Title

The Roles of Motivation and Attention in Lifelong Learning

Author

Silaj, Katie Marie

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Roles of Motivation and Attention in Lifelong Learning

A dissertation submitted in partial satisfaction

of the requirements for the degree Doctor of Philosophy

in Psychology

by

Katie Marie Silaj

2024

ABSTRACT OF THE DISSERTATION

The Roles of Motivation and Attention in Lifelong Learning

by

Katie Marie Silaj

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2024

Professor Alan Dan Castel, Chair

Rewards can enhance memory for important information; however, intrinsic motivation is also an important component of long-term learning. My dissertation explores extrinsic motivation to learn such as point values awarded on memory tasks and grades assigned in classroom settings, while considering intrinsic factors that influence learning like curiosity and interest in the material being studied. I also examined how individual differences in attention, age, and study strategies impact how learners navigate what information they should prioritize when engaging with learning materials. Value-directed Remembering (VDR; Castel et al., 2002) demonstrates the potent effects of rewards on memory for important information. Point values of varying magnitudes paired with information can motivate strategic allocation of cognitive resources that can mitigate age-related deficits in memory recall. Extrinsic rewards often accompany real-world situations to motivate better performance: grades in the classroom, bonuses in the work force, points in video games, etc. However, desired behavior and information associated with rewards are not always easy to identify in real-world contexts. Schematic support or context can make rewards more meaningful, and this may be especially

true for older adults who experience age-related declines in cognitive functioning (Castel, 2005). Additionally, extrinsic incentives may not always be enough to motivate all people. Some learners may need intrinsic sources of motivation to reach a goal such as curiosity, interest, or social connection. Thus, I explored whether learners could predict the value of information using rewards and schematic support to guide them, how being able to prioritize and identify important information relates to success in classroom learning, and how prior knowledge and curiosity influence what people remember. Overall, I find evidence that both younger and older adults can benefit from extrinsic rewards paired with explicit schematic knowledge to predict important information (Chapter 2), that selectivity in study strategies can be related to success in real classroom contexts (Chapter 3), and other factors like prior knowledge, curiosity, and collaboration can benefit learning (Chapter 4). Taken together, these findings suggest that learners may decide what is important to learn and remember through various extrinsic and intrinsic factors.

The dissertation of Katie Marie Silaj is approved.

Barbara Knowlton

Elizabeth Ligon Bjork

James W. Stigler

Alan Dan Castel, Committee Chair

University of California, Los Angeles

2024

For

Bob and Cash

who taught me so much

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 4

LIST OF TABLES

Chapter 2

Chapter 3

ACKNOWLEDGEMENTS

Portions of Chapter 2 are taken directly from the following article:

Silaj, K. M., Agadzhanyan, K., & Castel, A. D. (2023). Value-directed learning: Schematic reward structure supports value prediction. *Memory & Cognition*, *51*(7), 1527-1546. https://psycnet.apa.org/doi/10.3758/s13421-023-01406-6

To my advisor, Dr. Alan Castel, thank you for the opportunity to study memory and conduct research as a part of the Memory & Lifespan Cognition lab. I have achieved more than I ever thought possible, and your kindness, encouragement, and wisdom have supported me through this incredibly difficult and rewarding journey to earning a PhD. Thank you to my committee members, Dr. Barbara Knowlton, Dr. Elizabeth Bjork, and Dr. Jim Stigler for sharing your insight and thoughtful feedback on my work over the years. It has been an honor and a pleasure to teach, learn, and conduct research under your guidance. Thank you to Dr. Melissa Paquette-Smith for taking me under your wing early on in my graduate career. I learned to mentor undergraduate students by studying the example you modeled for me and have benefitted from the opportunity to explore other research topics with you.

Thank you to my co-authors on the work presented in this dissertation: Dr. Alan Castel, Karina Agadzhanyan, Dr. Barbara Knowlton, Dr. Tara Patterson, Fiona Gao, Alejandro Carranza, Himanshu Chaudhary, Ishita Kaloti, and Ziyi Meng. Thank you to my current and former research assistants for all the time and dedication you put into our research: Joy Hsieh, Itzel Nuñez, Emily Huang, Liora Rabizadeh, Julia Beasley, Natalie Ha, Celia Gleason, Julie Aguilar, Christina Alam, Brianna Wright, Celine Tran, and Yusuf Bello.

Thank you to Dr. Iris Firstenberg for trusting me as a Lead TA, allowing me to explore my research questions with your students, and for influencing my teaching practice. Thank you, Dr. Katie Dixie, Dr. K. Supriya, Dr. Rachel Kennison, and Dr. Virginia Sklar Muscatello for working with me through the CIRTL program and making me a better teacher. Rachel, thank you for teaching me about change leadership for inclusive teaching and learning. I am more confident than ever in my power to contribute to the positive change I want to see in institutions of higher education. Thank you to all the students and teachers I have worked with over the past 10 years who have inspired me to continue pursuing teaching and learning.

Thank you, Emily, my soul sister for life. The past 22 years would not have felt as bright or as warm without your friendship. Thank you, Aaron, for encouraging me to apply to UCLA when I didn't believe I had a chance of getting in and for being the best dad to our Cash man. Thank you, Jacks, for helping me make LA feel like home, teaching me that "home is the love you create inside you." Thank you, Lis, for always cheering me on and being silly with me. Thank you, Hunter, for all the talks, walks, and living room concerts. Thank you, Jordan, for the soccer games, writing sessions, and pool days. I would not have survived grad school without you! Dr. Kuan-Chou Chen, thank you for encouraging me to go to grad school and for your priceless advice. You believed I could do this long before I did. Thank you, mom, for always believing in me and for working so hard to give me the best life you could. I wouldn't be here without your love and hard work. Thank you, Matt Matt, for being my rock back home and for being the best dad to Hank. Thank you to my dad, Bob J. Silaj, and my dog son, Cash. This dissertation is dedicated to you both. I miss you every day and wish you could be here to celebrate with me. I wouldn't be where I am or who I am without your presence in my life.

VITA

2014	B.A., History, Social Studies Teaching Minor, Spanish Purdue University Calumet Hammond, Indiana
2014-2017	Social Studies Teacher Berkeley High School Moncks Corner, South Carolina
2017-2019	Social Studies Teacher Morton Senior High School Hammond, Indiana
2019-2024	Graduate Student Memory & Lifespan Cognition Lab University of California, Los Angeles
2020	M.A., Psychology University of California, Los Angeles
2019-2020 2020-2023 2023-2024	Teaching Assistant Teaching Associate Teaching Fellow University of California, Los Angeles
2021 2022 2024 2023-2024	CIRTL Associate CIRTL Practitioner CIRTL Scholar CIRTL Change Leader Center for the Integration of Research, Teaching, and Learning University of California, Los Angeles
2022	Project MATCH Intern Los Angeles Community College District
2023	C. Phil., Psychology University of California, Los Angeles
2023	Instructional Research Consultant Community Instructional Transformation Initiative Center for the Advancement of Teaching University of California, Los Angeles

SELECTED PUBLICATIONS

Silaj, K. M., Frangiyyeh, A., & Paquette-Smith, M. (2023). The impact of multimedia design and the accent of the instructor on student learning and evaluations of teaching. *Applied Cognitive Psychology*, *38*(1), e4143. doi:10.1002/acp.4143

Silaj, K. M., Agadzhanyan, K., & Castel, A. D. (2023). Value-directed learning: Schematic reward structure facilitates learning. *Memory & Cognition*, *51*, 1527-1546. doi:10.3758/s13421-023-01406-6

Murphy, D. H., **Silaj, K. M.**, Schwartz, S. T., Rhodes, M. G., & Castel, A. D. (2021). An own-race bias in the categorisation and recall of associative information. *Memory, 30*(2), 190-205. doi:10.1080/09658211.2021.1999982

Silaj, K. M., Schwartz, S. T., Castel, A. D., & McDonough, I. M. (2021). Is the future bright or bleak? Assessing past and future outlooks across the adult lifespan. *Gerontology and Geriatric Medicine, 7*, 1-10. doi:10.1177/23337214211046080

Silaj, K. M., Schwartz, S. T., Siegel, A. L. M., & Castel, A. D. (2021). Test anxiety and metacognition in the classroom. *Educational Psychology Review*, *33*(4), 1809-1834. doi:10.1007/s10648-021-09598-6

Whatley, M.C., Siegel, A. L. M., Schwartz, S. T., **Silaj, K. M.**, & Castel, A. D. (2020). Younger and older adults' mood and expectations regarding aging during COVID-19. *Gerontology and Geriatric Medicine, 6*, 1-9. doi:10.1177/2333721420960259

Whatley, M. C., Murphy, D. H., **Silaj, K. M.**, & Castel, A. D. (2021). Motivated memory for what matters most: How older adults (selectively) focus on important information and events using schematic support, metacognition, and meaningful goals. In G. Sedek, T. M. Hess, & D. R. Touron (Eds), *Multiple pathways of cognitive aging: Motivational and contextual influences.* (pp.40-65). Oxford University Press.

CHAPTER 1: INTRODUCTION

Learning in complex domains can be an arduous process and often learners struggle to grasp the most fundamental skills and concepts. Learning requires efficient allocation of cognitive resources such as working memory and attention. Prior work has shown that associating important information with higher extrinsic rewards can support memory, especially for older adults who face cognitive deficits due to aging (Knowlton & Castel, 2022). Further, rewarding learners for remembering important information can help them predict what new information will be important to remember in the future (Silaj et al., 2023). While rewards promote memory consolidation independently of motivation and attention (Murayama & Kitagami, 2014), these processes play an important role in long-term learning. Further, extrinsic motivation can be effective in short-term learning contexts, whereas intrinsic motivation is often needed for long-term learning and expertise (Lei, 2010). While both intrinsic and extrinsic motivation can facilitate learning, relying only on extrinsic incentives awarded based on performance can be harmful to learning outcomes. Evidence from computational models (Boedecker et al., 2013) and behavioral research (Murayama et al., 2016) suggests that extrinsic rewards can undermine intrinsic motivation, especially when rewards are contingent upon performance. When students rely on only extrinsic incentives, such as grades, to form predictions about what will be important to remember or to self-assess how well they have learned a particular concept, they may overemphasize the importance of previously learned material without forming connections between what they already know and what they still need to learn. Thus, sources of motivation to learn can also influence metacognitive processes important for academic success (Nolen, 1988).

Intrinsic sources of motivation, such as curiosity, may drive experts to continue to put effort into learning more about their own discipline because they appraise finding relationships between old and new material as exciting (Silvia & Kashdan, 2009). A novice on the other hand may feel frustrated when they are unable to see connections between old and new material. Further, learners experiencing cognitive deficits due to aging or executive dysfunction may require different types of motivational cues to make connections between concepts and identify and prioritize important information. Thus, the influence of extrinsic rewards on learning and memory may shift across the lifespan and vary within individuals based on differences in attentional processing (Simon-Dack et al., 2016). Metacognitive monitoring may support strategy use during learning and memory tasks and schematic support provides context for learners to bind to and create stronger episodic memories that enhance recall. Therefore, Chapter 2 explores value-directed learning and how it relates to metacognition and schematic support. As a measure of ecological validity, it is important to better understand how strategic allocation of attention relates to real-world outcomes, therefore Chapter 3 explores how strategic control of memory relates to classroom outcomes and how self-reported differences in attention relate to selectivity in study strategies and memory. Finally, intrinsic motivators such as curiosity can influence learning and memory, and Chapter 4 investigates this relationship. Chapter 4 also considers the effects of familiarity and prior knowledge on memory for medical and trivia facts.

Metacognition and Schematic Support

Students are often too focused on memorizing procedures instead of making connections between concepts (Fries et al., 2020), a practice which can lead to a shallower understanding of the discipline. In complex domains like math and physics, students often fail to achieve deep, transferable learning. When a student has not integrated important concepts into their general schema of the learned material, they have a more difficult time retrieving important information

(Cassady, 2004; Naveh-Benjamin, 1991). Deep, transferable learning happens over time and cannot be fully achieved or appropriately measured in one academic term. However, courses can be structured in a way to support students in making important connections when learning. For example, the "Practicing Connections" framework for organizing instruction in complex domains emphasizes making connections explicit for students, engaging students in productive struggles, and providing opportunities for students to have deliberate practice with a skill or concept to facilitate deep and transferable learning (Fries et al., 2020). Often educators use tools such as bolding or highlighting information or posting learning objectives to communicate to students what information is important to remember and what skills or connections, they want students to practice. Combined with these cues, students also typically receive grades as feedback on their progress. While these cues may be helpful for some skilled learners, others may require more explicit support to identify important information.

Schools in the United States often fail to teach students the basic skills needed to self-regulate their own learning. Even students who make it to the college level often have poor metacognitive awareness about what is important to learn, especially in courses like math (Givvin et al., 2011). Metacognition is generally defined as both an awareness of one's cognitive processes and an ability to use strategies to control these processes (Flavell, 1979). Two important types of metacognitive skillfulness are metacognitive monitoring and control. Metacognitive monitoring contributes to one's knowledge or awareness of what they already know and can be important for identifying gaps in learning (Dunlosky & Mueller, 2016; Nelson & Narens, 1990; Son & Metcalfe, 2000; Thiede & Dunlosky, 1999). Metacognitive control is involved in self-regulation of learning and is supported by active metacognitive monitoring. For example, when learning a new language, a student may recognize that they are struggling more with conjugating verbs in the past tense compared to the present tense. They may then choose to spend more time practicing the past tense in future study sessions.

Metacognitive monitoring has been studied in memory experiments. For example, when studying a series of word lists, participants may be asked to make judgments of learning (JOLs, Rhodes, 2016), such as predicting the likelihood they will remember each word on a later test (i.e., local or item-level JOL) or what percentage of words they will be able to recall (i.e., global JOL). Often, participants will receive feedback about their recall performance and can use that feedback to make decisions about what to study on a future list or whether they should adjust the memory strategy they are using during encoding.

In three experiments, Chapter 2 leverages the powerful effects of extrinsic rewards on memory and explores how they can be adaptive for making predictions about how important information will be to remember. Importantly, the memory tasks included in this chapter make important connections to participants' prior knowledge and the items used in these tasks are related to one another both in terms of semantic meaning and in their associated rewards. Further, because metacognition is an important process for adaptive learning and memory, metacognitive ratings of performance are related to actual performance on these tasks. Experts emphasize the importance of making connections between concepts explicit when teaching to help students as they expand their knowledge of a domain and provide opportunities for them to practice these connections over time (Fries et al., 2020). Therefore, in our first experiment, we provided participants with explicit instructions about the relationships between items on word lists as well as how valuable each item is based on those relationships. Through this work, we began to establish and test a schematic reward structure in which rewards meaningfully paired with point values support predictions of item importance.

Learning and Memory Strategies

Learning what is important to remember is an important skill for navigating our modern, fast-paced society where we are bombarded with extraneous information in all facets of life. Factors like age and attentional differences can make the task of remembering large amounts of information especially difficult. One way to motivate memory in a lab setting is to pair information with values of varying importance and instruct participants to prioritize higher-value items over lower-value ones, a strategy known as value-directed remembering. Importantly, though older adults typically remember less information in general, they often perform competitively with younger adults on value-directed remembering tasks, demonstrating that prioritizing important information is a strategy that can compensate for some age-related cognitive deficits (Castel et al., 2002). There is some evidence to suggest that attentional processes play a role in how values influence memory. For example, children diagnosed with Attention-Deficit/Hyperactivity Disorder (ADHD) demonstrate less strategic control of memory, suggesting a deficit in self-motivation (Castel et al., 2011). However, little is known about how the ability to selectively remember important information relates to real-world outcomes such as academic performance or how adult learners with attentional differences perform on value-directed remembering. While deficits in memory due to aging may be overcome by strategic encoding of important information, attentional deficits in adult learners may be better compensated for through intrinsic motivation.

Importantly, learning requires more than strategic control of memory and attention. Academic achievement is highly influenced by time management, study strategies, and other executive processes. Therefore, in Chapter 3, undergraduate students perform two types of memory tasks: one focused on selectively prioritizing important items (Experiment 4) and another focused on predicting what will be important to remember (Experiments 5). Further, we investigate relationships between memory performance, academic strategies, and classroom performance. This chapter also explores the relationship between symptoms of executive dysfunction in adults, such as inattentiveness, hyperactivity, and impulsivity, and strategic allocation of cognitive resources during learning and study habits.

Curiosity and Prior Knowledge

Universities and other organizations are increasingly more aware of the need for inclusive learning spaces. Though there have been efforts to increase inclusive practices for students with certain learning differences, there has been little work on how to create more inclusive learning spaces for older adult learners. This gap in the literature is problematic given the increasing numbers of older adult learners enrolled in universities and employed in the workforce (Barr, 2016; Bowen et al., 2022; Dauenhauer et al., 2022). One important consideration for age-inclusive education is to better understand how motivational goals change across the lifespan. For example, older adults may go back to school because they are interested in the material they are studying or in forming social connections whereas younger adults may be more focused on securing employment or increasing their income (Kim & Merriam, 2004). One explanation for age-related motivational shifts may be due to changes in future time perspectives across the lifespan. Typically, older adults have less time left in their lifespans compared to younger adults and this may lead older adults to be more guarded about their time, could lead them to being more selective about how they spend it, and can even impact their perceptions of the past and future (Silaj et al., 2021). The Socioemotional Selectivity Theory (Carstensen, 1995) posits that people become more selective as they age due to a perceived limitation of time and centers around two types of goals: information seeking (e.g., acquiring new information) and emotionally meaningful goals (e.g., feeling needed by others). Therefore, information that is more interesting, relevant, and connected to others may be perceived as more important to learn and remember, especially as we age.

Learning for older adults may be driven by both information seeking and emotional connection or other sources of intrinsic motivation. As mentioned previously, value-directed remembering, or learning to effortfully encode and retrieve important information, can be an adaptive strategy for overcoming certain cognitive deficits. However, older adult learners may also benefit from activating their curiosity when learning, connecting what they are learning to existing schemas, or collaborating with others during learning. These considerations are important when thinking about learning as we age. Chapter 4 examines the effect of prior knowledge on the ability to learn and remember important medical information, a topic especially important to older adult learners as medical ailments become more common as we age. This work allows for an examination of a potential interaction between age and familiarity on memory such that older adults may benefit more from explicit connections to existing schemas compared to younger adults. Then, we expand on this work by investigating the impact of collaborating to learn new information on later memory and curiosity for that information in younger adults.

