stars they earn on a trial. The more stars they earn, the closer their guess is to the correct setting.

The first block, baseline block A, consisted of participants only using IntelliBaker type A to earn 150 stars. The second block, baseline block B, consisted of participants only using IntelliBaker type B. These baseline blocks allowed participants to get a sense of how both types work. Then came three test blocks, which were randomized. In each test block, participants always started with a new version of type A, but once they earned a certain amount of stars, they were given the option to either switch to type B or stay with type A (Figure 4.2. This option was offered after either earning 15, 75, or 125 stars to correspond roughly with the beginning, middle, and end of the block. After participants made their choice, they were asked to type why they made that choice before finishing the rest of the block with their preferred machine.

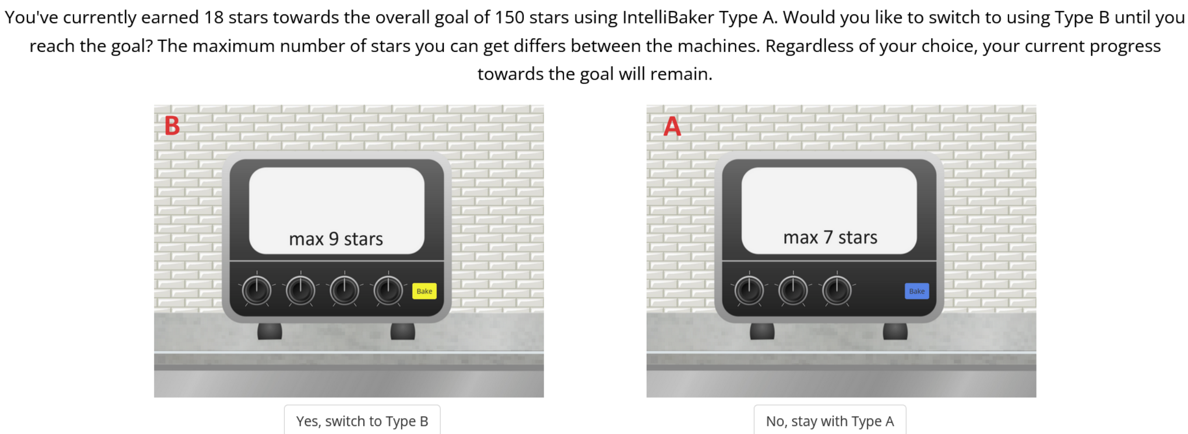

Figure 4.2: An example of the switch screen which appears during the test blocks after participants have earned either 15, 75, or 125 stars out of 150 stars.

Analysis

First, we'll confirm the results from experiment 1A in chapter 3 that there is no difference between age groups in how well they can use IntelliBaker types A and B. We'll use a Bayesian Mann-Whitney U test (van Doorn et al., 2020) to compare the two age groups on the number of trials needed to reach 150 stars in the baseline blocks. We will also confirm the absence of learning between blocks by using a Bayesian paired samples Wilcoxon test on participants' performance during their first test block and the last block (van Doorn et al., 2020).

We calculated the proportion of adults who chose to switch in each age group at the three different switch points (15, 75, and 125 stars). Then we compared the proportions at each time point for both age groups (e.g. % of older adults who switched after 15 stars vs % of older adults who switched after 75 stars) using a Bayesian A/B test (Hoffmann et al., 2022). We also used the Bayesian A/B test to compare the two age groups at the same time point.

In order to determine whether older adults were making optimal choices, we needed to approximate the counterfactual scenario. Then, we could compare the number of trials, or attempts, each participant needed to reach 150 stars during the test block to the counterfactual block. If a participant used more attempts during their test block than the imagined counterfactual, then their choice of IntelliBaker seems suboptimal. In order to simulate the data of what these counterfactual blocks could look like, we used the data from the first two baseline blocks. If a participant chose to switch to type B, we could simply compare that test block performance to the performance in baseline block A. If a participant chose to stay with type A, then we spliced together the rewards earned per trial from the test block up until the switch point and their rewards earned per trial from baseline block B, to simulate the scenario where the participant switched and started to learn type B from the beginning.

We used this counterfactual analysis method to create visualizations of group level performance after switching at 15, 75, and 125 stars. These visualizations made use of only the participants' baseline block data, such that rewards from the baseline A block were spliced together with rewards from participants' baseline B block to visualize the hypothetical switching scenario. We averaged the number of trials participants needed to reach or cross a certain reward threshold and then plotted this average at various reward levels (specifically at 5, 15, 25, 50, 75, 100, 125, and 150 stars). The error bands were constructed using the 95% confidence interval of the number of trials at each point.

Once we constructed the counterfactual data for each participant, we find the difference between the number of trials needed to reach 150 stars in the test block and the counterfactual block. If this difference is positive, that means the participant's choice resulted in them needing extra trials to finish the block. If the difference is negative, then that means their choice resulted in them needing less trials compared to the counterfactual. This difference is also known as regret in the machine learning literature (Neller & Lanctot, 2013). We then use a Bayesian t-test (Rouder, Speckman, Sun, Morey, & Iverson, 2009) to compare the regret between older adults who chose to switch and older adults who chose to stay at each time point to determine whether or not older adults are acting optimally. Since we hypothesize that older adults who choose to stay with IntelliBaker type A are acting optimally, we will define the alternative hypothesis as older adults who choose to stay will use fewer trials than older adults who had chosen to switch (i.e. their regret will be negative.) For younger adults, we used a two tailed alternative hypothesis since we are unsure which is the optimal choice for younger adults.

All of the Bayesian analysese were conducted using JASP (JASP Team, 2023). In the results section, we will report Bayes factors (BFs) for our analyses. Following the notation for the alternative hypothesis (1) against the null (0), $BF_{10} > 1$ indicates evidence for the alternative hypothesis while $BF_{10} < 1$ indicates evidence for the null hypothesis. If the alternative hypothesis is one sided, then the notation changes to BF_{+0} or BF_{-0}, depending on if the alternative hypothesis is that effect is greater or less for one group over the other. The value of the Bayes factor increases with the likelihood of the alternative hypothesis. For example, $BF_{10} = 10$ means that the data are 10 times more likely under the alternative hypothesis compared to the null hypothesis. Generally, BFs between 3 and 10 indicate moderate evidence against the null hypothesis, and BFs greater than 10 indicate strong evidence against the null (Kass & Raftery, 1995).

Additionally, we examined the free-text responses participants gave immediately after making their decision to switch or stay and coded these responses as one of five reasons for choosing a particular machine: ease of use, reward potential, time efficient, mastery focused, and other reasons not captured by the aforementioned categories, such as choosing based on a whim. Each participant was asked for a reason three times throughout the experiment and as such there are 177 total responses. We report the proportion of responses in each category in the results section.

4.3 Results

First, we compared the two age groups on the number of trials participants needed to reach 150 trials in the baseline blocks. There was little to no evidence pointing to a difference between age groups for both IntelliBaker type A ($BF_{10} < 1$) and type B ($BF_{10} < 3$). We also verified that both older adults ($BF_{-0} < 1$) and younger adults ($BF_{-0} < 1$) were not improving between the first test block and the last test block. Thus, we can conclude that both age groups are comparable when looking at their IntelliBaker performance.

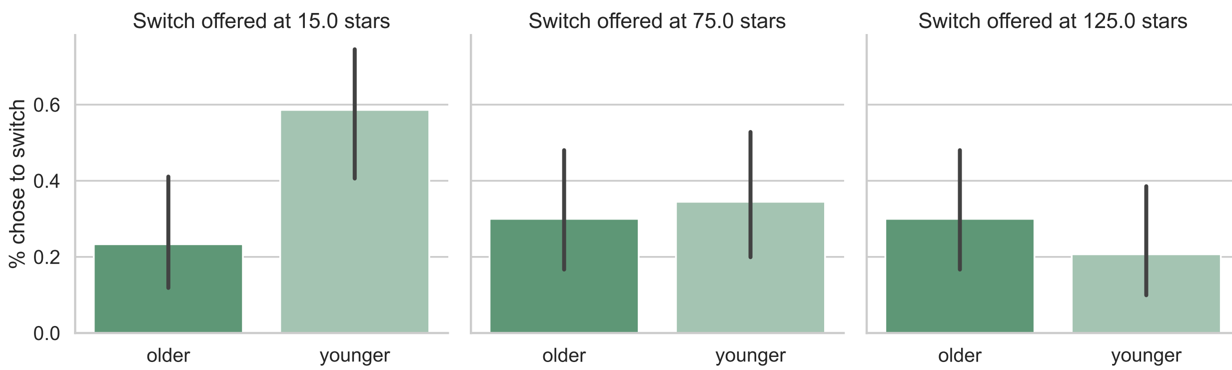

Figure 4.3: The proportion of each age group that chose to switch to IntelliBaker type B after either earning 15 stars, 75 stars, or 125 stars towards the overall goal of earning 150 stars. Error bars reflect the 95% credible interval.

Next, we compared the proportion of each age group who chose to switch to type B across the three different time points: at 15 stars, 75 stars, and 125 stars (Figure 4.3. There was no difference between the 7 out of 30 older adults who switched at 15 stars and the 9 older adults who switched at both 75 stars and 125 stars ($BF_{10} < 1$). There was moderate evidence for a difference between the 17 younger adults who chose to switch at 15 stars compared to the 10 younger adults who switched at 75 stars ($BF_{10} > 3$). There was very little evidence for a difference between these 10 younger adults at 75 stars and the 6 younger adults who switched after 125 stars ($BF_{10} > 1$). We also found a difference between age groups at 15 stars ($BF_{10} > 20$) but not at the other time points. The proportion of older adults who

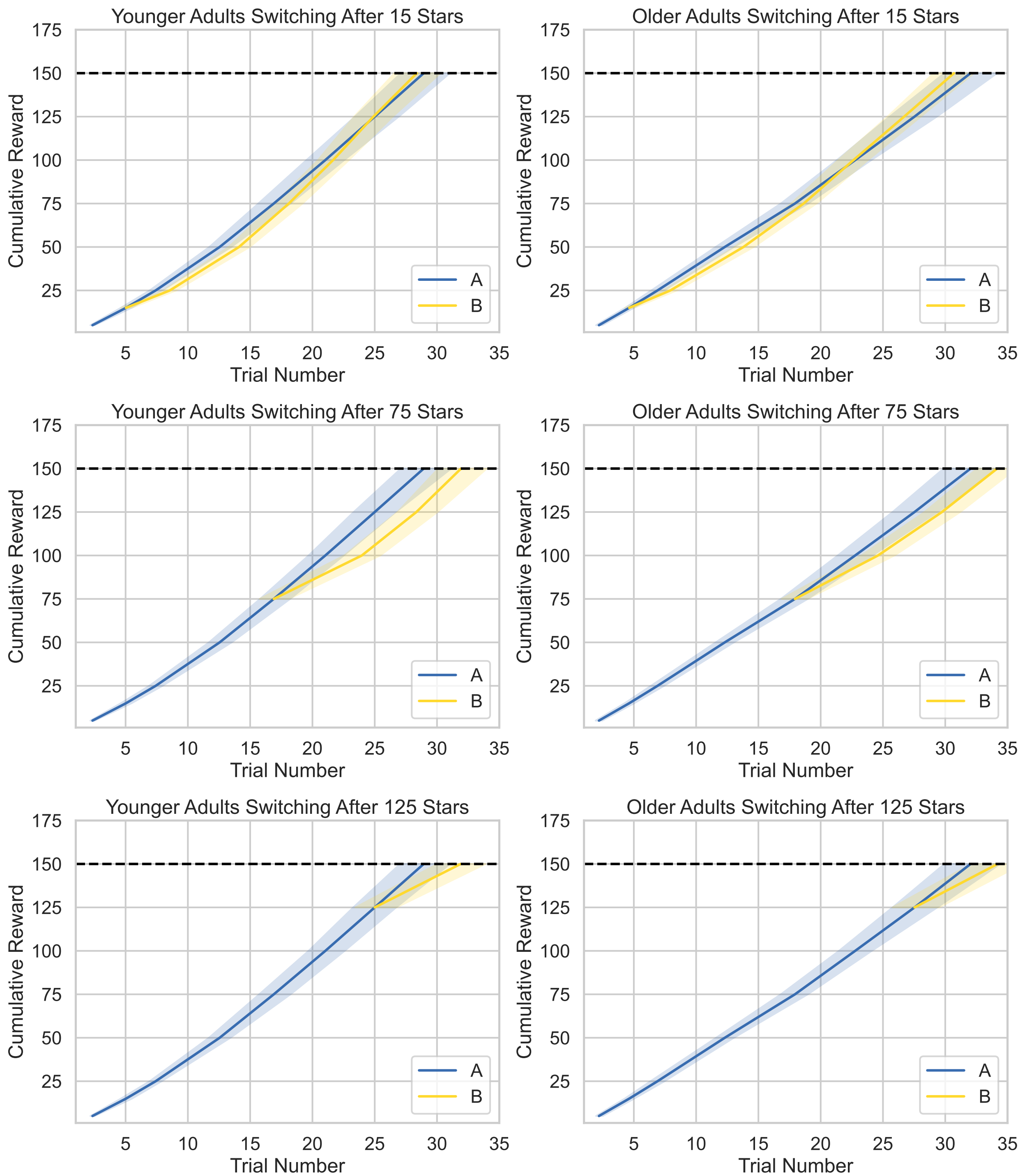

Figure 4.4: Plots of the predicted cumulative rewards for each age group if participants chose to switch to using IntelliBaker type B after earning 15, 75, and 125 stars out of 150 total, compared to if they stayed with IntelliBaker type A for the rest of the block. The earlier in the block participants switch, the better chance they have of finishing the block faster than if they chose to stay with the same IntelliBaker.

switched stayed relatively consistent at each time point, whereas the proportion of younger adults seems to decrease as the switch is offered later in the block.

Figure 4.4 illustrates the counterfactual performance for younger adults and older adults as a group at each of the possible switch points. Each group's baseline block A performance is plotted alongside for visual comparison of the reward trajectories. Based on these results, it seems like both younger and older adults have a chance at finishing the test block faster if they choose to switch to IntelliBaker type B early on in the block, after earning 15 stars (Figure 4.4 top panel). As the switch occurs later and later in the block it appears more likely to take younger and older adults more time to finish the block compared to if they just continued using IntelliBaker type A (Figure 4.4 middle and bottom panels).

When we compared participants' actual test block performance to their simulated counterfactual performance, we found no strong evidence that older adults who stayed with IntelliBaker type A were acting suboptimally. Specifically, when we compared the number of extra trials needed by older adults who chose to stay with type A compared to the older adults who chose to switch to type B, there was no difference at 15 stars ($BF_{-0} < 1$) nor at 75 stars ($BF_{-0} < 3$). While the older adults who stayed with type A weren't acting suboptimally, we can't conclude that they were acting optimally either, at least at the earlier switch points. At 125 stars, staying with type A was clearly the better option for older adults ($BF_{-0} > 20$; Figure 4.5). Turning our attention to the younger adults, there was no difference between those who chose to stay and those who chose to switch at 15 stars ($BF_{10} < 1$). There was strong evidence for differences in younger adults at 75 stars ($BF_{10} > 20$) and 125 stars ($BF_{10} > 200$). Figure 4.5 reveals that staying with type A is the optimal choice for younger adults at these later time points.

We also coded the free-form text responses to the prompt "please tell us why you chose this machine" into five categories: ease of use, reward potential, time efficient, mastery focused, and other. Reward potential was the code given to any responses which mentioned

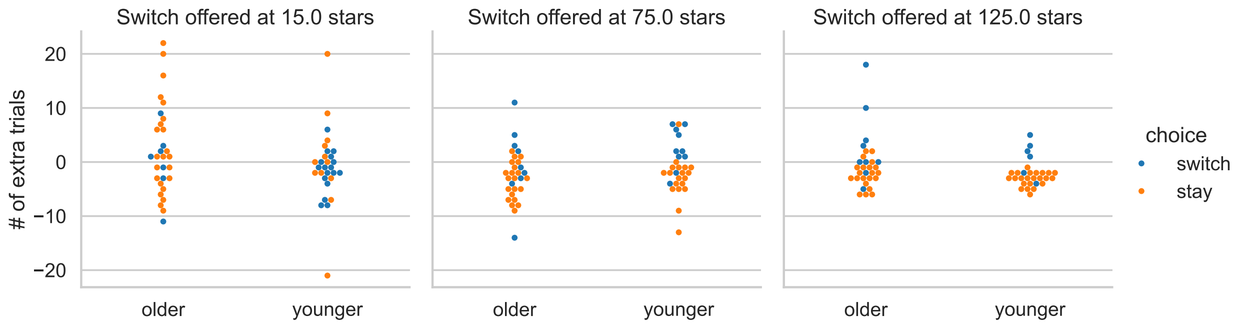

Figure 4.5: The number of extra trials a participant needed to reach the end of the block based on their decision to switch or stay, compared to their counterfactual data. Values less than or near 0 can be interpreted as having made an optimal choice since these participants finished the block as fast or faster than if they had made the other choice.

being able to get a higher amount of stars while ”time efficient” was the category for all responses mentioning an aspect of time, such as being able to finish the experiment faster. ”Mastery focused” describes those responses which mention finding the best combination for a particular IntelliBaker type, since achieving this would indicate mastery of the IntelliBaker. Examples of these responses and of the other categories can be found in Table 4.1, as well as the proportion of responses which fall into each category. The reasons most often given for staying with type A fell into the category “ease of use”, which made up 36.7% of all responses. The reasons most often given for switching to type B were in the “time efficient” category and made up 11% of all responses.