In summary, Experiments 1-3 explore the effectiveness of tying extrinsic rewards to stimuli based on semantic relationships to activate prior knowledge and support predictions of item importance. These experiments also explore how metacognitive monitoring and control support recall and value predictions in adults across the lifespan. Experiments 4 and 5 investigate relationships between selective memory, accuracy in value predictions, and metacognitive awareness in a laboratory task and test whether performance on these tasks relate to actual classroom outcomes in UCLA undergraduate students, such as exam performance and study strategies. We also measure factors that could impact these relationships such as differences in self-reported motivation and attention. Finally, Experiments 6 and 7 consider the needs of older adult learners through testing memory for useful information, taking into account the impact of curiosity and familiarity when learning medical facts and trivia questions.

CHAPTER 2: VALUE-DIRECTED LEARNING

As we are often exposed to large amounts of information, people must be selective in what they choose to remember, often at the cost of other information. Research has demonstrated that participants can selectively remember important information when paired with a numeric value, a phenomenon known as value-directed remembering (VDR; Castel et al., 2002; see Knowlton & Castel, 2022 for a review). Value has a direct influence on the selective encoding of more important information over less important information (Castel et al., 2007), and the ability to succeed on a typical VDR task is related to the strategic control of memory processes (Hennessee et al., 2019).

Meaningful learning occurs when a person can interpret new information, incorporate it with prior knowledge, and apply it to novel problems (Lujan & DiCarlo, 2006). One way to make novel information more meaningful is to incorporate existing schemas, or general knowledge structures consisting of bits of information obtained through experience that guide a person's understanding of a particular concept. Schemas support learning through their influence on retrieval processes and memory reconstruction and their role in attentional and encoding processes (Bartlett, 1932; Graessner & Nakamura, 1982; Webb & Dennis, 2019). When schemas are used during learning, they can provide the background knowledge necessary to make inferences and formulate predictions in novel situations (Norman & Bobrow, 1976). For example, prior knowledge (a form of "schematic support," Craik & Bosman, 1992) can influence memory performance when learning the prices of common grocery items (Castel, 2005). Specifically, when grocery items were associated with realistic prices (market value), older adults showed similar memory performance for the studied prices as younger adults, but when studying items associated with unrealistic prices (overpriced), younger adults outperformed older adults (Castel, 2005). Furthermore, both age groups were able to identify the general category of the prices of each item and use prior knowledge to predict the new item values. Other work has shown that younger adults also benefit from schematic support when learning (Kuhns & Touron, 2020), and schemas can make learning easier even without employing strategic control processes (Whatley & Castel, 2022).

When accompanying schemas, value can communicate meaning beyond item importance. In VDR tasks, words are paired with values, and participants are instructed to prioritize high-value over low-value words (Castel et al., 2002). Typically, the words used in VDR tasks are unrelated to each other; however, if participants are presented with a series of words paired with point values based on category membership, they may notice that values repeated in the word list are connected to similarities between words sharing the same value. This process may lead to a realization of the existence of categories within the word lists, as categories are used in classifying new objects into known groups of distinct items that share similar properties (Markman & Ross, 2003).

For example, when encountering words, such as “parrot,” “owl,” and “raven” paired with a high numeric value indicating their importance, one might notice similarities between them, leading to a grouping of those words into a category of “birds,” which share similar properties, like the presence of feathers. One might also notice differences between words from the “bird” category and other words, like “carp,” “tuna,” and “shark” paired with a lower numeric value, which belong to the category “fish,” sharing similar properties, like the presence of fins. Therefore, by allocating attention towards the higher value words, one learns not only that high-value words are important to remember, but also learns which words are paired with high values.

Categories can also be used to make predictions about new items using previous knowledge about the category to which each word belongs (Anderson, 1990) and experimental research has shown that people can learn categories without prior knowledge of category labels (Fried & Holyoak, 1984). Pairing numerical values with categories creates a *schematic reward structure* in which participants may learn how values are meaningfully paired with categories through being guided by the points they earn upon recall, extending the VDR paradigm to "value-directed learning (VDL)". Thus, if the next word "trout" is presented alone without a value, the participant may use their knowledge about this schematic reward structure combined with their prior knowledge about fish properties to predict the word to be associated with a low value, demonstrating a *transfer of learning* (e.g., Perkins & Salomon, 2012; Salomon & Perkins, 1989) of the schematic rewards structure of the word lists. Such evidence of transfer of learning would demonstrate the combined effect of numerical value cues and schematic support on learning and memory.

Specifically, because numerical values are often used in rewards (e.g., course grades, bonuses, scores in a baseball game), and prior research has demonstrated that rewards can be used to selectively guide attention (Chelazzi et al., 2013), these reward-dependent effects are strategic. Rewards can be used to allocate attention to objects, features, and locations that have been accompanied by rewards. This process requires active metacognitive processes, such as metacognitive monitoring of memory processes and control of future behavior based on this monitoring (Dunlosky & Metcalfe, 2009). If rewards are assigned to items based on category membership, metacognition should play a role in decision-making about which new items will be valuable to remember based on experience studying the schematic reward structure of previously encountered items and should aid in identifying the association between stimuli and the reward associated with them. The use of rewards to facilitate selective attention in subsequent tasks requires active monitoring of performance, but metacognitive monitoring during study may also contribute to learning. Previous work has demonstrated that making global JOLs before the learning session can result in a higher transfer of learning compared to making local item-level judgments after studying each word (Lee & Ha, 2019). Thus, global JOLs may be an effective way to monitor learning.

Although engaging in metacognitive judgments is potentially important for recognizing patterns within trials and applying them in novel situations, differences in fluid intelligence may also play a role. Raven's Progressive Matrices (RPM; Hall, 1957) is a test of fluid intelligence, which measures the ability to reason and succeed in tests that require adaptation to novel situations (Cattell, 1963). Prior work has shown that higher fluid intelligence scores attained through the RPM test are related to higher selectivity in a typical VDR task when the study time for the word lists was fixed (Murphy et al., 2021). We collected measures of fluid intelligence in Experiment 1 to explore whether extracting a schematic reward structure from a series of word lists relates to abilities such as problem solving, abstract thinking, and reasoning. Here we aimed to investigate the following research questions: (1) In a value-directed remembering experiment using value cues associated with categories (as opposed to being randomly paired with individual items), does metacognitive monitoring and control impact recall performance with task experience? Does fluid intelligence relate to the proportion of high-value words recalled with task experience? (2) On a value-directed learning task where participants are asked to predict the values of items based on their experience studying related items, is fluid intelligence related to word-value pairing accuracy? Does being given specific instructions of the categories present in the word lists prior to beginning the experiment lead to higher accuracy? Do visible value cues paired with words on the studied lists result in higher accuracy on the transfer of learning task?

Does the effect of value cues on transfer performance depend on the type of schema instructions provided?

In the current study, we examined the effects of value cues and schematic support on learning in a VDL task. Specifically, in Experiment 1, participants studied word lists in which each word belonged to a specific category. Within the lists, words from a given category were associated with a point value indicating their importance. Half of the participants received specific instructions about the schematic reward structure of the word lists before beginning the task, while the other half were not made explicitly aware of the categories. Furthermore, half of the participants studied words paired with visible values during encoding while the other half studied words alone. We chose to scaffold support provided to participants to model how learning in classroom contexts and other realistic environments is often facilitated by different types of motivation. For example, some learners may benefit from the extrinsic reward of points earned upon recalling high-value words. This may lead them to notice similarities between words sharing the same value. Other learners may benefit from being reminded of their prior knowledge of a topic. For example, being told that to be studied content will contain items from categories the learner has prior knowledge of may make them more aware of the categories as they study.

After encoding each of the lists, participants provided global JOLs and completed a free recall test. After following this procedure for five lists, participants were then presented with novel words belonging to the studied animal categories and were asked to assign a value to each item based on the prior lists (immediately in Experiments 1a and 3 and after a short delay in Experiment 1b), measuring their transfer of learning. In Experiment 2, we did not provide explicit instructions about the schematic reward structures of the word lists but did manipulate the presence of value between participants. Furthermore, participants were presented with a new theme with each trial, requiring them to learn the schematic reward structures with fewer trials and adapt to new categories throughout the task. In Experiment 3 we tested the combined effects of value cues and schematic support on recall performance and transfer of learning of the word-value pairs in an adult lifespan sample.

Experiment 1a

In Experiment 1a, the type of instruction and the presence of value was manipulated between participants. Participants were either given general or specific instructions about the schematic nature of the lists and either studied the words paired with visible values or alone. After studying and recalling five lists of animal words divided into three categories where each category was associated with a low-, medium- or high-value, participants engaged in a final transfer task. Participants also completed a test of fluid intelligence after the transfer task to examine whether fluid intelligence is related to transfer of learning.

When given specific schema instructions, we expected participants to demonstrate a higher transfer of learning than those who were given general instructions as participants may benefit from an explicit cue to activate their prior knowledge of the categories within the word lists (Castel, 2005). Similarly, we expected participants who studied words paired with visible values to demonstrate a higher transfer of learning than those who studied the words alone. Allocating attention towards words paired with high values may motivate participants to notice similarities between words paired with the same value, activating their prior knowledge of the categories. Participants receiving general instructions with no value cues were included as a control condition, thus we expected them to perform at chance in both the VDR task and VDL task with no difference between the other conditions in recall performance. Furthermore, we expected all other conditions to perform significantly better than this control condition on the VDL task. Finally, we expected participants receiving both specific instructions and visible value cues to demonstrate a performance advantage on the transfer task compared to all other conditions. Being given information about the categories present in the word lists prior to beginning the task eliminates the need for the participant to discover these categories as they study the word lists. Furthermore, being reminded of what points are associated with each category throughout the task by studying the words paired with visible value cues frees up space in working memory so the participant's attention will not be divided between discovering the categories present in the word lists and binding the categories with their associated point values.

Extensive prior work using the VDR paradigm has shown that value cues influence selectivity in recall (Castel et al., 2002; Knowlton & Castel, 2022; Middlebrooks & Castel, 2018). Therefore, though this application of the VDR paradigm is novel (binding categories of words with certain values), we expected participants receiving value cues to recall a higher proportion of high-value words compared to participants who did not receive value cues during encoding. We expected participants who made higher global JOLs to recall more words with task experience as metacognitive monitoring can lead to metacognitive control when feedback is provided on performance (Lee & Ha, 2019). Murphy and colleagues (2021) found that higher fluid intelligence was related to recalling more high-value words in a VDR task. Here we explore the influence of fluid intelligence on our transfer task where participants are expected to predict the values that are associated with each word based on their experience with similar items. This work is exploratory as there is no prior work investigating this specific association; however, we expected higher fluid intelligence to be associated with higher transfer scores.

Method

Participants

Participants were 120 undergraduate students (age: 18-38, M = 20.03, SD = 2.60; gender identity: 90 women, 27 men, 1 nonbinary, 2 prefer not to say) recruited from the University of California Los Angeles (UCLA) Human Subjects Pool who were tested online and received course credit for their participation[1]. Because our task involves categorizing English nouns, we asked participants whether they were fluent in English and how old they were when they began learning English. On average, participants began learning English at 1.83 years (SD = 2.87). The sample size was selected based on prior exploratory research and the expectation of detecting a medium effect size (Knowlton & Castel, 2022; Schwartz et al., 2023). A sensitivity analysis based on the observed sample was conducted using G*Power (Faul et al., 2009). For a multiple linear regression (MLR) with 6 predictors, assuming alpha = .05, power = .80, the smallest effect the design could reliably detect is $\eta 2$ = .11.

Materials

Stimuli used in the experiment consisted of 90 English animal names (see Appendix A for word lists used in Experiment 1). When schemas already exist, memory consolidation can happen more quickly (Tse et al., 2007), so using well-known categories may be a more effective way to assess whether a schematic reward structure can be learned and applied in a relatively short laboratory task than using nonwords and novel categories. Because our sample consisted of college students who were fluent in English, we expected them to be familiar with English animal words and be able to identify common categories of animals such as mammals, birds, and fish. These words were submitted to the English Lexicon Project (ELP; Balota et al., 2007)

[1] Exclusion criteria in all studies included removing participants from the final sample who admitted to cheating on a post-task questionnaire. No participants were excluded in Experiment 1a.

database to generate measures of length (*M* = 6.02 letters per word, *SD* = 1.75), frequency in the Hyperspace Analogue to Language corpus (HAL; Lund & Burgess, 1996; *M* = 6.86 occurrences per million, *SD* = 1.64), and concreteness (*M* = 4.76, *SD* = 0.27). Each animal name belonged to one of three categories: mammals, birds, or fish. There were five animals from each category per list, and each word was associated with a value of either 1, 3, or 5, signifying the importance of the word (1 = low-importance, 3 = medium-importance, 5 = high-importance) based on animal group. Category-value pairings were counterbalanced between participants.

Design and Procedure

A 2 (Value: No Value Cue, Value Cue) x 2 (Schema: General Instructions, Specific Instructions) design was used, with all factors manipulated between participants. All participants were told that they would study six lists of words that they would later be asked to recall and that each word was associated with a value of either 1, 3, or 5. They were also told that their goal was to maximize their scores which would be based on the sum of the points associated with the words they recalled and to try to remember as many words as they could. Additional instructions were provided to participants based on their randomly assigned conditions (see Table 1): No Support, Value Support, Schema Support, and Dual Support. Value cues during encoding were either present or absent. If value cues were present, participants were instructed that each word would be paired with a value of 1, 3, or 5 and that words paired with 5 were most important. If value cues were absent, participants were told they would not be able to see the values paired with each word but were aware that some words were worth more points than others. Instructions about the schematic reward structure were either specific or general. Participants receiving specific schema instructions were informed that each word belonged to one of three categories: animals, birds, or fish. They were also told that how many points each word was worth depended on its category and were given the category-value pairings (e.g., “mammals are worth 5 points”). Participants receiving general schema instructions were not informed of the animal categories.

Condition	Additional Instructions
No Support elephant	**No Value Cue and General Schema Instructions:** *You will not be able to see the value that each word is associated with. Some words are more important than other words.*
Value Support elephant 5	**Value Cue and General Schema Instructions:** *Words paired with the value 5 are most important. Words paired with the value 3 are of medium importance. Words paired with the value 1 are least important.*
Schema Support elephant	**No Value Cue and Specific Schema Instructions:** *Each word fits into one of the three categories: mammals, birds, or fish. Words in the mammal category are worth 5 points. Mammals are animals that have hair on their bodies and drink milk when they are young (examples: rhinoceros, guinea pig, chimpanzee). Words in the bird category are worth 3 points. Birds are animals that have feathers and are born out of hard-shell eggs (examples: robin, puffin, seagull). Words in the fish category are worth 1 point. Fish are animals that live in water and have gills, scales, and fins on their bodies (examples: piranha, goldfish, tilapia).*
Dual Support elephant 5	**Value Cue and Specific Schema Instructions:** *Each word fits into one of the three categories: mammals, birds, or fish. Words in the mammal category are worth 5 points. Mammals are animals that have hair on their bodies and drink milk when they are young (examples: rhinoceros, guinea pig, chimpanzee). Words in the bird category are worth 3 points. Birds are animals that have feathers and are born out of hard-shell eggs (examples: robin, puffin, seagull). Words in the fish category are worth 1 point. Fish are animals that live in water and have gills, scales, and fins on their bodies (examples: piranha, goldfish, tilapia).*

Table 1. *Instructions for the study phase of the value-directed learning task for each condition in Experiments 1a, 1b, and 3.*

The procedure for Experiment 1a is illustrated in Figure 1. After studying all 15 words within a list, participants were asked to make a global JOL: *“What percentage of words do you think you will be able to recall in a few minutes?”* Immediately following the JOL, participants completed a 30-second distractor task where they had to reorder randomly generated sets of three numbers from largest to smallest (Unsworth, 2007). Following the distractor task, participants had 1 minute to complete a free recall test by typing as many words as they could remember from the previously studied list. Participants were then presented with their score out of a possible 45 points (five 5-point, 3-point, and 1-point words per list). We used a real-time textual similarity algorithm to account for typographical errors in participants' responses on the free recall tests for all experiments presented in this paper. Responses with at least 75% similarity to the studied word were counted as correct (Garcia & Kornell, 2014). Participants followed the same procedure for a total of five lists.

After List 5, the encoding phase ended and participants received additional instructions for the transfer task: "In this final list, you will see a series of words, each paired with an empty box. Your goal is to predict which value belongs with each word based on the five previous lists you studied. You will have five seconds to type your value prediction into the empty box. You should assign each word a value of either 1, 3, or 5." Participants had five seconds to type their prediction into the box next to each new exemplar to demonstrate transfer of learning. If participants failed to type a prediction in the box within 5 seconds, the trial was scored as incorrect and they moved on to the next item (on average, participants failed to type a prediction on 3.56% of trials in Experiment 1a, 5.39% of trials in Experiment 1b, and 4.89% of trials in Experiment 2). They then made a global JOL, completed the distractor task, free recall test, and lastly were presented with their recall score. Participants were never told how many values they correctly paired on the final list. After the transfer task, to measure their fluid intelligence, participants completed the RPM test (e.g., Jarosz et al., 2019; Staff et al., 2014) consisting of 12 patterns of varying difficulty, each of which had a piece missing. Participants were instructed to select the correct missing piece from eight multiple-choice options, and the timing was self-paced such that participants could spend as much time on each item as they liked.

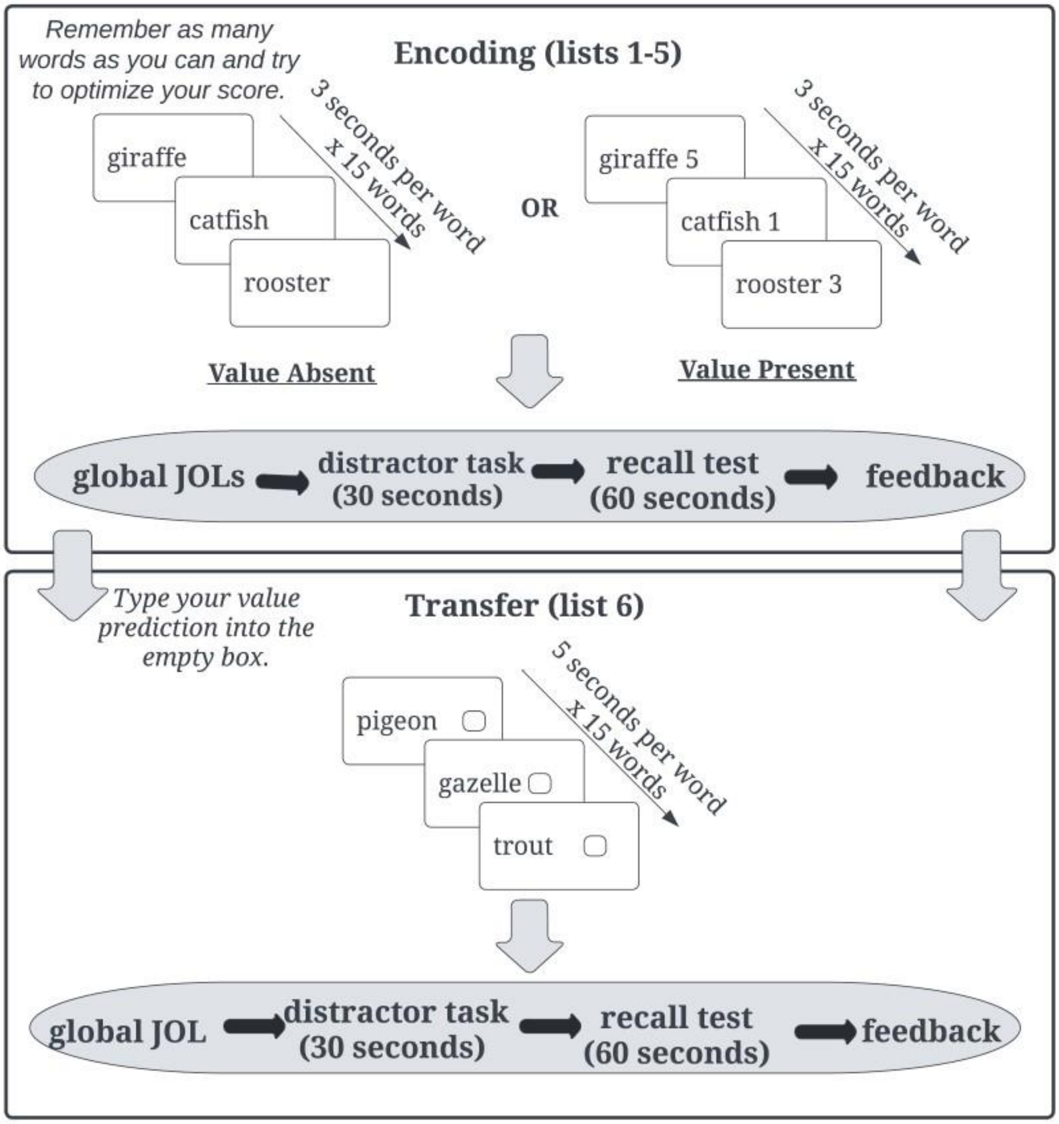


Figure 1. *Procedure for the encoding and transfer phases of the value-directed learning task in Experiments 1a, 1b, and 3.*

Results

We collected several measurements across Experiment 1a including global JOLs to measure metacognitive monitoring after studying each list. On the final list (i.e., the transfer test), participants were presented with novel animal exemplars falling into one of the same three categories present on the five studied lists: mammals, birds, and fish. For each presented item, participants were asked to type a value of either 1, 3, or 5 into the box next to the word to demonstrate a transfer of learning of the schematic reward structure of the lists. Scores on this task used in the following analyses are presented as the proportion of correct word-value pairings as a function of associated value (out of 5 trials per value). Finally, we also included fluid intelligence scores in some of the analyses which were calculated as the proportion correct (out of 12 trials) on the RPM task.

Recall. First, we sought to examine recall performance as a function of value cue, schema instructions, global JOL, list, and point value. We fit a MLR to model recall scores with value cue condition (no value cue = 0, value cue = 1), schema instruction condition (general instructions = 0, specific instructions = 1), point value, list, and global JOL. We also included interaction terms to examine both how value cue impacts the relationship between point value and recall and how metacognition impacts performance across lists. The model's explanatory power (R^2) was .23. The model's intercept was at .34, $t(1792) = 10.10$, $p < .001$. The effect of schema instructions, $b = .002$, $t(1792) = .19$, $p = .85$, was non-significant, suggesting that those receiving specific instructions performed similarly to those receiving general instructions. All other predictors were significant: The effect of value cue was significant and negative, $b = -.12$, $t(1792) = -5.30$, $p < .001$, the effect of point value was significant and positive, $b = .01$, $t(1792) = 2.87$, $p = .004$, and the interaction between value cue and point value was significant and positive, $b = .04$, $t(1792) = 6.49$, $p < .001$. Therefore, while those receiving value cues during encoding recalled significantly fewer words on average, a one-point increase in point value resulted in a .01 increase in recall score and this effect was dependent on whether value cue was present during encoding (see Figure 2). A simple slopes analysis revealed the effect of point value on recall was dependent on value cue condition such that those receiving value cues at encoding showed an increase of .06 in words recalled on average with each increase in point value, $b = .06$, $t(1792) = -12.05$, $p < .001$, and those studying the words alone still showed an increase in recall for higher-value words, but with a smaller slope, $b = .01$, $t(1792) = 2.87$, $p = .004$.

Furthermore, the effect of average JOL was significant and positive, $b = 0.34$, $t(1792) = 5.53$, $p < .001$, the effect of list was significant and negative, $b = -.03$, $t(1792) = -3.70$, $p < .001$, and the interaction between JOL and list was significant and positive, $b = .05$, $t(1792) = 2.62$, $p = .01$. These findings suggest that holding all other predictors constant, on average, a one-unit increase in average JOL on the studied lists predicted a .34 unit increase in recall performance, and recall performance decreased by .03 units with each additional list. A simple slopes analysis revealed the effect of list on recall was dependent on JOL such that those with average JOLs at the mean ($M = .39$, $SD = .18$), $b = -.01$, $t(1792) = -3.14$, $p = .002$, and 1 standard deviation below the mean, $b = -.02$, $t(1792) = -3.93$, $p < .001$, recalled fewer words with each additional list, while those with average JOLs 1 standard deviation above the mean recalled a similar number of words across lists, $b = -.002$, $t(1792) = -.45$, $p = .65$ (see Figure 2).

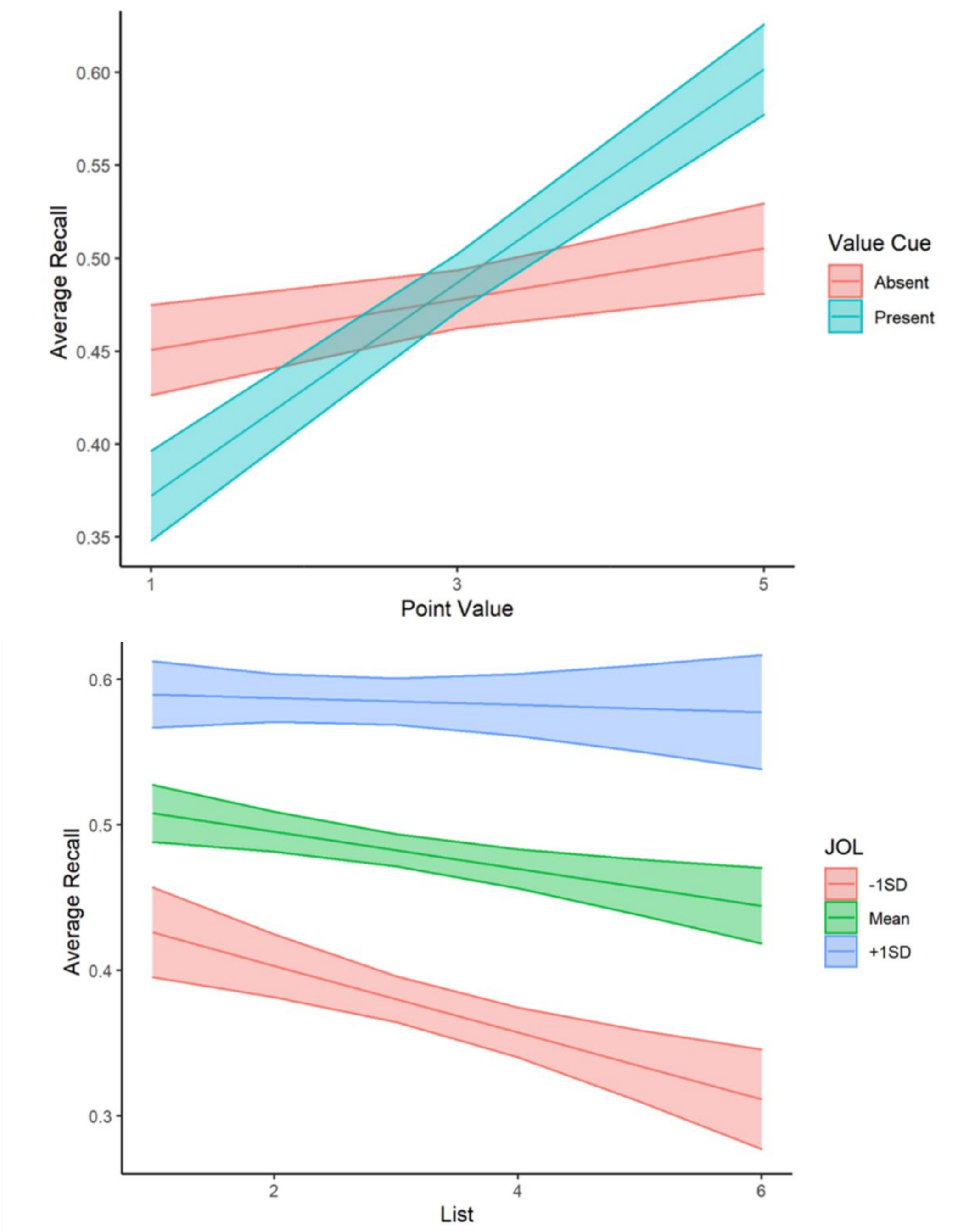


Figure 2. *Recall performance in Experiment 1a. Top graph: Average recall as a function of point value and value cue condition. Bottom graph: Average recall as a function of list and average global judgment of learning. Confidence bands represent 95% confidence intervals for the predicted values of the mean.*

Transfer of Learning. Next, we sought to examine transfer performance as a function of value cues, schema instructions, point value, and fluid intelligence. We fit a MLR to predict transfer of learning scores with value cues, schema instructions, point value, and fluid intelligence. We added interaction terms between schema instructions and value cues to evaluate whether the effect of value cues on transfer performance was dependent on the type of schema instructions participants received before beginning the task. We also added an interaction term between value cues and point value to test whether the effect of value cue on transfer performance was dependent on whether value cues were present during encoding. The model's explanatory power (R^2) was .33. The model's intercept was at 0.29, $t(323) = 6.05$, $p < .001$. The effect of point value, $b = -.01$, $t(323) = -.41$, $p = .69$, fluid intelligence, $b = .002$, $t(323) = 1.24$, $p = .22$, and the interaction between value cue and point value, $b = .004$, $t(323) = .23$, $p = .82$, were not significant. Therefore, transfer performance was not significantly influenced by fluid intelligence, how many points each word was worth upon recall, and the effect of point value did not depend on the presence of value cues during encoding. All other predictors were significant: value cues, $b = .27$, $t(323) = 3.76$, $p < .001$, schema instructions, $b = .43$, $t(323) = 10.19$, $p < .001$, and the interaction between value cues and schema instructions, $b = -.25$, $t(323) = -4.08$, $p < .001$. On average, participants receiving value cues during encoding performed significantly better on the transfer task than those who studied the words alone. Similarly, participants receiving specific schema instructions had significantly higher transfer scores than those receiving general instructions.

Furthermore, the effect of value cues on transfer performance depended on the type of schema instructions that were provided at the beginning of the experiment. Specifically, a simple slopes analysis revealed that when the schema instructions were specific, there was no additional effect of value cues on transfer performance, $b = .03$, $t(1792) = .67$, $p = .50$. However, when schema instructions were general, the presence of value cues during encoding resulted in significantly higher transfer performance compared to studying the words alone, $b = .28$, $t(1792) = 6.45$, $p < .001$ (see Figure 3).

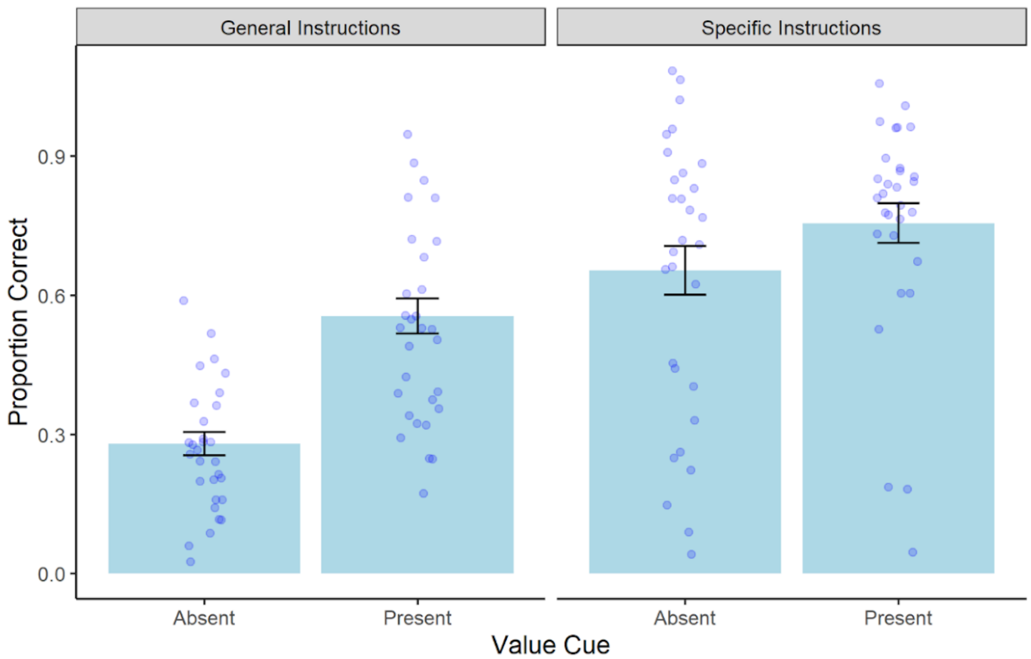


Figure 3. *Average transfer scores as a function of value cue and schema instructions in Experiment 1a. Error bars represent the standard error of the mean.*

Because there were three possible point values participants were instructed to use as predictions of items on the transfer task, performing at chance on this task would be .33, or five items correctly paired with the appropriate point values. We conducted within-condition one-sample *t*-tests to examine whether each group performed better than chance and found that all groups receiving some form of support performed significantly better than chance on this task: Value Support ($M = .56$, $SD = .28$), $t(89) = 7.42$, $p < .001$, $d = .78$, Schema Support ($M = .65$, $SD = .34$), $t(89) = 8.96$, $p < .001$, $d = .94$, and Dual Support ($M = .76$, $SD = .29$), $t(89) = 13.65$, $p <$

.001, d = 1.44. However, the No Support group who studied the words alone with general instructions performed below chance (M = .28, SD = .22), $t(89)$ = -2.28, p = .03, d = -.24.

Experiment 1a Discussion

In Experiment 1a, we aimed to extend the VDR paradigm to category learning to investigate whether participants could learn the assignment of values to words based on category membership and could transfer their learning of the schematic reward structure of the lists on a final transfer task. We scaffolded instructions about the schematic nature of the word lists to either explicitly inform participants about the existence of categories and values within the lists or to provide general instructions about how some words were more important to remember than others. Results revealed that on average, participants who studied the words paired with visible value cues performed better on the transfer task than those who studied the words alone, confirming our hypothesis that value cues would direct attention to the schematic structure of the word lists. However, those receiving value support recalled fewer words overall, but more high-value words compared to those studying the words alone. Thus, while value cues resulted in better value-based learning, it is unclear specifically through what mechanisms these cues facilitated performance on the transfer task. Additionally, specific schema instructions at the beginning of the task supported performance on the transfer task compared to having general instructions. Furthermore, it seems that the effect of value cues on transfer performance depended on the type of schema instructions participants received at the beginning of the task. Those receiving both value cues and specific schema instructions performed significantly better than those receiving value cues and general schema instructions, but similar to those studying the words alone with specific schema instructions.

We also examined whether measures of metacognition during encoding influenced recall performance and found that overall, those with higher average JOLs also recalled more words. Furthermore, though recall decreased with each list, this was only the case for participants with low to average JOLs whereas those with higher JOLs maintained similar recall scores across lists. This result was not in line with our prediction that higher JOLs would lead to higher recall with task experience, though metacognitive processes do seem to play a role in maintaining recall performance across lists. Finally, we were interested in how individual differences in fluid intelligence may relate to learning in our VDL task. Contrary to our prediction, results showed that on average, fluid intelligence did not significantly impact transfer of learning. Thus, surprisingly, differences in the ability to think abstractly and solve problems in novel situations as measured by RPM was not related to the ability to succeed in learning the schematic reward structure and applying it in novel settings. Based on our findings, in Experiment 1b, we moved the RPM test between the encoding and transfer phases of our task to act as a distractor task as opposed to using it to measure fluid intelligence.

Experiment 1b

Experiment 1b used the same materials and procedure as Experiment 1a except for two main changes: (1) Participants took the fluid intelligence test after completing the study phase (Lists 1-5). Then, after completing the fluid intelligence test, they completed the transfer task, creating a delay between the study and test phases of the experiment. (2) The pacing of the fluid intelligence test was fixed at 15 minutes to examine keep the delay between study and test constant for all participants.

In line with our results from Experiment 1a, we expected both value cues and specific schema instructions to support accuracy in the transfer task. We again expected a significant interaction between value and schema support such that value cues will provide a performance advantage when general instructions are given more so than when specific instructions are provided. We expected all conditions receiving some type of support to perform better than our control condition. Like in other VDR experiments (Castel et al., 2002; Knowlton & Castel, 2022; Middlebrooks & Castel, 2018) and Experiment 1a, we expected to observe an effect of value on recall when value cues were present during encoding demonstrating value-directed remembering. In line with our results in Experiment 1a, we expected higher JOLs to contribute to maintenance of recall performance across lists, whereas lower JOLs would be related to recalling fewer words with task experience.

Method

Participants

Participants were 120 undergraduate students (age: 18-30, *M* = 20.08, *SD* = 1.61; gender identity: 91 women, 23 men, 1 nonbinary, 1 transgender, 4 prefer not to say) recruited from the UCLA Human Subjects Pool who were tested online and received course credit for their participation[2]. On average, participants began learning English at 1.89 years (*SD* = 3.02). A sensitivity analysis based on the observed sample was conducted using G*Power (Faul et al., 2009). For a MLR with 5 predictors, assuming alpha = .05, power = .80, the smallest effect the design could reliably detect is η2 = .10.

Materials, Design, and Procedure

The design in Experiment 1b was identical to Experiment 1a. The materials and procedure in Experiment 1b were like those in Experiment 1a. However, all participants completed the RPM task after list 5 and before the final transfer task. On the RPM task, which

[2] No participants were excluded in Experiment 1b.

served as the distractor task, instead of having unlimited time for completion, participants had a time limit of 15 minutes to complete the RPM task.