Category	**Example Response**	**Stay**	**Switch**
Ease of Use	“It is easier”/“has less knobs”	0.367	0.017
Reward Potential	“More possible points.”	0.017	0.085
Time Efficient	“to reach the goal with fewer attempts”	0.062	0.113
Mastery Focused	“I know the settings on this machine”	0.175	0.028
Other	“Just a random guess...”	0.107	0.085

Table 4.1: The proportion of responses given for why a participant chose to stay with IntelliBaker type A or switch to type B, broken down by choice. Each participant made three choices, so there were 177 total responses.

4.4 Discussion

We conducted the current experiment to understand how future time perspective affects older adults' decisions to invest mental effort towards learning a new and challenging technological device. The main task was to use an IntelliBaker to earn rewards totalling 150 stars. The speed at which participants moved through the task was directly tied to their ability to learn the IntelliBaker. At three different points in the task, we offered participants the option to switch to a more difficult yet more rewarding IntelliBaker. We expected that participants would choose to switch if they believed they had enough time left in the task to benefit from the switch, otherwise they would stay with the easier, more familiar IntelliBaker.

Because older adults have shorter future time perspectives on average compared to younger adults (Carstensen, 2006, 2021), we expected that the proportion of older adults who choose to switch will decrease earlier in the experiment compared to younger adults. We found that the proportion of older adults who switched stayed relatively constant throughout the task. We also found that the proportion of younger adults who switched was highest near the beginning of the task, before dropping during the middle of the task and staying relatively constant till the end. We can interpret this finding in one of two ways. The first is that older adults are not affected by the amount of time they have to learn something and will always prefer to conserve their efforts and stick to the easier option. This interpretation is in line with previous results in the literature where older adults preferred known options (Spreng & Turner, 2021; Sherratt & Morand-Ferron, 2018). The second is that older adults never felt that they had enough time to learn the new IntelliBaker at the time points we chose to test and thus it remains possible that a higher proportion of older adults are willing to switch at an even earlier point in the task, much like we see in the younger adults. The switch point of 15 stars may have been the stopping point we were hoping to observe where older adults largely choose to stay with the easier option compared to younger adults who were more likely to switch at this point. Because there were age group differences at this point, we cannot rule out the possibility that our hypothesis holds at an earlier time point. However, participants earn 15 stars pretty quickly in the task, after using the IntelliBaker only about 4 or 5 times. Offering participants a switch even earlier may be risky since participants may choose to stay with the easier IntelliBaker because they may be curious about how well it works and have not had adequate time to learn about it. Thus, it would be better to extend the goal beyond 150 stars instead and lengthen the task. Perhaps a goal that is farther away will assure participants that they have enough time to learn a new IntelliBaker. With this option though, care must be taken to avoid enticing people to switch due to boredom. With the current design, at least, very few responses noted boredom as the reason for switching (boredom made up a fraction of the "Other" category in 4.1).

We also wondered whether older adults who choose to stay with the easier IntelliBaker are doing so optimally. The results are mixed. Towards the end of the task, older adults who stay with the easier IntelliBaker definitely made the correct choice in terms of finishing the task faster, but at earlier time points, there is little difference between the people who chose to switch and people who chose to stay. This was also true for younger adults at the earliest switch point of 15 stars, but later in the task, it was clear that staying with the easier machine was the best option for most people. Figure 4.5 indicates that early in the task, whether a particular choice is optimal is dependent on the individual's performance regardless of their age. Although these results are based on crude approximations of the counterfactual, we believe that our method is a reasonable starting point. We found very little evidence for learning between blocks so we can equate performance from the earlier blocks and later blocks. Even if we suppose that participants get better at using the IntelliBaker each time they use one, we can then assume that the baseline blocks represent participants' worst possible performance. In this scenario, participants who took more time to complete the test block clearly did not make the best choice because they were unable to best their worst performance despite getting better at the task.

Overall, these results neither prove nor disprove our hypotheses and thus, further research is needed. Future experiments can modify several parameters of the IntelliBaker task or similar "skilled bandit" tasks (e.g. Hotaling, Navarro, & Newell, 2021) to see how they affect participants' switch inclinations, such as when to offer a switch, the number of switch points, and the length of a block. The counterfactual analysis can also be improved by modeling participants' performance curves on these skilled bandit tasks and using the curves to predict performance after switching. Until these new experiments and analyses are conducted however, we interpret the current results to mean that older adults are inclined to stick with easier, more familiar options rather than learn a new one regardless of the amount of time given to them to learn. Whether this behavior is to their benefit or their detriment remains to be seen. If it is to their benefit, then the stereotype of older adults avoiding new technologies takes on a positive light. However, if older adults are missing opportunities to increase their quality of life due to feeling they have limited time to learn useful technologies, then this phenomenon should be established quickly so that solutions can be found.

CONCLUSION

The research included in this dissertation explores the various ways that adults can adapt to cognitive changes brought by aging, and also demonstrates that older adults are capable of learning complex strategies and tools as well as younger adults. In chapters 1 and 2, I discussed how some cognitive tutoring or extra practice on a task can help older adults perform as well as younger adults on planning, reasoning, and problem solving tasks. Like many other cognitive training programs that have come before this research, my results likely only hold for the particular set of tasks that I tested participants on. Strong evidence of transferable cognitive skills from a training task to activities of daily living remains elusive (Owen et al., 2010; Rebok et al., 2014; Simons et al., 2016). Thus, while training metacognitive skills such as teaching people how to choose a good planning strategy rather than training people to solve particular problems sounds promising in helping people make better decisions overall, much more careful research is needed to demonstrate that this kind of training will transfer beyond the training task. I believe such research will require sustained training of a particular metacognitive strategy on a variety of tasks in order for people to learn the particular features of a task in which the strategy is useful. Additionally, older adults will need much more training compared to younger adults in order to use these strategies proficiently. Despite these limitations, the work in these chapters adds to previous scientific work which finds that older adults can reach age equivalence with younger adults by adapting the way they learn, such as trying a different decision making strategy or giving themselves more time to learn.

In chapters 3 and 4, I investigated the reasons behind older adults' resistance to learning new technology and found that older adults are capable of learning complex tools and will approach them at the same rate as younger adults if the benefits of doing so are sufficiently high. Contrary to previous studies which reported that older adults explore less than younger adults (Queen et al., 2012; Mata et al., 2013; Wiegand et al., 2019; Spreng & Turner, 2021), we found in chapter 3 that older adults switch between IntelliBakers as often as younger adults do and they don't shy away from the hardest IntelliBaker either. However, in chapter 4, we found a different pattern of results – older adults were much less willing to switch to a different and more challenging IntelliBaker type compared to younger adults, which indicates a reluctance to explore and is in line with previous studies. The two experiments were set up similarly in terms of the goal and the types of IntelliBaker that were available, though chapter 4's experiment reduced the number of IntelliBaker types and did not include the most preferred IntelliBaker type from chapter 3 experiment 1B, type C. This was because the pace of earning rewards with type C is so slow that it would never make sense to choose type C over type A at any point in the experimental blocks in chapter 4. With type B, there was at least a chance of performing better after switching as opposed to staying with type A. However, type B had an earning potential of 9 stars compared to type C's 11 stars and so older adults may have felt that the reward potential was not high enough to invest the effort of learning the more difficult type B and chose to stay with type A. The other, major, difference between the two experiments is that the experiment in chapter 3 allowed participants to switch between IntelliBaker types at will, multiple times, while during the experiment in chapter 4, participants could only switch once per block and they had no control over when the switch option would appear. Having to commit to an IntelliBaker may have raised the stakes considerably and caused older adults to consider more deeply which IntelliBaker they wanted to spend time learning. In the situation where participants could switch at any time, older adults may have felt that there were no consequences to their decision because they could always switch to a different IntelliBaker if they felt they had chosen suboptimally. Not fearing severe consequences may have emboldened older adults to interact more with the challenging IntelliBaker type C.