Results

Recall. First, we sought to examine recall performance as a function of value cues, schema instructions, global JOL, list, and point value. We fit a MLR to model recall scores with value cues, schema instructions, point value, list, and global JOLs. We also included interaction terms to examine both how value cues impact the relationship between point value and recall and how metacognition impacts performance across lists. The model's explanatory power (R^2) was .14. The model's intercept was at .49, $t(1792) = 8.08$, $p < .001$. The effect of point value, $b = .01$, $t(1792) = 1.16$, $p = .25$, was non-significant, suggesting that on average recall performance did not depend on the point value associated with each word. All other predictors were significant: The effect of specific schema instructions (coded as 1) was significant and negative, $b = -.09$, $t(1792) = -4.81$, $p < .001$, the effect of value cues (coded as 1) was significant and negative, $b = -.24$, $t(1792) = -6.24$, $p < .001$, and the interaction between value cues and point value was significant and positive, $b = .06$, $t(1792) = 4.93$, $p < .001$. Therefore, those receiving specific schema instructions recalled significantly fewer words on average compared to those receiving general instructions. Similarly, those receiving value cues during encoding recalled significantly less words on average, but they recalled significantly more high-value words compared to low-value words, $b = 0.07$, $t(1792) = 8.13$, $p < .001$, while recall did not depend on point value for those studying the words alone, $b = 0.01$, $t(1792) = 1.16$, $p = .25$ (see Figure 4).

Furthermore, the effect of average JOL was significant and positive, $b = 0.24$, $t(1792) = 1.98$, $p = .048$, the effect of list was significant and negative, $b = -.07$, $t(1792) = -4.27$, $p < .001$, and the interaction between JOL and list was significant and positive, $b = .14$, $t(1792) = 3.95$, $p <$

.001. These findings suggest that holding all other predictors constant, on average, a one-unit increase in average JOL on the studied lists predicted a .24 unit increase in recall performance, and recall performance decreased by .07 with each additional list. However a simple slopes analysis revealed that the effect of list on recall was dependent on JOL such that those with average JOLs at the mean (M = .36, SD = .15), b = -.01, $t(1792) = -1.93$, $p = .05$ and 1 standard deviation below the mean, b = -.04, $t(1792) = -3.98$, $p < .001$, recalled fewer words with each additional list, while those with average JOLs 1 standard deviation above the mean recalled a similar number of words across lists, b = .01, $t(1792) = 1.39$, $p = .16$ (see Figure 4).

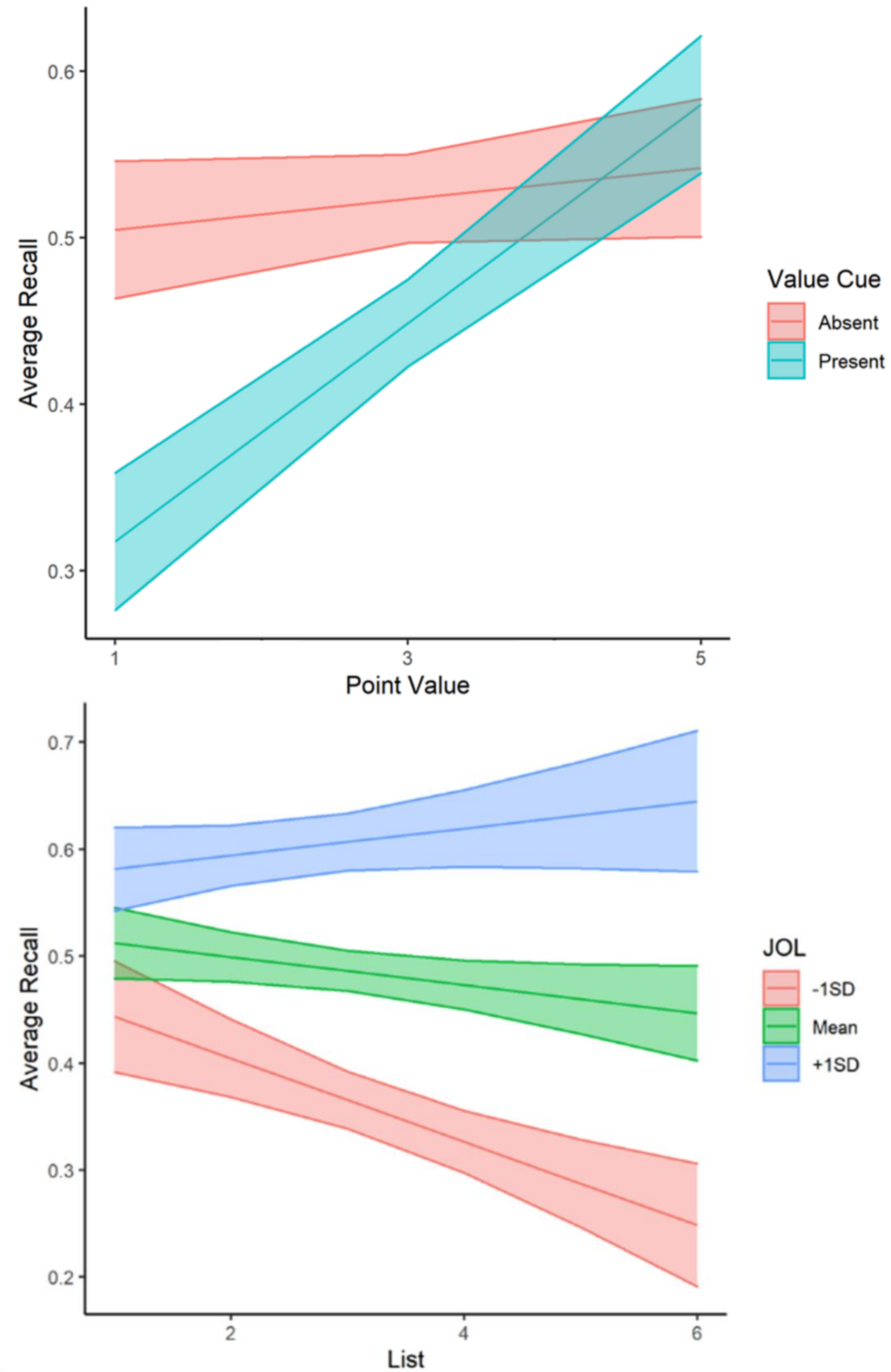


Figure 4. *Recall performance in Experiment 1b. Top graph: Average recall as a function of list and average global judgment of learning. Bottom graph: Average recall as a function of point value and value cue condition. Confidence bands represent 95% confidence intervals for the predicted values of the mean.*

Transfer of Learning. Next, we sought to examine transfer performance as a function of value cues, schema instructions, and point value. We fit a MLR to predict transfer of learning scores with value cues, schema instructions, and point value. We added interaction terms between schema instructions and value cue to test whether the effect of value cues on transfer performance was dependent on the type of schema instructions participants received prior to beginning the task. We also added an interaction term between value cues and point value to test whether the effect of value cues on transfer performance was dependent on whether value cues were present during encoding. The model's explanatory power (R^2) was .17. The model's intercept was at 0.28, $t(354) = 4.77$, $p < .001$. The effect of point value, $b = -.01$, $t(354) = -.64$, $p = .52$, and the interaction between value cues and point value, $b = .003$, $t(354) = .11$, $p = .91$, were not significant. Therefore, transfer performance was not significantly influenced by how many points each word was worth upon recall, and the effect of point value did not depend on the presence of value cues during encoding. All other predictors were significant: value cues, $b = .37$, $t(354) = 4.43$, $p < .001$, schema instructions, $b = .24$, $t(354) = 4.73$, $p < .001$, and the interaction between value cues and schema instructions, $b = -.26$, $t(354) = -3.53$, $p < .001$. On average, participants receiving value cues during encoding performed significantly better on the transfer task than those who studied the words alone. Similarly, participants receiving specific schema instructions had significantly higher transfer scores than those receiving general instructions.

Furthermore, the effect of value cues on transfer performance depended on the type of schema instructions that were provided at the beginning of the experiment. A simple slopes analysis revealed that when the schema instructions were specific, there was an effect of value cues on transfer performance, $b = .12$, $t(354) = 2.43$, $p = .02$; however, this effect was larger than when schema instructions were general, *b* = .38, *t*(354) = 7.41, *p* < .001 (see Figure 5).

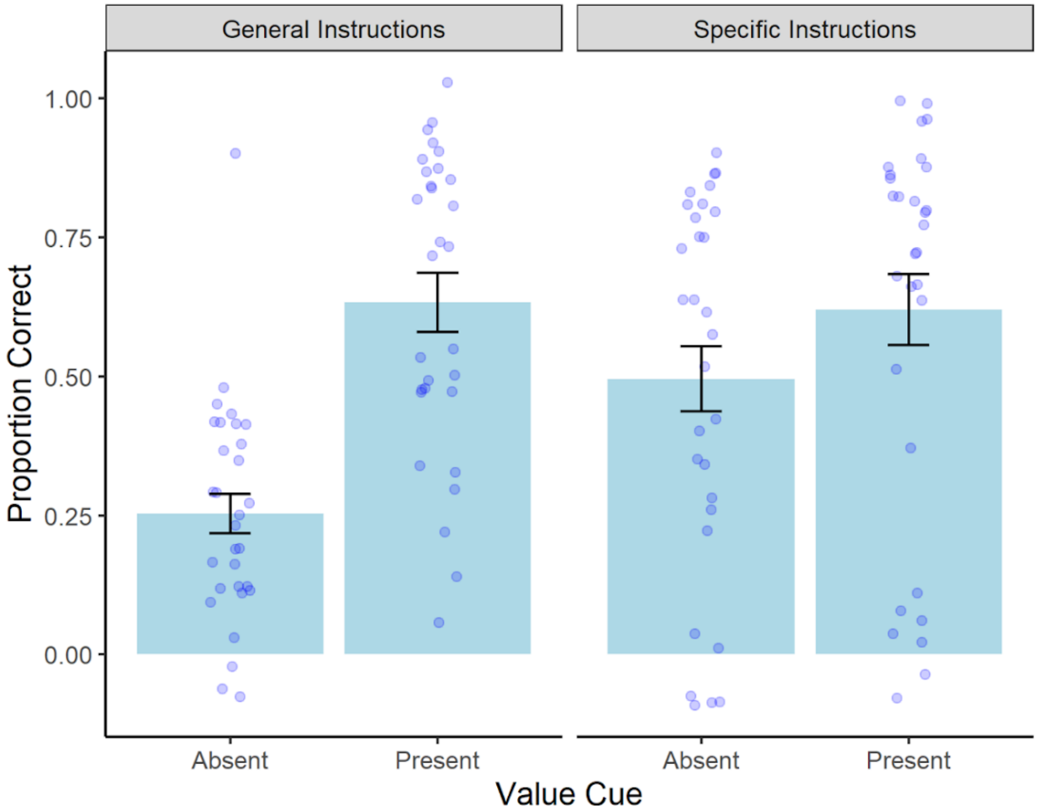


Figure 5. *Average transfer scores as a function of value cue and schema instructions in Experiment 1b. Error bars represent the standard error of the mean.*

Like in Experiment 1a, there were three possible point values participants were instructed to use as predictions of items on the transfer task, performing at chance on this task would be .33, or five items correctly paired with the appropriate point values. We conducted within-condition one-sample t-tests to examine whether each group performed better than chance and found that all groups receiving some form of support performed significantly better than chance on this task: Value Support (*M* = .63, *SD* = .33), *t*(89) = 8.69, *p* < .001, *d* = .92, Schema Support (*M* = .50, *SD* = .38), *t*(89) = 4.04, *p* < .001, *d* = .43, and Dual Support (*M* = .62, *SD* = .38), *t*(89) = 7.17, *p* < .001, *d* = .76. However, the No Support group who studied the words alone with general instructions performed below chance (M = .25, SD = .27), $t(89) = -2.77$, $p = .007$, $d = -.29$.

Experiment 1b Discussion

In Experiment 1b, we sought to replicate the findings from Experiment 1a and demonstrate that both schema instructions and value cues enhance learning in our novel VDL task even after a short delay between the study and the test. Consistent with Experiment 1a, results revealed that having specific schema instructions at the beginning of the task predicted higher transfer on the final test compared to receiving only general instructions. We also replicated the finding that studying the words paired with values predicted higher rates of transfer on the final test. Therefore, even after a short delay, participants were able to successfully demonstrate learning from the schematic reward structure when provided with either value or schema support (or both). In other words, the ability to learn the schematic reward structure and apply it in a novel test is preserved even when a short delay is introduced between learning and applying the new knowledge.

We also found that receiving specific schema instructions resulted in lower recall performance than receiving general instructions and that studying the words paired with value cues also was associated with lower recall performance. However, consistent with Experiment 1a, participants receiving value cues during encoding recalled significantly more high-value words, demonstrating strategic encoding and recall of words that would maximize their gains. Furthermore, having higher JOLs was associated with higher recall performance suggesting that metacognitive monitoring plays a role in recall performance.

Experiment 2

In Experiment 1, participants receiving value cues and/or specific schema instructions were able to learn the schematic reward structure within the 5 word lists and applied their knowledge in the final transfer task. In Experiment 2, to investigate whether participants could learn the schematic reward structure with fewer study trials and generalizing the results of Experiments 1a and 1b to other categories beyond types of animals, we exposed participants to a new theme with each list. Specifically, participants studied six lists with each list having a unique theme (e.g., plants) with three categories (e.g., flowers, trees, herbs) and completed a transfer task after each list allowing us to investigate whether participants could adapt to a new theme with each list and learn its schematic reward structure.

In Experiment 2, participants did not receive specific schema instructions as we were interested in how they might learn the schematic reward structure with task experience with or without visible value cues compared to a control condition where no value support is provided. Prior research has shown that multiple tests can enhance learning, a phenomenon known as "the testing effect" (e.g., Karpicke & Aue, 2015; Storm et al., 2010). VDR research has shown that, with task experience, people learn to be more strategic and selective in their memory (Knowlton & Castel, 2022). Therefore, we expected participants to show an increase in transfer scores with task experience with the aid of visible value cues during encoding. Because there was only one transfer trial in Experiment 1, we could not examine how well participants could perform once they were aware of the type of test to be expected.

We further explored the relationship between metacognition and the transfer of learning. In Experiment 1 we found that making higher global JOLs during encoding the maintenance of recall performance with task experience. However, we did not have item-level measures of metacognition for either the encoding or transfer task. In Experiment 2, Participants provided item-level JOLs during the encoding phase and item-level confidence judgments after each transfer trial. We expected higher metacognitive judgments to be related to higher recall and transfer scores. Finally, we expected participants studying words paired with value cues to recall more high-value words compared to the control group.

Method

Participants

Participants were 66 undergraduate students (age: 18-31, *M* = 20.64, *SD* = 2.43; gender identity: 60 women, 6 men) recruited from the UCLA Human Subjects Pool who were tested online and received course credit for their participation[3]. On average, participants began learning English at 1.85 years (*SD* = 2.74). A sensitivity analysis based on the observed sample was conducted using G*Power (Faul et al., 2009). For a MLR with 6 predictors, assuming alpha = .05, power = .80, the smallest effect the design could reliably detect is η2 = .19.

Materials

Stimuli used in Experiment 2 consisted of 180 English nouns submitted to the ELP (Balota et al., 2007) database to generate measures of length (M = 5.80 letters per word, *SD* = 1.66), frequency in the HAL corpus (Lund & Burgess, 1996, M = 7.32 occurrences per million, *SD* = 1.60), and concreteness (M = 4.74, *SD* = 0.26). Participants were exposed to 12 lists and six themes, with one list of each theme used for the encoding phase and one used for the transfer task. The six themes used were animal names (categories: mammals, birds, and fish), food items (categories: fruit, vegetables, and meat), fashion items (categories: clothing, shoes, and jewelry), household items (categories: bedroom items, bathroom items, and kitchen items), vehicles (categories: air, land, and water), and plants (categories: flowers, trees, and herbs). See Appendix

[3] One participant was excluded from analyses for admitting to cheating on a post-task questionnaire.

B for a complete list of materials.

Design and Procedure

A 2 (Value: No Value Cue, Value Cue) x 6 (Category Theme: Animal names, Food items, Fashion items, Household items, Vehicles, Plants) design was used, with value being manipulated between participants and category theme manipulated within participants. Participants were informed that there were low-value (1 point), medium-value (3 points), and high-value (5 points) words and that their goal was to maximize their scores, the sum of values associated with the words they recall. During encoding, some participants viewed only the words (No Support) while others viewed the words paired with values of either 1, 3, or 5 (Value Support). These values were assigned based on category membership and value-category pairings were counterbalanced between participants. There were five items from each category on each list. Participants in the No Support condition were informed that they would not be able to see the values on the screen with the words. Participants made local JOLs after viewing each word, indicating how likely they would recall that item on a later recall test from 0 (not at all likely) to 100 (very likely). Immediately following the encoding phase for each list, participants completed the same distractor task used in Experiment 1 where they reordered number sequences. Following the distractor task, participants had 1 minute to complete a free recall test by typing as many words as they could remember from the previously studied list and were given feedback in a form of their score out of a possible 45 points.

Participants then proceeded to complete the transfer task for that list and were exposed to a set of new words belonging to the previous list's categories. Each word appeared next to an empty box and participants were prompted to predict which value belonged with each word to measure their transfer of learning. Participants had 5 seconds to enter their value predictions into the box for each item. After predicting a value for each item, they were asked to rate how confident they were in their answers from 0 (not at all confident) to 100 (very confident). Participants followed this procedure for a total of 6 encoding-transfer phases. The complete procedure for Experiment 2 is illustrated in Figure 6.