Future experiments can elucidate this pattern of results from chapters 3 and 4 a number of ways. For example, decision consequences can be made clearer in the switch-at-will experimental setup by limiting the number of switches possible, while keeping the timing of the switch under participants' control. I also think both experimental setups would benefit from being made longer. The goal for both experiments was to earn 150 stars total using the participant's IntelliBaker(s) of choice but it seems likely that 150 stars is too few to feel the consequences of choosing a harder IntelliBaker that slows you down. Put another way, even if someone picked the wrong IntelliBaker to use, they still often reached the goal in a reasonable amount of time. Increasing the goal to 200 or 250 stars or more might encourage participants to really consider their time and effort costs while making a decision, and will provide further clarity as to whether older adults are swayed by time costs when choosing to learn something new. While the results from chapters 3 and 4 may have left us with more questions than answers, I believe they still provide encouraging evidence that older adults are capable not only of learning complex tasks, but also of acting optimally at times by conserving their mental effort.

This dissertation also introduces or builds upon a few new tasks and methods which can greatly aid future research on cognitive aging's effects on learning and decision making. In chapter 2, I introduced a new type of analysis which harnessed the power of a large data set to quantify training benefits at an individual level rather than at the group level, allowing us to be much more precise in our measurements. I also made the data set of nearly 10,000 participants' scores on 57 cognitive training games publicly available for other researchers (`https://osf.io/wbrq2/`). As I noted in chapter 3, I created the IntelliBaker task due to a lack of complex learning tasks in the cognitive aging literature and demonstrated that both younger and older adults are able to learn this task. The ways in which I set up and used the IntelliBaker task are far from the only ways to use it. Not only the number of knobs and settings, but the maximum potential rewards, or even how rewards are earned per trial can all be varied and adjusted to suit researchers' questions of interest. I believe there is a lot of potential in using the task to study problem solving strategies in older adults and I have deposited the full data from chapters 3 and 4, which includes participants' guesses and the correct setting, on the Open Science Framework for interested researchers (`https://osf.io/k2ye8/`).

Although the work in this dissertation doesn't directly engage with it, there is an adjacent line of cognitive aging research which contends that differences between older adults and younger adults on cognitive tasks appear due to the wealth of experience and knowledge that older adults have built over their lives. For example, older adults struggle with vocabulary-based problem solving tasks compared to younger adults because they know many more words than younger adults and thus need more time to mentally consider all the possible options (Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014; Ramscar, Sun, Hendrix, & Baayen, 2017). Relatedly, the decision making strategies that older adults default to often work poorly on decision making tasks in the lab because many lab tasks treat each decision as independent, which are unlike the decisions people make in real life. When lab based decision making tasks are modified so that each decision depends on a previous outcome, older adults' strategies turn out to work quite well (Blanco et al., 2016; Worthy & Maddox, 2012). These results suggest that researchers should be very careful about attributing an age effect to cognitive decline without carefully considering how task characteristics might put older adults at an unfair disadvantage. To this end, researchers interested in the planning strategies of older adults can make great use of the Mouselab-MDP task used in chapter 1 because it is flexible enough to support many different reward environments. Based on the results from experiment 1 in that chapter, I suspect that older adults would do as well as younger adults on an environment which rewards depth-first planning.

This dissertation adds to the growing body of scientific literature which provides a more nuanced look at cognition in later life. While it's true that aging contributes to cognitive decline and limits the use of cognitive resources, older adults need not resign themselves to a life of cognitive struggles and failures. As research on cognitive aging progresses, I remain optimistic that we will continue to uncover the ways in which the brain adapts with age and devise solutions to support lifelong learners.

References

Allain, P., Nicoleau, S., Pinon, K., Etcharry-Bouyx, F., Barré, J., Berrut, G., … Gall, D. L. (2005, February). Executive functioning in normal aging: A study of action planning using the Zoo Map Test. *Brain and Cognition*, *57*(1), 4–7. doi: 10.1016/j.bandc.2004.08.011

Anguera, J. A., Boccanfuso, J., Rintoul, J. L., Al-Hashimi, O., Faraji, F., Janowich, J., … Gazzaley, A. (2013, September). Video game training enhances cognitive control in older adults. *Nature*, *501*(7465), 97–101. doi: 10.1038/nature12486

Baltes, P. B., Dittmann-Kohli, F., & Kliegl, R. (1986). Reserve Capacity of the Elderly in Aging-Sensitive Tests of Fluid Intelligence: Replication and Extension. *Psychology and Aging*, *1*(2), 172–177. doi: 10.1037/0882-7974.1.2.172.

Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology*, *28*(1), 121–125. doi: 10.1037/0012-1649.28.1.121

Blanco, N. J., Love, B. C., Ramscar, M., Otto, A. R., Smayda, K., & Maddox, W. T. (2016, January). Exploratory decision-making as a function of lifelong experience, not cognitive decline. *Journal of Experimental Psychology: General*, *145*(3), 284. (Publisher: US: American Psychological Association) doi: 10.1037/xge0000133

Bolenz, F., Kool, W., Reiter, A. M., & Eppinger, B. (2019, August). Metacontrol of decision-making strategies in human aging. *eLife*, *8*, e49154. (Publisher: eLife Sciences Publications, Ltd) doi: 10.7554/eLife.49154

Buchler, N. G., Hoyer, W. J., & Cerella, J. (2008, June). Rules and more rules: The effects of multiple tasks, extensive training, and aging on task-switching performance. *Memory & Cognition*, *36*(4), 735–748. doi: 10.3758/MC.36.4.735

Callaway, F., Jain, Y. R., van Opheusden, B., Das, P., Iwama, G., Gul, S., … Lieder, F. (2022, March). Leveraging artificial intelligence to improve people's planning strategies. *Proceedings of the National Academy of Sciences*, *119*(12), e2117432119. (Publisher: Proceedings of the National Academy of Sciences) doi: 10.1073/pnas.2117432119

Callaway, F., Lieder, F., Das, P., Gul, S., M Krueger, P., & L. Griffiths, T. (2018, February). A resource-rational analysis of human planning. In *Proceedings of the 40th Annual Meeting of the Cognitive Science Society.* doi: 10.13140/RG.2.2.15636.40326

Callaway, F., Lieder, F., Krueger, P., & Griffiths, T. (2017, July). Mouselab-MDP: A new paradigm for tracing how people plan.. doi: 10.31219/osf.io/7wcya

Callaway, F., van Opheusden, B., Gul, S., Das, P., Krueger, P. M., Griffiths, T. L., & Lieder, F. (2022, August). Rational use of cognitive resources in human planning. *Nature*

Human Behaviour, *6*(8), 1112–1125. (Number: 8 Publisher: Nature Publishing Group) doi: 10.1038/s41562-022-01332-8

Carstensen, L. L. (2006, June). The Influence of a Sense of Time on Human Development. *Science*, *312*(5782), 1913–1915. (Publisher: American Association for the Advancement of Science) doi: 10.1126/science.1127488

Carstensen, L. L. (2021, November). Socioemotional Selectivity Theory: The Role of Perceived Endings in Human Motivation. *The Gerontologist*, *61*(8), 1188–1196. doi: 10.1093/geront/gnab116

A Consensus on the Brain Training Industry from the Scientific Community. (2014, August). (http://longevity3.stanford.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community/)

Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology and Aging*, *23*(4), 720–730. doi: 10.1037/a0014296

Davidson, D. J., Zacks, R. T., & Williams, C. C. (2003, June). Stroop Interference, Practice, and Aging. *Aging, Neuropsychology, and Cognition*, *10*(2), 85–98. doi: 10.1076/anec.10.2.85.14463

Derksen, B. J., Duff, M. C., Weldon, K., Zhang, J., Zamba, K. D., Tranel, D., & Denburg, N. L. (2015, May). Older adults catch up to younger adults on a learning and memory task that involves collaborative social interaction. *Memory*, *23*(4), 612–624. doi: 10.1080/09658211.2014.915974

Devine, S., Neumann, C., Otto, A. R., Bolenz, F., Reiter, A., & Eppinger, B. (2021, November). Seizing the opportunity: Lifespan differences in the effects of the opportunity cost of time on cognitive control. *Cognition*, *216*, 104863. doi: 10.1016/j.cognition.2021.104863

Donner, Y., & Hardy, J. L. (2015, October). Piecewise power laws in individual learning curves. *Psychonomic Bulletin & Review*, *22*(5), 1308–1319. doi: 10.3758/s13423-015-0811-x

Frank, M. J., & Kong, L. (2008). Learning to avoid in older age. *Psychology and Aging*, *23*(2), 392–398. (Place: US Publisher: American Psychological Association) doi: 10.1037/0882-7974.23.2.392

Fung, H. H., & Carstensen, L. L. (2004). Motivational Changes in Response to Blocked Goals and Foreshortened Time: Testing Alternatives to Socioemotional Selectivity Theory. *Psychology and Aging*, *19*(1), 68–78. doi: 10.1037/0882-7974.19.1.68

Goldstone, R. L., & Lupyan, G. (2016). Discovering Psychological Principles by Mining Naturally Occurring Data Sets. *Topics in Cognitive Science*, *8*(3), 548–568. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12212) doi: 10.1111/tops.12212

Griffiths, T. L. (2015, February). Manifesto for a new (computational) cognitive revolution. *Cognition*, *135*, 21–23. doi: 10.1016/j.cognition.2014.11.026

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015, April). Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229. doi: 10.1111/tops.12142

Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*. Elsevier.