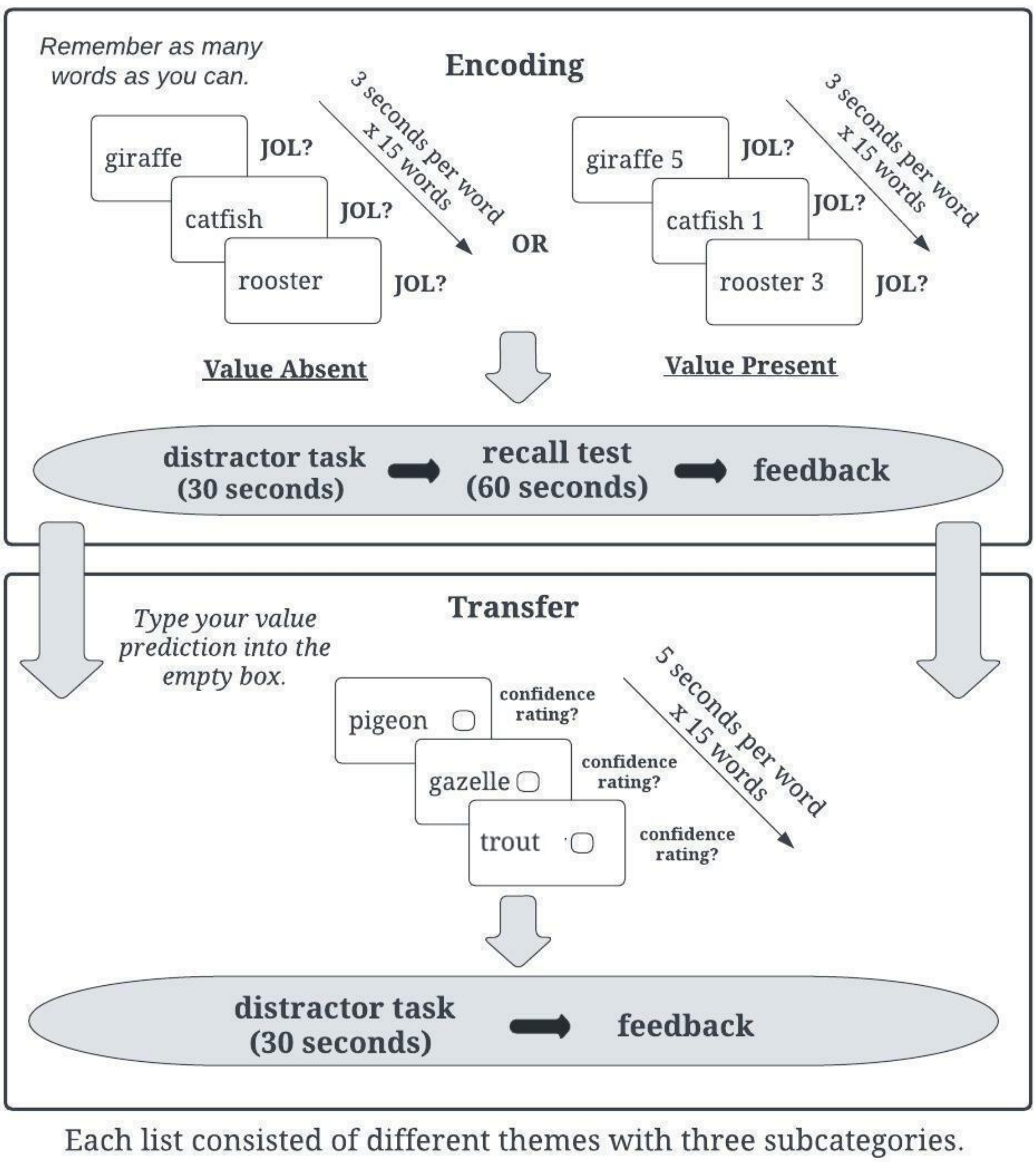


Figure 6. *Procedure for the value-directed learning task in Experiment 2.*

Results

Measures used in the following analyses include local JOLs and confidence judgments, recall performance, and transfer of learning scores. All measurements were averaged across lists by associated point value before being entered into the analyses.

Recall. First, we sought to examine recall performance as a function of value cues, local JOLs, list, and point value. We fit a MLR to model average recall scores with value cues condition, point value, list, and local JOLs. We also included interaction terms to examine both how value cues impact the relationship between point value and recall and how local JOLs impact performance across lists. The model's explanatory power (R^2) was .15. The model's intercept was at .59, $t(1177) = 13.37$, $p < .001$. The effects of value cue condition, $b = .06$, $t(1177) = 1.82$, $p = .07$, point value, $b = -0.002$, $t(1177) = -0.30$, $p = .76$, and JOL, $b = .02$, $t(1177) = .23$, $p = .82$, were non-significant, suggesting that receiving value cues at encoding did not significantly enhance recall. On average, recall performance did not change with the point value associated with each. Furthermore, local JOLs did not significantly influence average recall and the effect of point value on recall did not depend on whether value cues were present during encoding, $b = .02$, $t(1177) = 1.91$, $p = .06$.

Because we had expected an effect of value on recall for the value support condition due to prior work in VDR and our results from Experiment 1, and this interaction was of theoretical interest, we probed the interaction by conducting a post-hoc simple slopes analysis which revealed that there was a significant effect of value on recall for the Value Support condition, $b = .02$, $t(1177) = 2.39$, $p = .02$, but not for the control condition, $b = -.002$, $t(1177) = -.30$, $p = .76$ (See Table 2 for descriptive statistics). However, recall did decrease with each additional list, $b = -.05$, $t(1177) = -5.22$, $p < .001$, and the effect of list on recall was influenced by local JOLs, $b =$

.09, $t(1177) = 5.13$, $p < .001$. A simple slopes analysis revealed that those with JOLs at the sample mean ($M = .48$, $SD = .26$) recalled a similar number of words across lists, $b = -.01$, $t(1177) = -1.39$, $p = .17$. In contrast, those with JOLs 1 SD above the mean showed better recall performance with each additional list, $b = .02$, $t(1177) = 2.63$, $p = .01$ and those with JOLs 1 SD below the mean showed worse recall performance with each additional list, $b = -.03$, $t(1177) = -4.66$, $p < .001$ (see Figure 7).

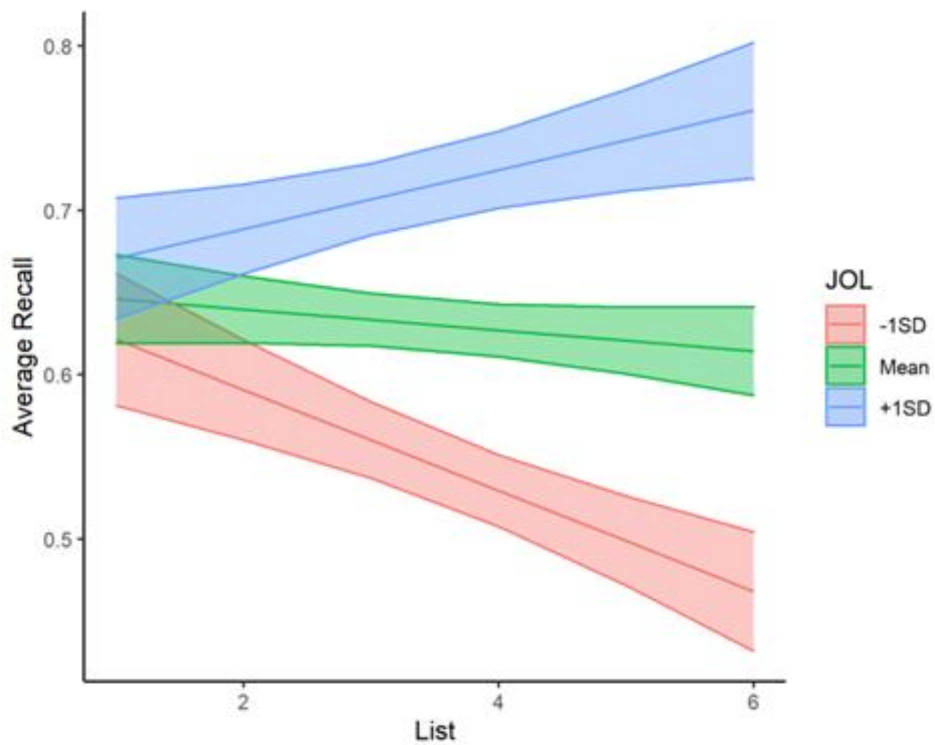


Figure 7. *Recall performance in Experiment 2 as a function of average local judgment of learning and list. Confidence bands represent 95% confidence intervals for the predicted values of the mean.*

Table 2. *Means presented as proportion correct (with standard deviation in parentheses) for recall performance as a function of point value and condition in Experiment 2.*

Condition	1-point items	3-point items	5-point items
No Support	.57 (.31)	.59 (.30)	.57 (.29)
Value Support	.63 (.28)	.67 (.26)	.72 (.26)

Transfer of Learning. To examine average transfer performance as a function of condition (no value cue = 0, value cue = 1), local confidence judgment, list, and point value. We fit a MLR to model transfer scores with value cue condition, point value, list, and local confidence judgments. We also included interaction terms to examine both how local confidence judgments and value cue condition impacted transfer performance across lists. The model's explanatory power (R^2) was .35. The model's intercept was at .10, *t*(1181) = 2.32, *p* = .02. The effects of list, *b* = -.01, *t*(1181) = -1.11, *p* = .27, and point value, *b* = .01, *t*(1181) = 1.90, *p* = .06, were non-significant, suggesting that on average, transfer performance did not increase with task experience and was not significantly impacted by the point value associated with each word. Furthermore, the effect of list did not depend on value cue condition, *b* = .02, *t*(1181) = 1.81, *p* = .07. As expected, studying the words paired with visible value cues resulted in significantly higher transfer performance, *b* = .24, *t*(1181) = 6.35, *p* < .001. Having higher average confidence judgments did significantly influence average transfer performance, *b* = .24, *t*(1181) = 3.73, *p* < .001 and the effect of list on transfer performance was dependent on average confidence judgments, *b* = .05, *t*(1181) = 3.01, *p* = .003. A simple slopes analysis revealed that those with confidence judgments 1 SD below the mean recalled a similar number of words across lists, *b* =

.01, $t(1181) = 1.08$, $p = .28$. In contrast, those with confidence judgments at the mean (M = .50, $SD = .30$) showed better transfer performance with each additional list, $b = .02$, $t(1181) = 4.50$, $p < .001$, and as did those with JOLs 1 SD above the mean, $b = .04$, $t(1181) = 5.16$, $p < .001$ (see Figure 8).

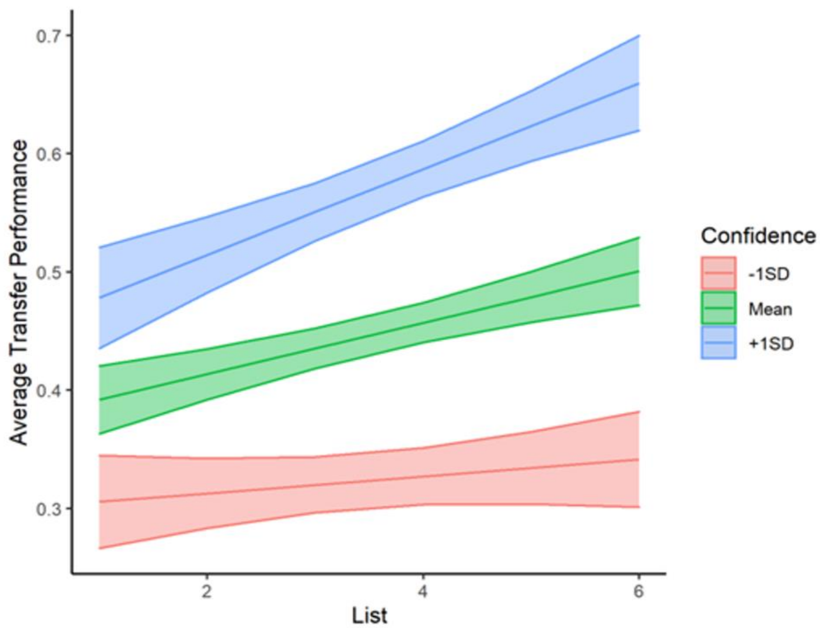


Figure 8. *Transfer performance in Experiment 2 as a function of average confidence judgment and list. Confidence bands represent 95% confidence intervals for the predicted values of the mean.*

Like in Experiment 1, chance performance on the transfer task would be 5 out of 15 correct (33%) as participants had three options for predicted values of each item. To test whether each group performed above chance, we conducted one-sample *t*-tests, which revealed that the group receiving value cues performed significantly better than chance, $t(593) = 17.94$, $p < .001$, $d = .74$, while the group studying the words alone performed significantly worse than chance, $t(593) = -4.30$, $p < .001$, $d = -.18$.

Experiment 2 Discussion

In Experiment 2, we expected transfer performance to increase with each list for participants in the value support condition as their prior knowledge of the semantic relationships between category items (McGillivray & Castel, 2017) paired with the value cues studied during encoding would enhance performance with task experience (Knowlton & Castel, 2022). As expected, participants were only able to learn the schematic reward structure when the words were paired with a visible value cue.

The value category (i.e., low, medium, high) paired with each word did not significantly influence transfer of learning scores on average, suggesting that performance on the value-pairing task did not depend on the value category. One difference between our design and typical VDR tasks is that we only used three values and these values each repeated five times on each list, while VDR tasks often use values ranging from 1 to 20 or 1 to 12 that do not repeat values within lists (Stefanidi et al., 2018). Therefore, participants in our new paradigm are learning a gist category associated with the word as opposed to an Item-level value. Therefore, with this more discrete range of values compared to the continuous values in typical VDR tasks, we may have been underpowered to detect the value-directed remembering effect in this study. Even though we did not observe a significant interaction between value cue and condition in Experiment 2, we did probe the interaction post hoc using a simple slopes analysis and found that value did impact recall for the value support condition. We also found that average confidence judgments significantly impacted average transfer of learning scores such that having average to high confidence resulted in better performance with task experience while having lower confidence was related to no increase in performance with task experience. Similarly, higher local JOLs were related to recalling more words with each list whereas lower local JOLs were related to recalling fewer words with each list.

Experiment 3

In Experiment 1 we found that both schematic and value support contributed to accurate value predictions. In Experiment 2 we observed an increase in transfer performance with task experience and established that participants formed a schematic reward structure without the presence of specific schema instructions and were able to apply the structure to new semantic themes with each list. Prior work has shown that while older adults have an overall lower memory capacity compared to younger adults, they are able to allocate their available resources towards items deemed as more important. However, according to the associative deficit hypothesis, older adults often display impairments in binding information together (Naveh-Benjamin, 2000). Though VDR tasks do require the binding of items with point values, our VDL task may introduce an additional strain on memory for older adults as they are not only required to recall higher-value items but must also recall the specific values each item is associated with. On the other hand, older adults benefit from schematic support (Castel, 2005) and our task is designed to activate prior knowledge of semantic relations between items on the word lists. Therefore, we were interested in evaluating transfer performance of older adult participants compared to a younger adult sample on a VDL task. We expected older adults to perform better than chance, demonstrating an ability to predict novel item values based on experience studying related items; however, we do expect younger adults to outperform younger adults on this task. We also analyzed the recall data and as is commonly observed in the VDR literature, we expected participants to show recall more high-value items compared to low-value items when value cue is present. Furthermore, we expected older adults to recall fewer words overall, but to recall just as many high-value words as younger adults.

Method

Participants

Participants were 200 adults recruited from Cloud Research Prime Panels (Chandler et al., 2019) and split into two age groups: 100 younger adults (age: 18-35, *M* = 28.61, *SD* = 5.29; gender identity: 53 women, 42 men, 3 nonbinary, 2 other) and 100 older adults (age: 60-99, *M* = 70.14, *SD* = 6.56; gender identity: 57 women, 43 men) who were tested online and received course credit for their participation. Participants' education backgrounds varied from having some high school to graduate degrees with 3 younger and 1 older adults reporting having some high school, 19 younger and 17 older adults being high school graduates, 24 younger and 21 older adults having some college experience with no degree, 10 younger and 15 older adults having associates degrees, 33 younger and 29 older adults having bachelor's degrees, and 11 younger and 17 older adults having graduate degrees. This sample consisted of 60 younger adult and 93 older adults participants identifying as white or Caucasian, 13 younger and 3 older adults identifying as Black or African American, 14 younger and 1 older adults identifying as Latinx or Hispanic, 5 younger and 2 older adults identifying as Asian, 3 younger adults identifying as Native American, 2 younger adults identifying as multiracial, 1 younger adult identifying as other, and 2 younger and 1 older adults who marked "Prefer not to say" for their racial identity. An additional 10 participants were excluded based on their responses to a post-experiment survey: 7 younger adult participants reported engaging in other activities during the experiment and 3 older adults reported cheating (writing down the words as they came up on the screen). Because our task involves categorizing English nouns, we asked participants whether they were fluent in English and how old they were when they began learning English. On average, participants began learning English at 0.94 years (*SD* = 3.76). The sample size was selected based on prior exploratory research and the expectation of detecting a medium effect size (Knowlton & Castel, 2022; Schwartz et al., 2023; Silaj et al., 2023).

Materials, Design, and Procedure

The design, materials (see Appendix A), and procedure (see Figure 1) in Experiment 3 was identical to Experiment 1a except that participants in Experiment 3 did not complete the fluid intelligence test.

Results

For the following analyses, we conducted multilevel models (MLMs) with items clustered within participants to account for individual-level variance. Recall and transfer performance for each item was binary (either correct or incorrect), so the regression coefficients in the models for these analyses are reported as logit units, or the log odds of correct recall. We report exponential betas (e^B) and their 95% confidence intervals ($CI_{95\%}$), which give the coefficient as an odds ratio (i.e., the odds of correctly recalling a word divided by the odds of not recalling a word). Thus, e^B can be interpreted as the extent to which the odds of recalling a word changed where values greater than 1 represent an increased likelihood of recall while values less than 1 represent a decreased likelihood of recall.

Recall performance. To examine differences in selectivity for items paired with higher values, we calculated a generalized linear mixed effects model. Predictors in the model were list, age, point value, and value cue condition, and schema instruction condition was entered into the model as a covariate. We were interested in testing whether recall for high-value items improved with practice for participant studying words paired with value cues and whether there were any age-related differences in recall for high-value information. The model was fitted using maximum likelihood estimation with Laplace Approximation. The model demonstrated a good fit to the data with an Akaike Information Criterion (AIC) of 18580.3 and a Bayesian Information Criterion (BIC) of 18687. The fixed effects of the interaction between point value and age [e^B = 1.08, $CI_{95\%}$ = 1.01-1.05, z = 2.44, p = .015], and between point value, age, and value cue condition [e^B = 0.89, $CI_{95\%}$ = 0.82-0.97, z = -2.64, p = .008] were significant. No other predictors were significant (all *p*s > .05), suggesting that overall, recall did not change with list, point value, age, value cue condition or schema condition. However, the effect of points value on recall did depend on age and value cue condition. The random effects analysis revealed significant variability in intercepts across individual participants with a variance estimate of 1.05 and a standard deviation of 1.02. The adjusted intraclass correlation coefficient (ICC) was 0.24, which indicates that approximately 24% of the total variance in test performance can be attributed to differences between participants.

A follow-up test of simple slopes was conducted to probe the three-way interaction effect between point value (low, medium, high), value cue condition (value absent, value present), and age (young, old) on recall performance. For younger adults studying the words paired with value cues, the effect of point value was significant [b = 0.15, z = 6.77, p < .001], but not for younger adults studying the words alone [b = 0.04, z = 1.79, p = .07]. For older adults studying the words paired with value cues, the effect of point value was significant [b = 0.11, z = 4.74, p < .001]. Unexpectedly, there was an effect of point value on recall for older adults studying the words alone [b = 0.11, z = 5.20, p < .001]. This could suggest that older adults who were aware of the schematic structure of the lists were prioritizing high-value items based on the categories, though we did not test this in the model.

Transfer Performance. We conducted a generalized linear mixed effects model to examine the relationship between transfer performance on the value prediction task and point values, schema condition, value cue condition, age, and their interactions, while accounting for the nested structure of the data with items clustered within each participant such that there was a random intercept for each individual included in the model. The model was fitted using maximum likelihood estimation with Laplace Approximation. The model demonstrated a good fit to the data with an AIC of 3292.2 and a BIC of 3394.3. The fixed effects of schema instructions [e^B = 4.84, $CI_{95\%}$ = 1.66-14.14, *z* = 2.88, *p* = .004] and value cue condition [e^B = 5.31, $CI_{95\%}$ = 1.82-15.53, *z* = 3.05, *p* = .002] were significant predictors of transfer performance such that participants who studied words paired with values and/or received specific instructions were more accurate in predicting the word values on the transfer task compared to those receiving general instructions and/or studying the words alone (see Figure 9). The three-way interaction between schema instructions, point values, and age was a significant predictor of transfer performance [e^B = 0.69, $CI_{95\%}$ = 0.50-0.95, *z* = -2.31, *p* = .021], suggesting that the effect of point value on transfer performance varied by age and schema instructions provided. No other predictors were significant (all *p*s < .05). The random effects analysis revealed significant variability in intercepts across individual participants with a variance estimate of 1.74 and a standard deviation of 1.32. The adjusted ICC was 0.35, which indicates that approximately 35% of the total variance in transfer performance can be attributed to differences between participants.

A follow-up test of simple slopes was conducted to probe the three-way interaction effect between point value (low, medium, high), schema instruction condition (general, specific), and age (young, old) on transfer performance. For participants receiving general instructions, the effect of point value was not significant for younger adults [*b* = 0.06, *z* = 1.01, *p* = .310] or older adults [*b* = 0.10, *z* = 1.89, *p* = .060]. For participants receiving specific instructions, the effect of point value was not significant for younger adults [*b* = -0.08, *z* = -1.42, *p* = .150]. Unexpectedly, the effect of point value on transfer performance was significant for older adults [*b* = -0.30, *z* = -5.14, *p* < .001]. This suggests that older adults who were aware of the schematic nature of the word lists correctly paired fewer high-value words with the appropriate value compared to low-value words on the transfer task.

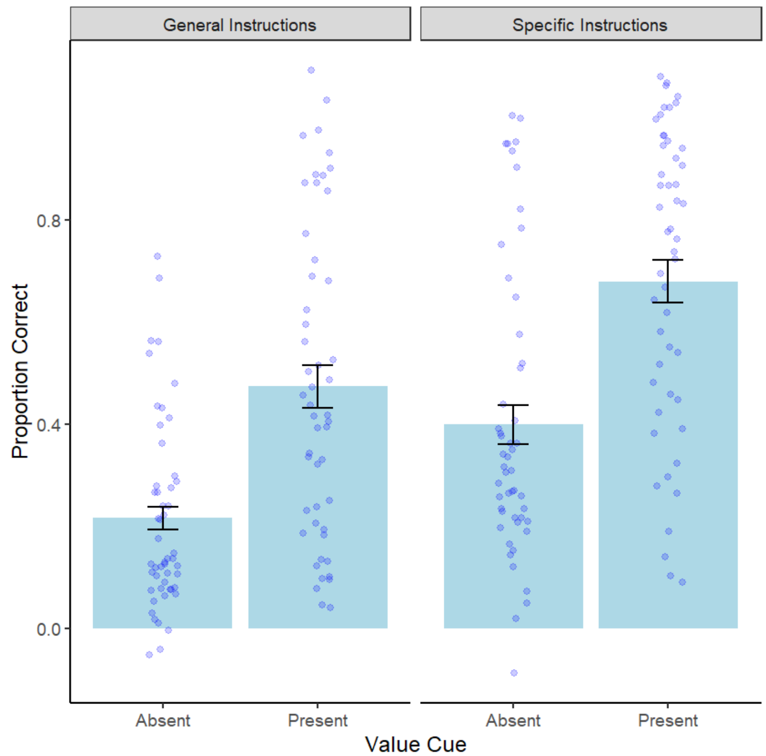


Figure 9. *Average transfer scores as a function of value cue and schema instructions averaged across age groups in Experiment 3. Error bars represent standard error of the mean.*

Experiment 3 Discussion

In Experiment 3, we expected recall performance to be higher for high-value words for those studying the words paired with value cues; however, we found no evidence of this trend. We expected older adults studying words paired with value cues to recall fewer words overall compared to younger adults, but to recall just as many high-value words. We again find no evidence for either of these predicted effects based on our model results. However, we did find that both older and younger adults recalled more high-value words than low-value words when studying the word paired with value cues as expected based on prior VDR and VDL studies.

Unexpectedly, older adults studying words alone also recalled more high-value words than low-value words. We suspect that this finding may be related to older adults either noticing or being explicitly informed of the schematic structures of the list. Future work should explore this finding further in other older adult samples to see if our finding replicates.

As expected, based on our findings in Experiments 1a, 1b, and 2, participants scored higher on the transfer task when they studied the words paired with values compared to studying the words alone and when they were explicitly informed of the schematic nature of the word lists compared to when they only received general instructions. However, we also found that there was a significant effect of point value on transfer performance for older adults when receiving specific instructions such that they correctly paired more low-value words with the appropriate point values compared to high-value words. Because we did not observe an effect of point value on transfer performance in Experiments 1a, 1b, and 2, we did not expect this finding. However, in both our recall and transfer data, there is evidence that age influences the relationship between point values and schema instructions. Future work should explore this further as older adults' familiarity with categories and larger accumulated vocabulary (Kavé & Halamish, 2015) may impact the way they engage with rewards when they are associated with semantic connections between items.

Future work should also investigate how older adults may perform on this task after a short delay. In Experiment 1b, UCLA undergraduate students were still able to use value and schematic support to provide more accurate value predictions on the transfer task, though older adults may experience more of a deficit in transfer performance after a short delay as they need to hold both the categories and their associated reward in memory and this binding may be challenging.

Chapter 2 Conclusions

The current study aimed to examine whether making participants aware of categories present within a series of word lists would facilitate a transfer of learning of the category-value pairings across lists. Prior work has shown that numerical values paired with words (Castel et al., 2002; Hennessee et al., 2019), item-location pairs (Siegel & Castel, 2018), and even faces (DeLozier & Rhodes, 2015) can enhance memory for important information. In Value-Directed Remembering experiments (VDR), values paired with words facilitate the strategic control of memory, while in Value-Directed Learning experiments (VDL) the value cues direct the learner's attention to how the words on each list are related to one another. As we used well-known semantic categories, participants could learn the schematic reward structure of the lists when provided with either explicit schematic instructions or value cues during encoding, but not when provided with no support at all.

The results of Experiment 1 suggest that both being aware of the schematic reward structure before encoding and receiving value support during encoding contributed to higher transfer of learning both with and without a short delay before encoding and transfer. However, in Experiment 1a, the effect of value cues on transfer performance was only beneficial when schema instructions were general. On the other hand, when tested after a short delay in Experiment 1b, the effect of value cues on transfer was beneficial for both types of schema instructions, but this effect was stronger for general instructions. In our transfer task, participants must make two decisions within the 5-second limit to properly predict each item's value. First, they must categorize the word within the theme of the list, and then they must recall the value belonging to each item's category. This process may utilize some form of working memory capacity (WMC), and relevant research has shown that performance in VDR tasks may be influenced by WMC (Griffin et al., 2019; Hayes et al., 2013; Knowlton & Castel, 2022) though other studies have reported little to no relationship (Castel et al., 2009; Cohen et al., 2014; Knowlton & Castel, 2022). It has been shown that people with high WMC demonstrate superior recall (Unsworth, 2007; Unsworth, 2016). Such individual differences in recall performance between people varying in WMC could be partially explained by the notion that individuals with low WMC are searching through a larger set of items than individuals with high WMC. Other work has looked at strategy use as a potential candidate for understanding the relationship between recall and WMC and results revealed that people with high WMC reported using more effective strategies, such as grouping or sentence generation, than people with low WMC (Unsworth, 2016). Decision-making in our transfer task may require a heavier cognitive load than selectivity in a free recall test and having a measure of WMC could help determine the additional benefit of value when also receiving specific schema instructions after engaging in an unrelated task as this distraction may lead to some forgetting of the category-value pairings, especially when relying on knowledge of the schematic structure alone without value cues.

We also examined recall performance and found that studying words paired with values led to lower recall overall, but higher rates of high-value words recalled demonstrating selectivity. Other work in VDR has shown that point values may cue the learner to engage in differential encoding strategies (Cohen et al., 2014; Knowlton & Castel, 2022). Having higher global JOLs resulted in stable recall across lists while having lower global JOLs resulted in a decrease in recall with task experience. These findings suggest that value cues provide support in determining what is important to remember and participants are metacognitively aware of their performance. The act of selectively recalling words may be a mechanism through which participants can notice how the words are related, which could support performance on the transfer task. This strategy may become more conscious and explicit with task experience, consistent with current models of human reward pursuit (Bijleveld et al., 2012), and suggests a metacognitive mechanism that may help guide learning.

In Experiment 2, we tested participants after each studied list and the theme of the list changed after each study-test phase. We found that receiving value support resulted in significantly higher transfer of learning scores compared to studying the words alone. Therefore, not only were participants in the Value Support condition able to learn and apply the schematic reward structures with only one study trial (compared to five trials in Experiments 1a, 1b, and 3), but they were also able to learn schematic reward structures of multiple lists each with a different theme, categories, and items.

Students are exposed to copious amounts of information and must be selective about what to study to be successful on assessments. Often, students struggle to decide what is most important to remember, though schemas and prior knowledge may guide what people tend to recall (McGillivray & Castel, 2017; Murphy & Castel, 2020, 2021). In Experiment 2, participants receiving value cues were able to adapt to new themes with each list and use both their prior semantic knowledge of the words and their task experience with the transfer test to learn not only which items were most important, but also what made an item important (i.e., category membership). Furthermore, as in Experiment 1, having higher Judgements of Learning facilitated recall during the encoding phase, suggesting metacognitive awareness of performance. Additionally, because participants made confidence judgments after each item on the transfer tests, we also observed that confidence was positively related to transfer performance. To achieve long-term learning in a domain, one must not only remember important facts and details but must also understand how important concepts and themes are connected resulting in transferable knowledge (Bransford & Schwartz, 1999; Fries et al., 2020; Greeno et al., 1993; Renkl et al., 1996). Though this deep learning occurs over a long period, we can see in our experiments that assigning items point values based on categorical features facilitated predictions of novel items' importance and metacognitive monitoring and control played an important role in this process.

A limitation of our stimuli is that some items were more prototypical of the categories they belong to than others (e.g., mammals: "giraffe" versus "whale"). Natural prototypical stimuli are typically learned more quickly than their non-prototypical counterparts (Rosch, 1973). We did not collect data on how familiar participants were with each item, which could be a factor in categorizing the words. Thus, we assumed prior knowledge of the words used in the study based on the demographics of our sample of fluent English speakers. Furthermore, while we conducted Experiment 2 in part to see whether participants could learn the schematic reward structure for categories other than types of animals, our results may not be generalizable beyond the specific well-known types of categories we chose for our experiments.

Numerous studies have found age-related differences in memory capacity; however, work using the VDR paradigm has demonstrated that older adults can be just as selective as younger adults and more selective than adolescents and children (Castel et al., 2011). However, our novel transfer task involves the binding of values to specific categories present on the word lists. The associative deficit hypothesis posits that older adults struggle with processing associative information (Naveh-Benjamin, 2000). However, Experiment 3 demonstrated that the older adults we sampled performed just as well as younger adults. However, the younger adults we sampled came from an online platform whereas the younger adults sampled in our other experiments were recruited from a university subject pool. Thus, future work should explore potential age differences in samples of older and younger adults from both university and broader community settings. However, a recent metanalysis found that age-related cognitive deficits are becoming smaller over time (Badham, 2024), thus our finding that older adults performed just as well as younger adults in Experiment 3 may replicate in other samples.

Furthermore, Castel and colleagues (2011) demonstrated that children with Attention-Deficit/Hyperactivity Disorder (ADHD) Combined type display deficits in the strategic and efficient encoding and recall of important information in a VDR task. Attention to the value-category pairings in VDL may be important in facilitating performance on the word-value pairing transfer task. Chapter 3 examines how individual differences may contribute to performance on the value-directed learning task to determine if similar attentional mechanisms contribute to performance.

Value-directed learning extends the VDR paradigm to category learning and demonstrates that the effect of value on recall persists even when there are more discrete value categories as opposed to continuous sets of values arbitrarily paired with words. We also explored how scaffolding instructions about to-be-studied items impacts the effectiveness of using value cues to identify a schematic reward structure across word lists. Using point values to group items may help learners identify what is most important to pay attention to and facilitate learning.

CHAPTER 3: REWARD-BASED LEARNING IN THE CLASSROOM

One area in which information overload can feel particularly overwhelming is in the classroom. Specifically, university students, especially on the quarter system, are expected to learn high-level concepts well enough to demonstrate mastery of them in a short period of time. To do this successfully, one must pay attention to important information and structure their individual learning and study time effectively. Patterson and colleagues (unpublished manuscript), explore the relationship between the ability to strategically control memory and academic strategies. In this work, 80 undergraduate students enrolled in a neuroscience course completed a VDR experiment and responded to a series of questions about their study habits called the Academic Strategies Questionnaire (ASQ; see Table 3). This questionnaire consisted of 10 questions where half of the questions were phrased in a way to demonstrate a more selective study habit (i.e., "When reading my assignments for school, I try to figure out what is important to take away from the text and disregard information I think is less important.") and half were phrased in a way that described less selective habits (i.e., "When reading my assignments for school, I sometimes find myself highlighting or underlining the majority of the text.").

More selective study habits may be beneficial, especially when under time pressure; however, there could be consequences to being more selective when studying if the student chooses to prioritize less important content. For example, some student may carefully read all or parts of a text where others may skim over portions of the material. Reading is typically defined as the processing of textual information with the goal of comprehending the meaning of each word, phrase, and sentence, while skimming is the process of moving one's eyes quickly through a body of text to glean the general idea of the piece or find a specific bit of information (Rayner et al., 2016). One study that surveyed 744 graduate students in clinical psychology doctoral programs reported that students on average read 330 pages a week (McMinn et al., 2009). Compliance ratings from this study suggest that graduate students completely read about 50 percent of their assigned material, while the rest was read thoroughly or skimmed. In contrast, undergraduate students may only read 20-30 percent of their required readings, which can lead to poor exam performance (Kerr & Frese, 2017). Some of the contributors to a failure to complete course readings include unpreparedness, a lack of motivation, limitations in time, and a lack of perceived reading importance. On the other hand, when under time pressure, it may be an important skill to be able to skim readings to find the most important information. Here, we were interested in better understanding how students manage their time and select their study strategies related to classroom performance as well as their ability to selectively recall high-value items.

Patterson et al. (unpublished manuscript) found that students demonstrated an overall effect of value on memory such that they recalled more high-value words compared to low-value words. Further, they found that scores on the ASQ were positively related to how selective they were on the recall test. In other words, students who reported more selective study habits also recalled more high-value words compared to low-value words. Following up on this work, the authors sought to better understand how self-reported academic strategies related to classroom exam performance. They found a quadratic relationship such that high exam performers reported high selectivity on the ASQ or low selectivity. The authors postulated that perhaps the amount of time students spent on coursework could explain how less selective studiers demonstrated high exam performance. In Experiment 4, we asked students to self-report how many hours a week they devoted to coursework to test this possibility.

Experiment 4

Participants in Experiment 4 were UCLA undergraduates enrolled in Cognitive Neuroscience, an upper-level Psychology course covering complex information. Students at this stage of their academic careers have likely developed their own routines for studying and preparing for assessments. In a course with so much information to remember, the ability to strategically allocate resources towards the most important information is an adaptive skill. Here, we were interested in whether strategic control of memory in a VDR experiment would be related to real world allocation of time and self-regulation of learning and how these metacognitive skills relate to classroom performance. Therefore, we used the ASQ developed by Patterson et al. (unpublished manuscript) to assess how selective students were regarding how they study and administered a VDR experiment mid-quarter to measure students' selective memory abilities.

Importantly, there are other factors besides study habits and memory abilities that contribute to exam performance and other course outcomes. Prior work exploring the effects of test anxiety on exam performance demonstrates that high-stakes events like exams can induce higher anxiety in students before an exam leading to lower performance outcomes, especially for students who are more anxious in general (Silaj et al., 2021). Therefore, in addition to examining the relationship between strategic control of memory, classroom performance, and study habits, we also collected self-report measures of anxiety before each exam in Experiment 4.

Metacognitive awareness can moderate the negative effect of test anxiety on exam performance such that more skilled metacognitive learners with higher test anxiety may not experience the same expected deficit in performance as their less metacognitively skilled peers with higher test anxiety. In this experiment we also asked participants to predict their score before each exam to account for the possibility that metacognitive awareness of their learning may contribute to exam performance.

In this study, we explored the following research questions: How does memory selectivity relate to self-regulation of learning? Specifically, we expected higher selectivity on the VDR task to be positively related to more strategic study habits as measured by the ASQ. Further, we examined the relationship between exam performance and strategic study habits and while controlling for self-reported study time.

Method

Participants

Participants were students enrolled in Cognitive Neuroscience at UCLA in Spring 2021 (N = 34; age: 19-28, *M* = 21.06, *SD* = 1.57; gender identity: 27 female, 7 male), Spring 2022 (N = 15; age: 20-31, *M* = 21.60, *SD* = 2.94; gender identity: 13 female, 2 male), and Spring 2023 (N = 26; age: 21-23, *M* = 21.39, *SD* = 0.59; gender identity: 12 female, 6 male, 8 preferred not to say). The students enrolled in Spring 2022 and 2023 were also asked whether they identified as first-generation college students and/or transfer students. There were 5 first-generation college students and 3 transfer students in Spring 2022 and 8 first-generation college students and 2 transfer students in Spring 2023. In all three quarters students had the same instructor and learned the same material.

Materials

Word lists. Stimuli used in the experiment consisted of 120 English nouns. These words were submitted to the English Lexicon Project (ELP; Balota et al., 2007) database to generate measures of length (*M* = 4.98 letters per word, *SD* = .97), frequency in the Hyperspace Analogue to Language corpus (HAL; Lund & Burgess, 1996; *M* = 9.07 occurrences per million, *SD* =

1.39), and concreteness (*M* = 40.05, *SD* = 21.14).