Hess, T. M., Smith, B. T., & Sharifian, N. (2016, November). Aging and effort expenditure: The impact of subjective perceptions of task demands. *Psychology and Aging*, *31*(7), 653–660. doi: 10.1037/pag0000127

Hinault, T., & Lemaire, P. (2020). Aging Effects on Brain and Cognition: What Do We Learn from a Strategy Perspective? In A. Gutchess & A. K. Thomas (Eds.), *The Cambridge Handbook of Cognitive Aging: A Life Course Perspective* (pp. 127–146). Cambridge: Cambridge University Press. doi: 10.1017/9781108552684.009

Hoffmann, T., Hofman, A., & Wagenmakers, E.-J. (2022, December). Bayesian Tests of Two Proportions: A Tutorial With R and JASP. *Methodology*, *18*(4), 239–277. doi: 10.5964/meth.9263

Hotaling, J. M., Navarro, D. J., & Newell, B. R. (2021, June). Skilled bandits: Learning to choose in a reactive world. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(6), 879–905. doi: 10.1037/xlm0000981

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012, March). Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLOS Computational Biology*, *8*(3), e1002410. (Publisher: Public Library of Science) doi: 10.1371/journal.pcbi.1002410

Kable, J. W., Caulfield, M. K., Falcone, M., McConnell, M., Bernardo, L., Parthasarathi, T., ... Lerman, C. (2017, August). No Effect of Commercial Cognitive Training on Brain Activity, Choice Behavior, or Cognitive Performance. *Journal of Neuroscience*, *37*(31), 7390–7402. doi: 10.1523/JNEUROSCI.2832-16.2017

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*(6), 978–990. doi: 10.1111/j.1467-7687.2009.00846.x

Kass, R. E., & Raftery, A. E. (1995, June). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. doi: 10.1080/01621459.1995.10476572

Kliegl, R., Smith, J., & Baltes, P. B. (1989). Testing-the-Limits and the Study of Adult Age Differences in Cognitive Plasticity of a Mnemonic Skill. *Developmental Psychology*, *25*(2), 247–256. doi: 10.1037/0012-1649.25.2.247

Kool, W., & Botvinick, M. (2018, December). Mental labour. *Nature Human Behaviour*, *2*(12), 899–908. (Number: 12 Publisher: Nature Publishing Group) doi: 10.1038/s41562-018-0401-9

Kramer, A. F., Hahn, S., & Gopher, D. (1999, April). Task coordination and aging: explorations of executive control processes in the task switching paradigm. *Acta Psychologica*, *101*(2), 339–378. doi: 10.1016/S0001-6918(99)00011-6

Kray, J., & Ferdinand, N. K. (2014, June). Task Switching and Aging. In J. Grange & G. Houghton (Eds.), *Task Switching and Cognitive Control* (p. 0). Oxford University Press. doi: 10.1093/acprof:osobl/9780199921959.003.0014

Kray, J., & Lindenberger, U. (2000). Adult age differences in task switching. *Psychology and Aging*, *15*(1), 126–147. doi: 10.1037/0882-7974.15.1.126

Kumar, A., Benjamin, A. S., Heathcote, A., & Steyvers, M. (2022, October). Comparing models of learning and relearning in large-scale cognitive training data sets. *npj Science of Learning*, *7*(1), 1–11. doi: 10.1038/s41539-022-00142-x

Lampit, A., Hallock, H., & Valenzuela, M. (2014, November). Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. *PLOS Medicine*, *11*(11), e1001756. doi: 10.1371/journal.pmed.1001756

Lang, F. R., & Carstensen, L. L. (2002, March). Time counts: Future time perspective, goals, and social relationships. *Psychology and Aging*, *17*(1), 125–139. (Num Pages: 125-139 Place: Washington, US Publisher: American Psychological Association (US)) doi: 10.1037/0882-7974.17.1.125

Lemaire, P. (2010, December). Cognitive Strategy Variations During Aging. *Current Directions in Psychological Science*, *19*(6), 363–369. (Publisher: SAGE Publications Inc) doi: 10.1177/0963721410390354

Lieder, F., Callaway, F., Krueger, P. M., Das, P., Griffiths, T. L., & Gul, S. (2018). Discovering and Teaching Optimal Planning Strategies. In (p. 2).

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. (Publisher: Cambridge University Press) doi: 10.1017/S0140525X1900061X

Lieder, F., Krueger, P. M., Callaway, F., & Griffiths, T. (2017). A reward shaping method for promoting metacognitive learning.. doi: 10.31234/osf.io/qj346

Löckenhoff, C. E. (2011). Age, time, and decision making: from processing speed to global time horizons. *Annals of the New York Academy of Sciences*, *1235*(1), 44–56. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.2011.06209.x) doi: 10.1111/j.1749-6632.2011.06209.x

Löckenhoff, C. E., & Rutt, J. L. (2015, January). Chapter 11 - Age Differences in Time Perception and Their Implications for Decision Making Across the Life Span. In T. M. Hess, J. Strough, & C. E. Löckenhoff (Eds.), *Aging and Decision Making* (pp. 213–233). San Diego: Academic Press. doi: 10.1016/B978-0-12-417148-0.00011-X

Mata, R., Josef, A. K., & Lemaire, P. (2015, January). Chapter 6 - Adaptive Decision Making and Aging. In T. M. Hess, J. Strough, & C. E. Löckenhoff (Eds.), *Aging and Decision Making* (pp. 105–126). San Diego: Academic Press. doi: 10.1016/B978-0-12-417148-0.00006-6

Mata, R., Schooler, L. J., & Rieskamp, J. (2007). The aging decision maker: Cognitive aging and the adaptive selection of decision strategies. *Psychology and Aging*, *22*(4), 796–810. doi: 10.1037/0882-7974.22.4.796

Mata, R., & von Helversen, B. (2015). Search and the Aging Mind: The Promise and Limits of the Cognitive Control Hypothesis of Age Differences in Search. *Topics in Cognitive Science*, *7*(3), 416–427. (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/tops.12139) doi: 10.1111/tops.12139

Mata, R., von Helversen, B., & Rieskamp, J. (2010, June). Learning to choose: Cognitive aging and strategy selection learning in decision making. *Psychology and Aging*, *25*(2), 299–309. (Num Pages: 299-309 Place: Washington, US Publisher: American Psychological Association (US)) doi: https://doi.org/10.1037/a0018923

Mata, R., Wilke, A., & Czienskowski, U. (2013). Foraging across the life span: is there a reduction in exploration with aging? *Frontiers in Neuroscience*, *7*.

Mell, T., Heekeren, H. R., Marschner, A., Wartenburger, I., Villringer, A., & Reischies, F. M. (2005, January). Effect of aging on stimulus-reward association learning. *Neuropsychologia*, *43*(4), 554–563. doi: 10.1016/j.neuropsychologia.2004.07.010

Murman, D. L. (2015, August). The Impact of Age on Cognition. *Seminars in Hearing*, *36*(03), 111–121. doi: 10.1055/s-0035-1555115

Myers, C., & Conner, M. (1992). Age differences in skill acquisition and transfer in an implicit learning paradigm. *Applied Cognitive Psychology*, *6*(5), 429–442. doi: 10.1002/acp.2350060507

Neller, T. W., & Lanctot, M. (2013). An Introduction to Counterfactual Regret Minimization. In *Proceedings of Model AI Assignments, The Fourth Symposium on Educational Advances in Artificial Intelligence* (Vol. 11).