End of Experiment Survey. The ASQ items are included in Table 3. Each question asks about a particular study habit such as reading or studying for an exam and is framed as either a more selective habit (e.g., When I have limited time to read my assignments for school, I tend to scan over the text and only pay attention to the main ideas) or a less selective habit (e.g., When I have limited time to read my assignments for school, I tend to read as much of the text as I can carefully and closely, even though I probably will not finish the whole text). Each participant's score on this questionnaire is a sum of their ratings for each of the 10 statements with items 2, 4, 6, 8, and 10 reverse scored. Scores can range from 7-70 with higher scores indicating more selective study habits (*M* = 44.39, *SD* = 8.46). We also asked students to report their GPAs (*M* = 3.62, *SD* = 0.33) as a measure of academic success outside of Cognitive Neuroscience. In addition to state-anxiety ratings collected on the pre-exam surveys, we asked participants to report their trait anxiety (*M* = 4.28, *SD* = 1.68) on a single-item measure used in prior work (Silaj et al., 2021) from 1 (not at all anxious) to 7 (very much anxious) to be able to differentiate between state anxiety related to the exams and more general trait anxiety. Further we asked participants to rate how motivated they were to do coursework (*M* = 4.66, *SD* = 1.52) and how interested they were in the course (*M* = 5.43, *SD* = 1.44) from 1 (not at all) to 7 (very much). We acknowledge that participants were students earning extra credit for their participation. Thus, these students may be more motivated to do coursework and be more interested in the course content than students who did not participate. Participants also estimated how many hours they spent on coursework each week on average (*M* = 5.29, *SD* = 3.11).

Pre- Exam Surveys. We asked participants to complete a pre-exam survey within 30 minutes of beginning their exam. This survey consisted of single-item measures including a selfrating of state anxiety (i.e., “How nervous/anxious do you feel about this exam right now from 1 = not at all anxious to 7 = very much anxious?”) and a prediction of exam performance (“Please predict your score on this exam out of 100%”).

Academic Strategies Questionnaire Items
Please respond to the following questions about how you study for classes in general. Indicate the extent to which you agree or disagree with the following statement, from 1=strongly disagree to 7=strongly agree: 1. When reading my assignments for school, I try to figure out what is important to take away from the text and disregard information I think is less important. 2. When reading my assignments for school, I sometimes find myself highlighting or underlining the majority of the text. (R) 3. When I have limited time to read my assignments for school, I tend to scan over the text and only pay attention to the main ideas. 4. When I have limited time to read my assignments for school, I tend to read as much of the text as I can carefully and closely, even though I probably will not finish the whole text. (R) 5. When taking notes in class, I try to figure out what is important to remember and only write down the things I think are important. 6. When taking notes in class, I try to write down or record everything the professor says. (R) 7. When studying for exams, I try to figure out what is most important to remember and focus more time on that material. 8. When studying for exams, I try to memorize everything that has been covered in class. (R) 9. When taking exams, I tend to allocate my time according to what each question is worth, spending more time answering questions that are worth more. 10. When taking exams, I tend to treat all questions equally, regardless of what each question is worth. (R)

Table 3. *Items from the Academic Strategies Questionnaire used in Experiments 4 and 5. Items 2, 4, 6, 8, and 10 were reverse scored.*

Procedure

The procedure for the Spring 2021 course did differ slightly from the Spring 2022 and Spring 2023 courses. Most importantly, students participating in the VDR task in Spring 2021 were monitored on Zoom by a research assistant whereas participants in Spring 2022 participated independently online. To account for this difference, we conducted multi-level models with students' measures clustered within students as a predictor of each dependent measure and controlled for the quarter during which each student was enrolled. With only three quarters in this sample, we did not have enough sections to cluster students within classrooms but adding quarter as a covariate controlled for any potential effects of quarter. There were no other major differences between the procedures in Spring 2021 and the other two samples.

Participants participated in the study for extra credit. Each participant was emailed a unique study ID to use throughout the quarter. Exams were spread out across the quarter. Before each exam, participants were reminded by email to fill out the pre-exam survey. Between Exam 2 and Exam 3 participants participated in a standard VDR task consisting of 6 lists of 20 words where each word was paired with a value of 1-10. After studying each list, participants completed a free recall task and received feedback in terms of a score after each trial. After completing this procedure for 6 lists, they completed the post-experiment questionnaire.

Results

Selectivity on the VDR task. To examine selectivity in memory for recalled items, we calculated a selectivity index score for each participant on each of six lists. The selectivity index (Watkins & Bloom, 1999) has been used in prior VDR experiments (e.g., Eich & Castel, 2016; Robison & Unsworth, 2017) to measure differences in sensitivity based on the magnitude of reward paired with each word. This index is based on the participants' overall score on each word list which is calculated by summing the points paired with the words they recalled. The selectivity index is calculated using the participant's actual score, ideal score, and chance score: (actual score – chance score)/(ideal score – chance score) = selectivity index score. In general, a score closer to 1 indicates higher selectivity, whereas a score closer to -1 indicates the participant recalled more words paired with low-magnitude rewards, and a score closer to 0 indicates the participant performed at chance and was not selective.

A linear mixed model analysis was conducted to investigate the relationship between study selectivity scores as measured by the ASQ and selective memory scores as measured by the selectivity index. List was also entered into the model to test whether participants became more selective as they progressed through the experiment. The quarter each student was enrolled in the course was entered into the model to control for any potential differences between groups. Finally, students' ratings of interest in the course and motivation to do coursework were entered into the model to examine intrinsic motivation to perform well on the task. Students participated for extra credit, so these factors could influence their attention and effort during the experiment. The model demonstrated convergence with an REML criterion of 254.9. Our key variable of interest, study selectivity as measured by scores on the Academic Strategies Questionnaire, was not a significant predictor of selectivity index scores [$b = 0.003$, $SE = 0.003$, $t(67.55) = 0.83$, $p = .410$], such that memory selectivity on the VDR task was not predicted by self-reported study selectivity. Participants did become more selective across lists [$b = 0.03$, $SE = 0.01$, $t(367.04) = 3.49$, $p = .001$]. Interest in the course was a significant predictor of memory selectivity on the VDR task [$b = 0.06$, $SE = 0.02$, $t(68.08) = 2.34$, $p = .022$] suggesting that students who reported being more interested in the course were more selective on the VDR task (see Figure 10). Motivation to do coursework and quarter students were enrolled were not significant predictors of selectivity index scores (all ps > .05). The adjusted ICC for the random intercept model was 0.35 indicating that approximately 35% of the total variance in selectivity can be attributed to between-participant differences.

Though selective study habits were not significantly related to selectivity index scores, we were interested in better understanding whether selective memory was related to measures of academic achievement. Therefore, we conducted a Pearson correlation between selectivity index scores on the VDR task and self-reported GPA. Results suggested that students with higher GPAs were more selective on the VDR task, [$r = 0.23$ $t(72) = 2.03$, $p = .046$]. We also conducted a correlation between the selectivity index and average exam performance and found a significant positive relationship, [$r = 0.25$ $t(72) = 2.21$, $p = .030$], such that students who demonstrated more selective memory performance on the VDR task also performed better on the class exams on average.

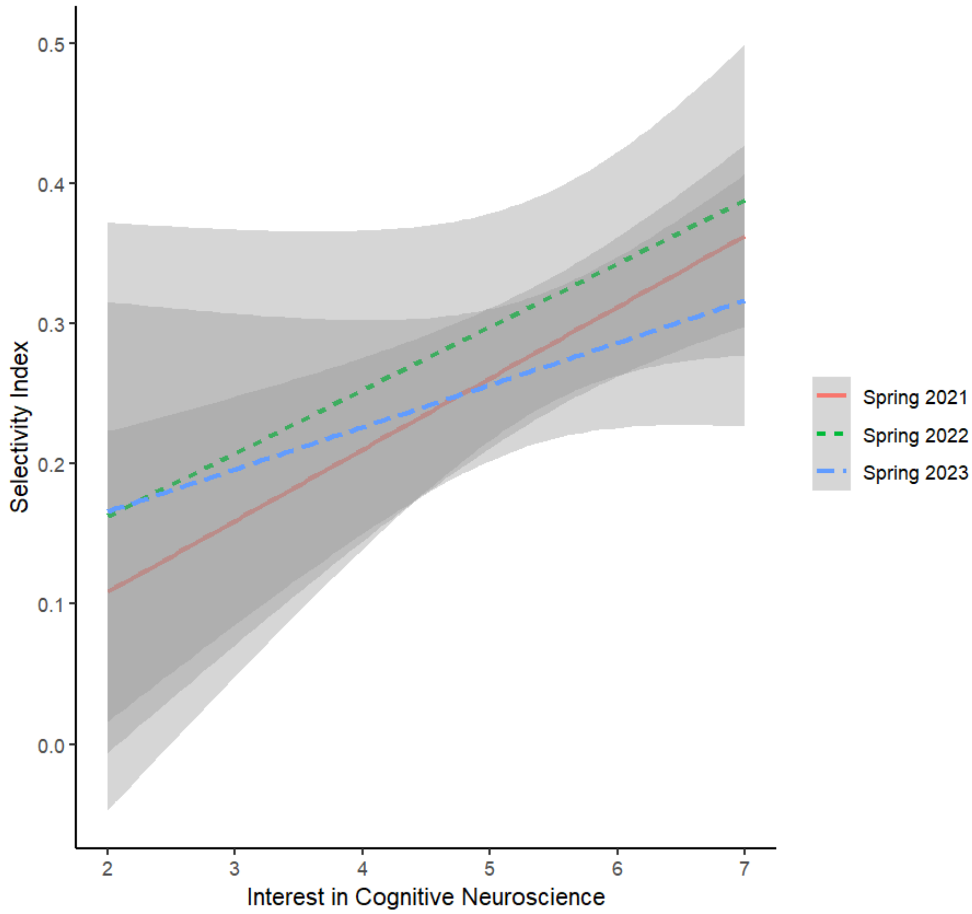


Figure 10. *Average selectivity index scores as a function of self-reported interest in Cognitive Neuroscience. Confidence bands represent 95% confidence intervals for the predicted values of the mean for each class. The lines represent slopes of interest for each class.*

Exam Performance. To investigate which variables predicted exam performance, we calculated a linear mixed model analysis. Predictors entered into the model were study selectivity as measured by the Academic Strategy Questionnaire, quarter, motivation, interest, time spent studying, GPA, predictions, selectivity index scores, pre-state anxiety, and trait anxiety. We also included exam number to see how scores changed across the quarter. The model demonstrated convergence with an REML criterion of 1210.7. Our key variable of interest, study selectivity as measured by scores on the ASQ, was not a significant predictor of exam scores [b = -0.15, SE = 0.18, t(45.12) = -0.86, p = .40], such that exam performance was not predicted by self-reported study selectivity. On average, students exam scores improved across the quarter [b = 5.19, SE = 1.02, t(106.13) = 5.08, p < .001]. Motivation to do coursework was a significant predictor of exam performance [b = 4.11, SE = 1.26, t(45.34) = 3.27, p = .002] such that those who were more motivated to do coursework performed better on the exams. Unsurprisingly, GPA was a significant predictor of exam performance [b = 35.22, SE = 4.29, t(49.36) = 8.22, p < .001] such that students with higher GPAs performed better on the exams. Finally, score predictions significantly predicted exam scores [b = 0.27, SE = 0.09, t(146.98) = 2.90, p = .004] such that students with higher metacognitive predictions of exam performance performed better on the exams. No other predictors were significant (all ps > .05), suggesting that exam scores did not significantly vary by quarter, trait anxiety, pre-state anxiety, interest in the course, or time spent studying. The adjusted ICC for the random intercept model was 0.31 indicating that approximately 31% of the total variance in exam scores can be attributed to between-student differences.

Metacognitive Accuracy. Because metacognitive predictions were a significant positive predictor of exam performance, we were interested in examining metacognitive accuracy. Here, we calculated accuracy using calibration scores by subtracting students' actual exam scores from their predictions. Thus, positive values indicate over confidence in exam performance and negative values indicate under confidence, while a score of zero represents perfect calibration. We calculated a linear mixed model analysis with the interaction between GPA and exam number as predictors of metacognitive accuracy as low-performing students have been found to be overconfident in their performance (Karaca et al., 2023) and to test for potential "underconfidence with practice" effects (Koriat et al., 2002; Silaj et al., 2021) where students become less confident with each subsequent exam. The model demonstrated convergence with an REML criterion of 1286.8. GPA was a significant predictor of metacognitive accuracy [b = -27.77, SE = 8.85, *t*(154.53) = -3.14, *p* = .002], such that students with higher GPAs were less confident in their predictions of exam score (see Figure 11). No other predictors were significant (all *p*s > .05), suggesting that accuracy did not change with each additional exam and change in accuracy with practice did not vary by GPA. The adjusted ICC for the random intercept model was 0.35 indicating that approximately 35% of the total variance in exam scores can be attributed to between-student differences.

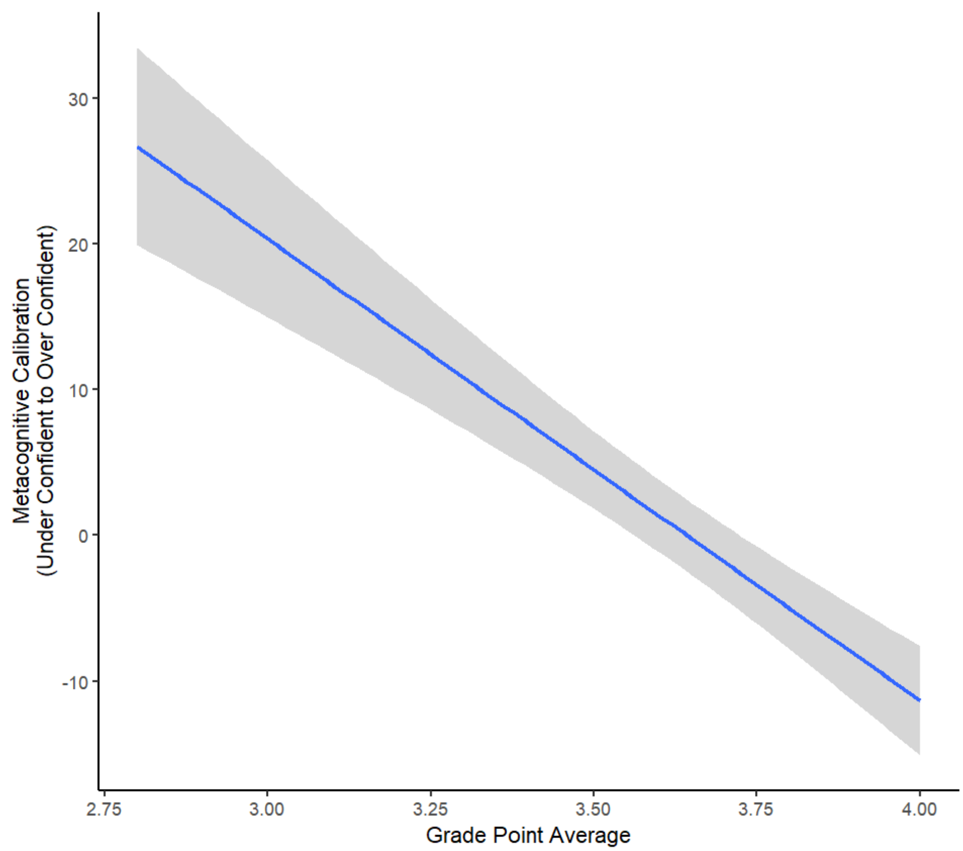


Figure 11. *Metacognitive calibration as a function of grade point average. Confidence bands represent 95% confidence intervals for the predicted values of the mean.*

Experiment 4 Discussion

Students in college courses are exposed to large amounts of new information and need to make decisions about how to regulate their learning as they prepare for exams. Strategic allocation of study time and effective strategy use is crucial to performing well in many courses. Value-directed remembering (VDR) research has shown that people can strategically prioritize memory for high-value information compared to less important information. This may be an important skill in real-world settings when learning new information as novices may struggle to distinguish important information from less relevant details. However, it is unclear how memory selectivity relates to how students study in university settings or how they perform in their courses. Following up on the work of Patterson and colleagues (unpublished manuscript), we investigated the relationship between memory selectivity, study habits, and performance in a Cognitive Neuroscience course at UCLA.

In our sample, memory selectivity as measured by selectivity index scores on a VDR task was not related to self-reported study selectivity as measured by the Academic Strategies Questionnaire (ASQ) created by Patterson and colleagues (unpublished manuscript). However, our sample may have been underpowered to detect an effect. Future work should further examine the construct that this questionnaire is measuring and how it relates to memory selectivity. Interestingly, selectivity index scores were related to interest in the course suggesting that there may have been an intrinsic incentive to be selective in our VDR task in addition to the extrinsic rewards earned during the experiment and the extra credit earned for participating. We did find two interesting correlations between memory selectivity and GPA and memory selectivity and average exam performance. Therefore, it seems that students who prioritized high-value words on the VDR task tended to perform better on the course exams and reported higher GPAs. To our knowledge this is the first study to observe relationships between memory selectivity and real-world academic outcomes.

We did not find a relationship between selectivity in study strategies and exam performance. The exams taken in this course consisted of short-answer questions. Future work should explore the relationship between study selectivity and performance on other exam formats. Self-reported motivation to do coursework predicted exam scores, suggesting a relationship between intrinsic motivation and exam grades. We did not ask students why they were motivated to do coursework, though interest in the course and motivation to do coursework were significantly correlated [$r = 0.64$, $t(52) = 6.00$, $p < .001$]. However, these students may also be extrinsically motivated to earn a good grade.

Metacognitive predictions were significant predictors of exam performance and students with higher grade point averages were more underconfident in their predictions compared to students with lower grade points averages who were more overconfident. This finding is in line with other work that has found that low-performing students are overconfident in their predictions of performance despite practice and feedback from multiple testing sessions (Karaca et al., 2023). However, we did not observe a change in metacognitive accuracy with practice across the quarter.

Here, interest in course content seems to motivate students to perform well on a VDR experiment even though the extra credit points were not tied to how well they performed. Motivation and metacognition were positively related to exam performance on short-answer exams in a Cognitive Neuroscience course. In Experiment 5 we explore these variables further in a Research Methods course with a much larger sample size.