Newell, A., & Simon, H. A. (1972). *Human Problem Solving.* Prentice-Hall.

Nguyen, L., Murphy, K., & Andrews, G. (2019, July). Immediate and long-term efficacy of executive functions cognitive training in older adults: A systematic review and meta-analysis. *Psychological Bulletin*, *145*(7), 698–733. (Num Pages: 698-733 Publisher: American Psychological Association (US)) doi: http://dx.doi.org/10.1037/bul0000196

Noack, H., Lövdén, M., Schmiedek, F., & Lindenberger, U. (2013). Age-related differences in temporal and spatial dimensions of episodic memory performance before and after hundred days of practice. *Psychology and Aging*, *28*(2), 467–480. (Place: US Publisher: American Psychological Association) doi: 10.1037/a0031489

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., ... Ballard, C. G. (2010, June). Putting brain training to the test. *Nature*, *465*(7299), 775–778. doi: 10.1038/nature09042

Queen, T. L., Hess, T. M., Ennis, G. E., Dowd, K., & Grühn, D. (2012). Information search and decision making: Effects of age and complexity on strategy use. *Psychology and Aging*, *27*, 817–824. (Place: US Publisher: American Psychological Association) doi: 10.1037/a0028744

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The Myth of Cognitive Decline: Non-Linear Dynamics of Lifelong Learning. *Topics in Cognitive Science*, *6*(1), 5–42. doi: 10.1111/tops.12078

Ramscar, M., Sun, C. C., Hendrix, P., & Baayen, H. (2017, August). The Mismeasurement of Mind: Life-Span Changes in Paired-Associate-Learning Scores Reflect the "Cost" of Learning, Not Cognitive Decline. *Psychological Science*, *28*(8), 1171–1179. doi: 10.1177/0956797617706393

Ratcliff, R., Thapar, A., & McKoon, G. (2006, June). Aging, Practice, and Perceptual Tasks: A Diffusion Model Analysis. *Psychology and aging*, *21*(2), 353–371. doi: 10.1037/0882-7974.21.2.353

Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H.-Y., King, J. W., ... Willis, S. L. (2014). Ten-Year Effects of the Advanced Cognitive Training for Independent and Vital Elderly Cognitive Training Trial on Cognition and Everyday Functioning in Older Adults. *Journal of the American Geriatrics Society*, *62*(1), 16–24. doi: 10.1111/jgs.12607

Reijnders, J., van Heugten, C., & van Boxtel, M. (2013, January). Cognitive interventions in healthy older adults and people with mild cognitive impairment: A systematic review. *Ageing Research Reviews*, *12*(1), 263–275. doi: 10.1016/j.arr.2012.07.003

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009, April). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. doi: 10.3758/PBR.16.2.225

Ruel, A., Devine, S., & Eppinger, B. (2021). Resource-rational approach to meta-control problems across the lifespan. *WIREs Cognitive Science*, *12*(5), e1556. (_eprint:

https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1556) doi: 10.1002/wcs.1556

Salthouse, T. A. (1996). The Processing-Speed Theory of Adult Age Differences in Cognition. *Psychological Review*, *103*(3), 26. doi: 10.1037/0033-295X.103.3.403

Salthouse, T. A. (2010, September). Selective review of cognitive aging. *Journal of the International Neuropsychological Society*, *16*(5), 754–760. doi: 10.1017/S1355617710000706

Sanders, C., & Schmitter-Edgecombe, M. (2012, August). Identifying the nature of impairment in planning ability with normal aging. *Journal of Clinical and Experimental Neuropsychology*, *34*(7), 724–737. (Publisher: Routledge _eprint: https://doi.org/10.1080/13803395.2012.670210) doi: 10.1080/13803395.2012.670210

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. (Publisher: Institute of Mathematical Statistics)

Seaman, K. L., Abiodun, S. J., Fenn, Z., Samanez-Larkin, G. R., & Mata, R. (2022). Temporal discounting across adulthood: A systematic review and meta-analysis. *Psychology and Aging*, *37*, 111–124. (Place: US Publisher: American Psychological Association) doi: 10.1037/pag0000634

Sherratt, T. N., & Morand-Ferron, J. (2018, April). The adaptive significance of age-dependent changes in the tendency of individuals to explore. *Animal Behaviour*, *138*, 59–67. doi: 10.1016/j.anbehav.2018.01.025

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, *63*, 129–138. (Place: US Publisher: American Psychological Association) doi: 10.1037/h0042769

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016, October). Do “Brain-Training” Programs Work? *Psychological Science in the Public Interest*, *17*(3), 103–186. doi: 10.1177/1529100616661983

Sorel, O., & Pennequin, V. (2008, March). Aging of the Planning process: The role of executive functioning. *Brain and Cognition*, *66*(2), 196–201. doi: 10.1016/j.bandc.2007.07.006

Spreng, R. N., & Turner, G. R. (2021, December). From exploration to exploitation: a shifting mental mode in late life development. *Trends in Cognitive Sciences*, *25*(12), 1058–1071. doi: 10.1016/j.tics.2021.09.001

Steyvers, M., & Benjamin, A. S. (2019, August). The joint contribution of participation and performance to learning functions: Exploring the effects of age in large-scale data sets. *Behavior Research Methods*. doi: 10.3758/s13428-018-1128-2

Steyvers, M., Hawkins, G. E., Karayanidis, F., & Brown, S. D. (2019, September). A large-scale analysis of task switching practice effects across the lifespan. *Proceedings of the National Academy of Sciences*, *116*(36), 17735–17740. doi: 10.1073/pnas.1906788116

Steyvers, M., & Schafer, R. J. (2020, November). Inferring latent learning factors in large-scale cognitive training data. *Nature Human Behaviour*, *4*(11), 1145–1155. doi: 10.1038/s41562-020-00935-3

Team, J. (2023). *JASP (Version 0.18.0)[Computer software]*. (https://jasp-stats.org)

Vallat, R. (2018, November). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. doi: 10.21105/joss.01026 van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E.-J. (2018, October). Bayesian Inference for Kendall's Rank Correlation Coefficient. *The American Statistician*, *72*(4), 303–308. doi: 10.1080/00031305.2016.1264998

van Doorn, J., Ly, A., Marsman, M., & Wagenmakers, E.-J. (2020, December). Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and Spearman's ρ. *Journal of Applied Statistics*, *47*(16), 2984–3006. doi: 10.1080/02664763.2019.1709053

Verhaeghen, P., Marcoen, A., & Goossens, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging*, *7*(2), 242–251. doi: 10.1037/0882-7974.7.2.242

Westbrook, A., Kester, D., & Braver, T. S. (2013, July). What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference. *PLOS ONE*, *8*(7), e68210. (Publisher: Public Library of Science) doi: 10.1371/journal.pone.0068210

Wiegand, I., Seidel, C., & Wolfe, J. (2019). Hybrid foraging search in younger and older age. *Psychology and Aging*, *34*, 805–820. (Place: US Publisher: American Psychological Association) doi: 10.1037/pag0000387

Worthy, D. A., Cooper, J. A., Byrne, K. A., Gorlick, M. A., & Maddox, W. T. (2014, December). State-based versus reward-based motivation in younger and older adults. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(4), 1208–1220. doi: 10.3758/s13415-014-0293-8

Worthy, D. A., & Maddox, W. T. (2012). Age-Based Differences in Strategy Use in Choice Tasks. *Frontiers in Neuroscience*, *5*. (Publisher: Frontiers) doi: 10.3389/fnins.2011.00145

Appendix A

Supplementary Information for Chapter 2: "Older Adults Catch Up"

The Games

A list of all Lumosity games included in the analysis and a brief description of each game is provided in Table A.1. For a more complete description of the properties of each game and which cognitive processes are engaged, please refer to Extended Data Figures 1 and 2 in Steyvers and Schafer (2020). The extended description of each game's features is also available in a dummy coded format in the OSF repository for this paper (`https://osf.io/wbrq2`).

Table A.1: Table of all Lumosity games included in the main analyses and the cognitive domain they were classified as.

Game Name	Cognitive Domain	Game Type
Train of Thought	Attention	Divided Attention
Trouble Brewing	Attention	Divided Attention
Assist Ants	Attention	Divided Attention
Birdwatching	Attention	Field of View
Eagle Eye	Attention	Field of View
Eagle Eye 2	Attention	Field of View
Highway Hazards	Attention	Information Processing
Splitting Seeds	Attention	Information Processing
Lost in Migration 2	Attention	Selective Attention
Star Search	Attention	Selective Attention
Penguin Pursuit	Attention	Spatial Orientation
Speed Pack	Attention	Visualization
Color Match 2	Flexibility	Response Inhibition
Brain Shift 2	Flexibility	Task Switching
Brain Shift Overdrive 2	Flexibility	Task Switching
Disillusion	Flexibility	Task Switching
Disillusion 2	Flexibility	Task Switching
Ebb and Flow	Flexibility	Task Switching
Word Bubbles	Language	Verbal Fluency
Word Bubbles 3	Language	Verbal Fluency
Word Bubbles Rising	Language	Verbal Fluency
Editors Choice	Language	Vocabulary Proficiency
Taking Root	Language	Vocabulary Proficiency
Word Snatchers	Language	Vocabulary Proficiency
Raindrops	Math	Numerical Calculation
Raindrops 2	Math	Numerical Calculation
Halve Your Cake	Math	Numerical Calculation

Continued on next page

Game Name	Cognitive Domain	Game Type
Chalkboard Challenge 2	Math	Quantitative Reasoning
Top That	Math	Quantitative Reasoning
Playing Koi	Memory	Divided Attention
Familiar Faces	Memory	Face-Name Recall
Observation Tower	Memory	Field of View
River Ranger	Memory	Information Processing
Spatial Speed Match 2	Memory	Information Processing
Speed Match 2	Memory	Information Processing
Speed Match Overdrive	Memory	Information Processing
Speed Match Web	Memory	Information Processing
Memory Matrix 2	Memory	Spatial Recall
Moneycomb	Memory	Spatial Recall
Follow That Frog	Memory	Working Memory
Memory Match	Memory	Working Memory
Memory Match 2	Memory	Working Memory
Memory Match Overdrive	Memory	Working Memory
Pinball Recall	Memory	Working Memory
Tidal Treasures	Memory	Working Memory
Rhyme Workout	Memory	Working Memory
Rotation Matrix	Memory	Working Memory
Memory Match Overdrive	Memory	Working Memory
Rotation Matrix 2	Memory	Working Memory
Memory Serves Web	Memory	Working Memory
Fuse Clues	Reasoning	Logical Reasoning
By the Rules	Reasoning	Logical Reasoning
Organic Order	Reasoning	Logical Reasoning
Word Sort	Reasoning	Logical Reasoning
Pet Detective	Reasoning	Planning
Pirate Passage	Reasoning	Planning

Continued on next page

Game Name	Cognitive Domain	Game Type
Masterpiece	Reasoning	Spatial Reasoning

The Data

The original data set used for this analysis consists of 36,297 Lumosity users who have played at least 500 times across any of the games and played them broadly (defined as 5% or less of the user's gameplays were repeats of the previously played game.) Users played on the web platform in English and were located in either the US, Canada, or Australia. These users had signed up between August 1st, 2013 and December 31st, 2016 and the data was collected between August 1st, 2013 and June 30th, 2019. This data can be found on the OSF repository for Steyvers and Schafer (2020) at `https://osf.io/g9zkf/`.

Figure A.1 is included here to give a sense of the raw learning trajectories on a single Lumosity game and to highlight the individual differences within age groups.

Table A.2 provides the breakdown of the various age groups in the sample at three different levels of training. As the amount of training increases, users drop out and the sample skews older. The first row of the table represents a sample of 36,294 users. The last row of the table represents the sample used for all of the analyses in the main text (9,923 users).

Table A.2: The percentage of users in each age bin when the sample consists of all users who played any game up to 20, 60, and 100 times.

	Age Bin							
Gameplays	20s	30s	40s	50s	60s	70s	80s	90s
20	0.051	0.055	0.104	0.239	0.320	0.184	0.046	0.001
60	0.039	0.043	0.090	0.234	0.336	0.203	0.053	0.001
100	0.032	0.034	0.077	0.224	0.350	0.222	0.060	0.001

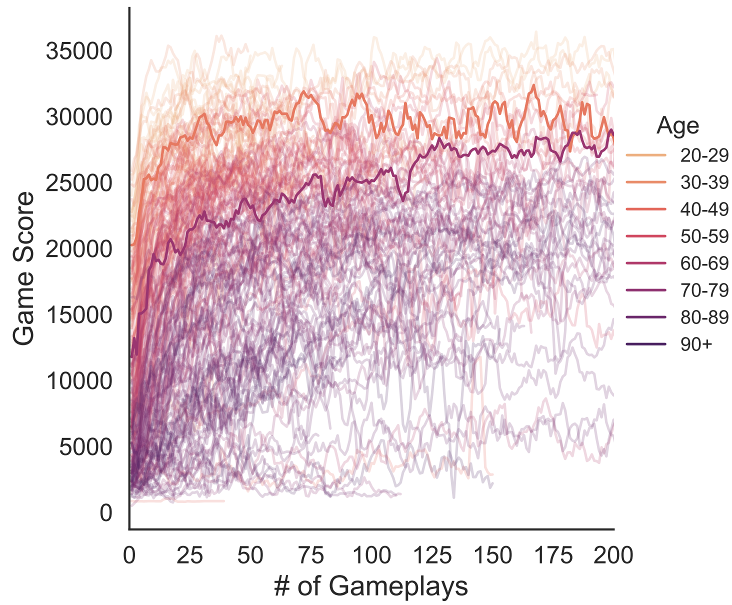

Figure A.1: A random sample of 207 players' learning trajectories during multiple plays of the task switching game Ebb and Flow. One individual from the 30-39 and 70-79 groups has been highlighted to show the initial performance gap and the later catch up potential of the older adult.

Score Normalization and Smoothing

Three methods were considered: z-scoring, min-max scaling, and the method used in Steyvers and Schafer (2020). All three produced similar results, and we chose min-max scaling due to its parsimony. First we considered all scores that were greater than 3 standard deviations above the mean score as outliers and set those scores to be 3 standard deviations above the mean. Then the scores were transformed using the min-max rule $(scores - scores_{min})/(scores_{max} - scores_{min})$ which transforms the scores such that the smallest value (0) is the lowest score and the largest value (1) is the highest score achieved (3 standard deviations above the mean).

To calculate the current best performance at every time point we took a rolling average of the normalized score over windows of 5 gameplays around the current time point. For example, to estimate a user's performance level at gameplay 10, we calculated the average

Table A.3: Difference in mean scores [95% confidence intervals] between 20 and 100 gameplays for each age bin, where "30s" means users who are 30-39 years old, "40s" means 40-49, etc.

Age	Attention	Flexibility	Memory	Reasoning	Language	Math
20s	0.137 [0.099 0.174]	0.091 [0.043 0.137]	0.125 [0.082 0.166]	0.187 [0.146 0.226]	0.15 [0.089 0.212]	0.082 [0.013 0.146]
30s	0.15 [0.122 0.177]	0.12 [0.087 0.154]	0.138 [0.106 0.17]	0.222 [0.19 0.254]	0.128 [0.079 0.181]	0.107 [0.043 0.169]
40s	0.157 [0.14 0.174]	0.148 [0.129 0.169]	0.158 [0.138 0.179]	0.204 [0.182 0.223]	0.143 [0.113 0.177]	0.105 [0.061 0.15]
50s	0.156 [0.148 0.164]	0.167 [0.158 0.176]	0.154 [0.144 0.163]	0.206 [0.196 0.216]	0.119 [0.103 0.134]	0.1 [0.083 0.119]
60s	0.135 [0.129 0.14]	0.177 [0.17 0.183]	0.137 [0.131 0.144]	0.188 [0.18 0.195]	0.096 [0.086 0.106]	0.086 [0.075 0.097]
70s	0.092 [0.087 0.098]	0.177 [0.169 0.184]	0.124 [0.117 0.131]	0.163 [0.153 0.171]	0.063 [0.054 0.073]	0.076 [0.066 0.086]
80s	0.058 [0.05 0.068]	0.151 [0.139 0.162]	0.097 [0.086 0.108]	0.139 [0.125 0.154]	0.039 [0.025 0.053]	0.057 [0.043 0.069]
90s	-0.021 [-0.069 0.025]	0.162 [0.117 0.21]	0.039 [-0.017 0.095]	0.106 [0.054 0.158]	-0.004 [-0.088 0.074]	0.05 [0.011 0.093]

of that user's scores for gameplays 8, 9, 10, 11, and 12. This helps smooth out the individual learning curve.

Additional Results

Table A.3 shows the difference in mean scores between 20 gameplays and 100 gameplays for each age bin. As reported in the main text, these differences were mainly positive and nonzero, showing that users are improving performance over practice. One exception involves the 90-95 age group on attention, memory, and language games.

Figure A.2 shows the catch up effect in greater detail. For select age comparisons we have visualized the increase in catch up probability per every extra 20 gameplays of training.

Table A.4 lists the catch up probabilities (along with the 95% credible intervals) per Lumosity game for the age group comparisons discussed in the main text.

Table A.4: Table of catch up probabilities [95% credible intervals] for the age comparisons reported in the paper on each Lumosity game. Blanks in the table mean there were less than 10 players in either age group for that particular game.

Game	80s vs 60s	70s vs 60s	70s vs 50s	60s vs 40s
Penguin Pursuit	0.469 [0.393 0.544]	0.569 [0.495 0.62]	0.363 [0.282 0.445]	0.512 [0.374 0.618]
Eagle Eye	0.453 [0.422 0.482]	0.607 [0.582 0.62]	0.41 [0.389 0.432]	0.42 [0.393 0.454]
Lost in Migration 2	0.371 [0.355 0.402]	0.583 [0.558 0.592]	0.366 [0.361 0.398]	0.427 [0.415 0.47]
Speed Pack	0.488 [0.475 0.531]	0.651 [0.633 0.668]	0.486 [0.471 0.513]	0.564 [0.518 0.583]
Train of Thought	0.398 [0.416 0.467]	0.623 [0.64 0.673]	0.477 [0.476 0.514]	0.573 [0.55 0.602]
Star Search	0.414 [0.389 0.444]	0.614 [0.599 0.635]	0.455 [0.438 0.48]	0.555 [0.518 0.588]
Trouble Brewing	0.356 [0.321 0.391]	0.564 [0.524 0.572]	0.398 [0.367 0.421]	0.469 [0.443 0.52]
Splitting Seeds	0.321 [0.262 0.377]	0.502 [0.449 0.547]	0.415 [0.349 0.475]	0.542 [0.428 0.632]
Highway Hazards	0.483 [0.451 0.53]	0.63 [0.614 0.666]	0.478 [0.446 0.515]	0.543 [0.475 0.588]
Eagle Eye 2	0.338 [0.296 0.4]	0.489 [0.451 0.541]	0.337 [0.287 0.395]	0.402 [0.299 0.503]
Color Match 2	0.526 [0.49 0.563]	0.708 [0.686 0.731]	0.577 [0.546 0.602]	0.66 [0.616 0.696]
Brain Shift 2	0.483 [0.434 0.522]	0.681 [0.645 0.701]	0.53 [0.498 0.566]	0.615 [0.558 0.656]
Ebb and Flow	0.445 [0.423 0.472]	0.633 [0.604 0.638]	0.43 [0.421 0.462]	0.519 [0.487 0.549]
Disillusion 2	0.368 [0.33 0.391]	0.546 [0.517 0.564]	0.374 [0.346 0.399]	0.47 [0.425 0.517]
Playing Koi	0.357 [0.268 0.523]	0.497 [0.422 0.641]	0.322 [0.206 0.458]	
Familiar Faces	0.636 [0.669 0.751]	0.805 [0.81 0.851]	0.782 [0.785 0.83]	0.863 [0.862 0.896]
Observation Tower	0.336 [0.275 0.397]	0.426 [0.389 0.484]	0.353 [0.305 0.426]	0.312 [0.277 0.474]
Pinball Recall	0.473 [0.424 0.512]	0.594 [0.562 0.623]	0.444 [0.408 0.476]	0.461 [0.409 0.508]
Memory Matrix 2	0.346 [0.309 0.375]	0.481 [0.453 0.503]	0.351 [0.325 0.38]	0.393 [0.348 0.428]
Speed Match 2	0.465 [0.435 0.489]	0.631 [0.597 0.631]	0.473 [0.448 0.489]	0.54 [0.502 0.568]
Spatial Speed Match 2	0.478 [0.417 0.536]	0.729 [0.675 0.746]	0.601 [0.549 0.639]	0.556 [0.481 0.611]
Tidal Treasures	0.473 [0.447 0.507]	0.584 [0.573 0.618]	0.513 [0.498 0.55]	0.586 [0.545 0.627]
River Ranger	0.416 [0.384 0.473]	0.599 [0.567 0.63]	0.4 [0.391 0.468]	0.528 [0.5 0.618]
Pet Detective	0.674 [0.653 0.719]	0.789 [0.777 0.81]	0.729 [0.71 0.749]	0.835 [0.811 0.85]
Organic Order	0.392 [0.378 0.501]	0.597 [0.588 0.673]	0.438 [0.388 0.518]	0.565 [0.461 0.67]
Masterpiece	0.465 [0.414 0.515]	0.574 [0.533 0.607]	0.466 [0.405 0.509]	0.611 [0.512 0.685]

Continued on next page

Game	80s vs 60s	70s vs 60s	70s vs 50s	60s vs 40s
Fuse Clues	0.457 [0.383 0.519]	0.654 [0.569 0.675]	0.56 [0.451 0.602]	0.629 [0.463 0.721]
Word Bubbles Rising	0.367 [0.348 0.401]	0.534 [0.52 0.559]	0.431 [0.413 0.456]	0.516 [0.481 0.546]
Word Bubbles 3	0.232 [0.188 0.286]	0.397 [0.36 0.444]	0.247 [0.208 0.308]	0.377 [0.292 0.508]
Raindrops	0.327 [0.265 0.374]	0.513 [0.457 0.557]	0.375 [0.316 0.419]	0.405 [0.315 0.455]
Chalkboard Challenge 2	0.333 [0.308 0.396]	0.476 [0.454 0.524]	0.393 [0.368 0.444]	0.496 [0.431 0.545]
Raindrops 2	0.242 [0.225 0.275]	0.423 [0.402 0.451]	0.303 [0.283 0.34]	0.432 [0.379 0.48]
Disillusion		0.689 [0.578 0.764]	0.522 [0.42 0.617]	0.465 [0.385 0.603]
Memory Match 2		0.619 [0.469 0.741]	0.442 [0.281 0.606]	
Editors Choice		0.461 [0.281 0.616]	0.586 [0.346 0.755]	
Taking Root		0.844 [0.631 0.937]		

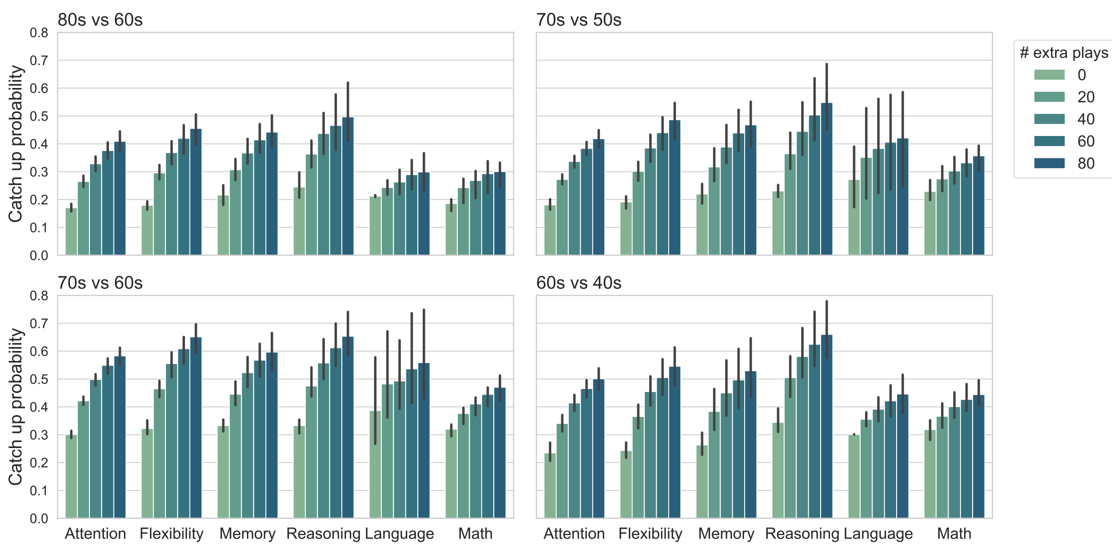

Figure A.2: The catch up potential for 60-69, 70-79, and 80-89 year olds versus younger age groups across different levels of training and different domains. Each bar represents the probability of the older adult meeting or exceeding the performance of the younger adult after randomly sampling an individual from the two age groups and comparing their performances. The younger group stays fixed at 20 gameplays while the older group is assessed at higher levels of training, up to 100 total gameplays. Catch up probability was averaged across the individual games and the error bars represent the 95% confidence interval on the average catch up probability.