Experiment 5

In Experiment 4 we began exploring the relationships between selective study habits, strategic control of memory, and classroom performance. In Experiments 1, 2, and 3 we found that values paired with category exemplars aided in more accurate value predictions of new exemplars on later lists of words. Experiment 5 aimed to further examine how selectivity in memory and in study habits relates to classroom performance. Further, because our VDL task is more novel than VDR tasks, we were interested in better understanding how the study schedule during the encoding phase relates to transfer performance on the final test. Specifically, in Experiments 1, 2, and 3, participants studied lists of words, completed a short distractor task, and then had the opportunity to recall the studied words in a selective order. Perhaps the recall test is when participants notice the categories present in the word lists and how they relate to the point values. Participants may then use the feedback provided after each trial to confirm how valuable the words they recalled were. On the other hand, participants could potentially notice the categories from simply studying the word lists without the opportunity to recall the words. In Experiment 5, we manipulated whether participants restudied or recalled each word list before the final transfer task between trials. Additionally, we were interested in how point values motivated different types of learners to selectively recall high-value words and accurately predict the values of new category exemplars, and how performance on these tasks related to their self-reported study habits and exam performance.

In Experiment 5, we collected data from undergraduate students enrolled in Research Methods in Psychology during the Winter and Spring Quarters of 2023. Students participated in an active learning module for 5 points of course credit. They completed a version of our VDL task from Experiment 2 and then completed a post-experiment survey before engaging in other course-related activities that were a part of the module. This group of students may be slightly less advanced in their academic careers compared to students in Experiment 4 as they were enrolled in a required, lower-level Psychology course. Fortunately, because Research Methods is a prerequisite for officially declaring the Psychology, Cognitive Science, and Psychobiology majors, over 400 students are typically enrolled each quarter, giving us a larger sample size to investigate the relationships between strategic memory, study habits, and exam performance. Additionally, this course had two high-stakes multiple-choice exams compared to the two short answer exams given in the course in Experiment 4. Study habits may vary based on experience with college courses and the type of exam, thus these students who were less advanced in their college careers and were preparing for more high-stakes exams may prepare differently than the more advanced students in Experiment 4.

Prior work has shown that students with symptoms of ADHD report using more surface compared to deep study approaches (Simon-Dack et al., 2016) and often are more motivated by external rewards like grades rather than a deeper interest in the material which can negatively impact their learning outcomes (Carlson et al., 2002). However, students with more ADHD symptoms do perform better on academic tasks that are more salient or interesting (Zentall & Shaw, 1980). Therefore, we were interested in exploring how individual differences in ADHD symptoms related to exam grades and memory selectivity in Experiment 5. We expected students with more symptoms of ADHD to perform worse on exams. However, we suspected that these same students may also perform worse on the VDR and VDL tasks as prior work has shown that children with ADHD combined type recalled fewer high-value words on a VDR task than those without ADHD and those with ADHD inattentive type (Castel et al., 2011). On the other hand, because students were participating in an engaging laboratory activity, we expected that students with more symptoms of ADHD could perform just as well as students with fewer symptoms.

A second aim of this experiment was to better understand how participants identify connections between each category and its associated reward in the VDL task. In Experiments 1-3, participants were always given the opportunity to recall studied words before engaging in the final transfer task where they were asked to predict the values of new, related words. One possibility is that this learning occurs during encoding as participants could notice the values are meaningfully tied to a particular category as they study each item. On the other hand, because participants are instructed to prioritize higher-value items over lower-value ones, when recalling words in a selective order, participants could notice groups of words as they type them into the text box on the free recall test. The testing effect (also known as retrieval practice) shows that engaging in retrieval of previously learned material through testing leads to greater retention of learned material in both laboratory and applied settings (McDaniel et al., 2007; Roediger & Karpicke, 2006). Prior work has shown that test-enhanced learning through retrieval practice can enhance learning transfer more than non-retrieval strategies, such as restudying (Pan & Rickard 2018).

A study conducted by Karpicke and Blunt (2011) had students engage in either retrieval practice, elaborative study, or simple restudy after studying a science text. For the elaborative study condition, researchers had participants draw a concept map, allowing them to connect nodes of information together. Results showed that retrieval practice led to the best performance on a final test a week after the experiment. In the current study, we manipulated whether participants restudy the same words in a random order or engage in free recall of the studied words before being presented with a novel list and asked to make value predictions of the new words.

First, regarding the study schedule manipulation, we expected higher transfer performance on the recall trials compared to the restudy trials. We expected the accuracy of value predictions to improve with each list and that this effect would be enhanced on recall trials compared to restudy trials. Furthermore, we were interested in exploring whether ASQ scores were related to selectivity index scores, transfer performance, and exam performance. We expected students with more symptoms of ADHD to have less selective study habits and lower selectivity index scores (Castel et al., 2011). Further, we expected these students to have lower scores on exams and were interested in looking at how interest and motivation impacted these relationships.

Method

Participants

Participants were 510 UCLA undergraduate students enrolled in Research Methods in Psychology during the Winter (N = 254; age: 18-41, *M* = 21.76, *SD* = 2.63; gender identity: 189 female, 54 male, 3 nonbinary, 7 other, 1 preferred not to say) and Spring (N = 256; age: 18-37, *M* = 20.71, *SD* = 2.56; gender identity: 197 female, 43 male, 6 nonbinary, 9 other, 1 preferred not to say) Quarters of 2023. There were 149 transfer students enrolled in the Winter section and 119 transfer students in the Spring section. See Table 4 for a breakdown of student racial demographics. Each section had a different instructor of record, but the assignments, class content, format, and policies were the same in both sections.

	Winter	Spring
Asian	99	90
White/Caucasian	74	81
Latinx/Hispanic	31	30
Multiracial	26	26
Black/African American	8	12
Middle Eastern/Arab	9	10

Native American/Indigenous	2	0
Other	0	2
Prefer not to say	5	5

Table 4. *Student self-reported racial demographics in Experiment 5.*

Materials

Word lists. Stimuli used in the experiment consisted of 80 English nouns adapted from Experiment 2 were submitted to the English Lexicon Project (ELP; Balota et al., 2007) database to generate measures of length (*M* = 5.46 letters per word, *SD* = 1.75), frequency in the HAL (Lund & Burgess, 1996; *M* = 7.71 occurrences per million, *SD* = 1.46), and concreteness (*M* = 7.70, *SD* = 0.30). There were four themes, each with two categories: vehicles (land, air), fashion (clothing, shoes), plants (flowers, trees), and animals (fish, birds). As such, there were eight total word lists (two lists per theme) with two lists being used in each block (one for study and one for transfer). Further, each list had five words from each category (e.g., 5 flowers, 5 trees). Appendix C contains a full list of words used in this experiment.

End of Experiment Survey. We used the same end of survey measures included in Experiment 4. In addition, we asked participants to complete the Adult ADHD Self-Report Scale (ASRS-v1.1; Kessler et al., 2005). This scale consists of 18 items where participants rate how they have felt over the past 6 months. Each item relates to a symptom of ADHD and can be categorized into “inattentiveness”, “motor hyperactivity/impulsivity”, and “verbal hyperactivity/impulsivity”. For each item, participants rate how often they have experienced the symptom in the past 6 months with the options of “never”, “rarely”, “sometimes”, “often”, and “very often”. Items are displayed in Table 5. Items in Part A are used to diagnose ADHD, while items in Part B are used to further probe the patient’s symptoms. In our study, we did not use these items to label students as having ADHD, but instead were interested in how these symptoms relate to academic performance, selectivity in study habits, and strategic control of memory. Items 1-4 and 7-12 relate to inattention whereas items 5-6 and 13-18 relate to hyperactivity/impulsivity.

Adult ADHD Self-Report Scale

PART A

1. How often do you have trouble wrapping up the final details of a project, once the challenging parts have been done?
2. How often do you have difficulty getting things in order when you have to do a task that requires organization?
3. How often do you have problems remembering appointments or obligations?
4. When you have a task that requires a lot of thought, how often do you avoid or delay getting started?
5. How often do you fidget or squirm with your hands or feet when you have to sit down for a long time?
6. How often do you feel overly active and compelled to do things, like you were driven by a motor?

PART B

7. How often do you make careless mistakes when you have to work on a boring or difficult project?
8. How often do you have difficulty keeping your attention when you are doing boring or repetitive work?
9. How often do you have difficulty concentrating on what people say to you, even when they are speaking to you directly?
10. How often do you misplace or have difficulty finding things at home or at work?
11. How often are you distracted by activity or noise around you?
12. How often do you leave your seat in meetings or in other situations in which you are expected to stay seated?
13. How often do you feel restless or fidgety?
14. How often do you have difficulty unwinding and relaxing when you have time to yourself?
15. How often do you find yourself talking too much when you are in social situations?
16. When you’re in a conversation, how often do you find yourself finishing the sentences of the people you are talking to, before they can finish it themselves?
17. How often do you have difficulty waiting your turn in situations when turn taking is required?
18. How often do you interrupt others when they are busy?

Table 5. *Items from the Adult ADHD Self-Report Scale used in Experiment 5.*

Procedure

Participants logged into their learning management system during Week 6 of the between Monday at 10:00 a.m. and Sunday at 11:59 p.m. Once logged in, participants clicked a link to the experiment and were randomly assigned to complete one of two study schedules: they either restudied the first two-word lists and completed a free recall test on lists 3 and 4 or began with free recall and finished with restudying (see Figure 12 for a schematic of the value-directed learning task in Experiment 5). Participants were given instructions to study the words on the screen, that the words would be paired with number values, and were told to prioritize high-value words over low-value words. Each word belonged to a category, but participants were not made explicitly aware of this like in Experiment 1. Categories were adapted from Experiment 2 and consisted of plants (trees, flowers), fashion items (clothing, shoes), animals (birds, fish), and vehicles (land, air). Each category was paired with either a high (10 points)- or low (1 point)-value. In the restudy condition, participants restudied the same words in a different random order before the final test. In the recall condition, participants participated in a free recall test before the final test. The final test was the same transfer task as in Experiments 1 and 2 except participants only had two values to choose from instead of three. Participants made a confidence judgment (Dunlosky & Metcalfe, 2009) after each value prediction on a scale from 0 ("not at all confident") to 100 ("very confident"). After completing four study-test trials, participants completed the post-experiment questionnaire.

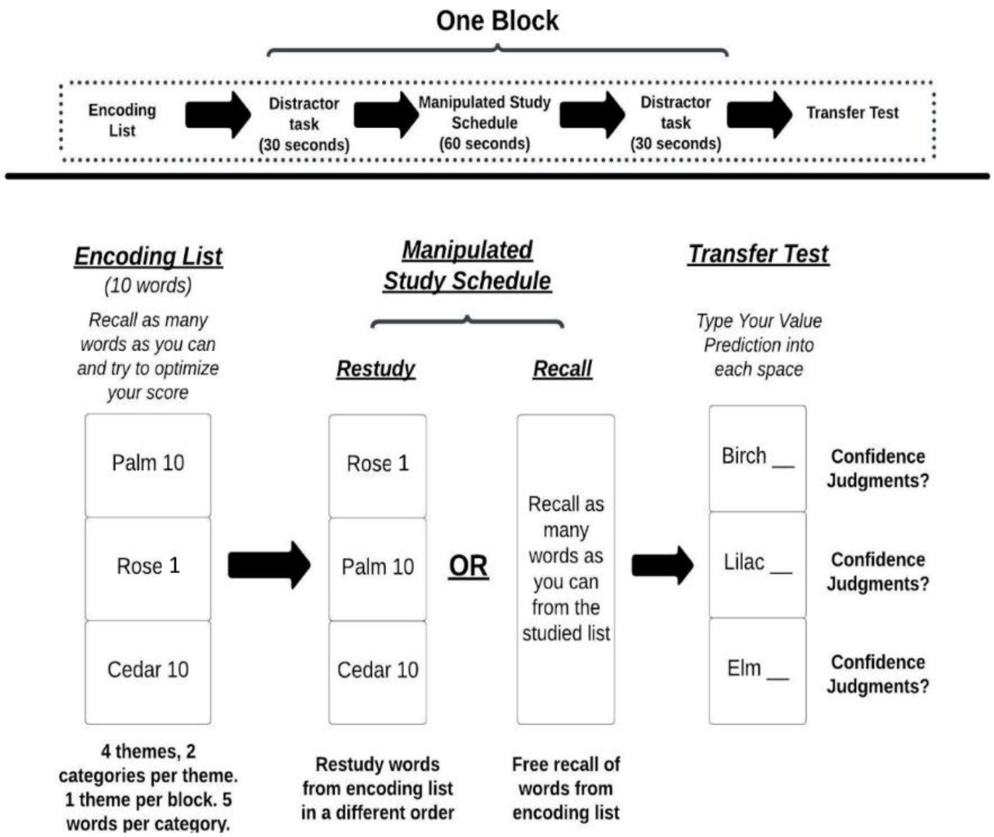


Figure 12. *Procedure for the value-directed learning task in Experiment 5.*

Results

Selectivity on the VDR task. A linear mixed model analysis was conducted to investigate the relationship between study selectivity scores and selective memory scores as measured by the selectivity index. List was also entered into the model to test whether participants became more selective as they progressed through the experiment. The quarter each student was enrolled in the course was entered into the model to control for any potential differences between groups. Students' ratings of interest in the course as we observed an effect of interest on selectivity in Experiment 4 and we also included scores on the ADHD ASRS to assess whether symptoms on inattention or hyperactivity influenced selectivity scores. The model demonstrated convergence with an REML criterion of 9183.1. Our key variable of interest, study selectivity as measured by scores on the Academic Strategies Questionnaire, was not a significant predictor of selectivity index scores [*b* = 0.001, *SE* = 0.002, *t*(498.60) = 0.63, *p* = .530], such that memory selectivity on the VDR task was not predicted by self-reported study selectivity. Participants did not become more selective with practice, but instead were less selective on list 1 compared to list 2 [*b* = -0.03, *SE* = 0.01, *t*(968.60) = -4.40, *p* < .001]. Interest in the course was not a significant predictor of memory selectivity on the VDR task scores [*b* = 0.01, *SE* = 0.01, *t*(498.10) = 0.94, *p* = .347] and there were no differences in selectivity scores between classes [*b* = 0.04, *SE* = 0.04, *t*(497.70) = 0.95, *p* = .341].The adjusted ICC for the random intercept model was 0.59 indicating that approximately 59% of the total variance in selectivity can be attributed to between-participant differences.

Transfer Performance on the VDL task. We conducted a generalized linear mixed effects model to examine the relationship between transfer performance on the value prediction task and study selectivity, symptoms of ADHD, class, list, interest, encoding condition (restudy or recall), and the interaction between list and encoding condition, while accounting for the nested structure of the data with items clustered within each participant such that there was a random intercept for each individual included in the model. The model was fitted using maximum likelihood estimation with Laplace Approximation. The model demonstrated a good fit to the data with an AIC of 18385.6 and a BIC of 18456.8. The fixed effects of study selectivity [e^B = 1.02, $CI_{95\%}$ = 1.00-1.03, *z* = 2.06 *p* = .039] was a significant predictor of transfer performance such that participants who reported being more selective when studying as measured by the ASQ were more accurate in predicting the word values on the transfer task (see Figure 13). List was also a significant predictor of transfer performance [e^B = 1.64, $CI_{95\%}$ = 1.53-1.77, *z* = 13.28, *p* < .001], suggesting transfer performance improved from list 1 to list 4. No other predictors were significant (all *p*s < .05), therefore, ADHD symptoms, class, interest, and encoding condition were not significant predictor of transfer performance and the effect of list did not vary by encoding condition. The random effects analysis revealed significant variability in intercepts across individual participants with a variance estimate of 1.87 and a standard deviation of 1.37. The adjusted ICC was 0.36, which indicates that approximately 36% of the total variance in transfer performance can be attributed to differences between participants.

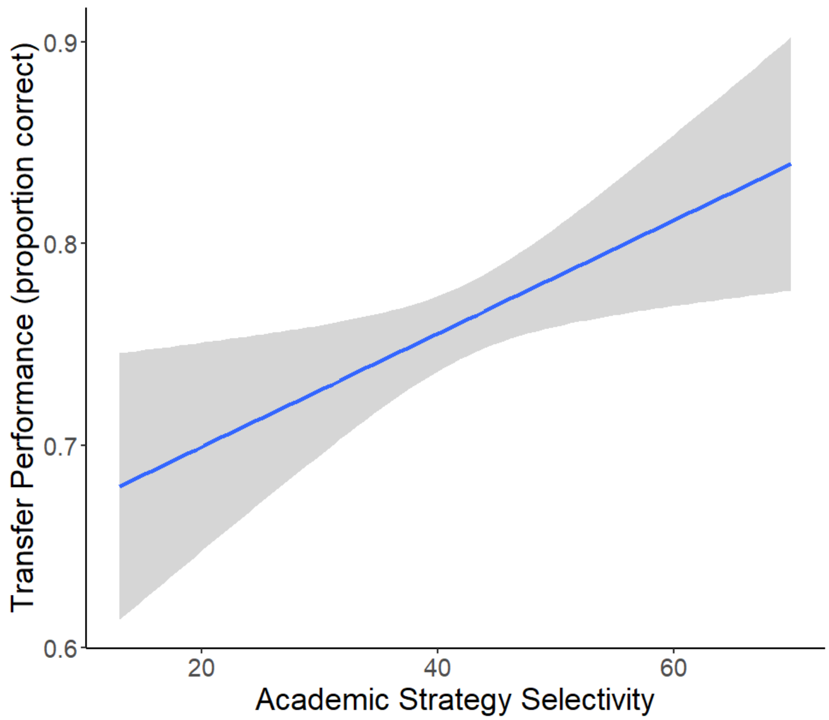


Figure 13. *Average transfer scores as a function of academic strategy selectivity as measured by the Academic Strategies Questionnaire in Experiment 5. Error bars represent the standard error of the mean.*

Exam Performance. We fit a MLR to model exam scores with study selectivity, motivation, trait anxiety, and ADHD symptoms as predictors and quarter students were enrolled as a covariate. The model's explanatory power (R^2) was .01. The model's intercept was at 85.25, $t(20360) = 136.69$, $p < .001$. The effect of study selectivity was a significant predictor of exam performance [$b = .06$, $t(20360) = 5.56$, $p < .001$], suggesting that more selective studiers performed better on the exams. The effects of motivation [$b = -0.15$, $t(20360) = -2.91$, $p = .004$], trait anxiety [*b* = -0.44, *t*(20360) = -7.53, *p* < .001], and ADHD symptoms [*b* = -0.16, *t*(20360) = -6.72, *p* < .001], were significant predictors of exam performance, suggesting students who reported being more motivated to do coursework, being more anxious, and having more symptoms of inattention and hyperactivity performed worse on the exams (see Figure 14). The quarter the students were enrolled was not a significant predictor of exam performance [*b* = -0.30, *t*(20360) = -1.65, *p* = .100].

Because ADHD was a significant negative predictor of exam performance and ASQ was a significant positive predictor of exam performance, we conducted a Pearson correlation between ASQ scores and ADHD symptoms which revealed a significant negative relationship, [*r* = -.09, *t*(508) = -2.06, *p* = .040] (see Figure 14).

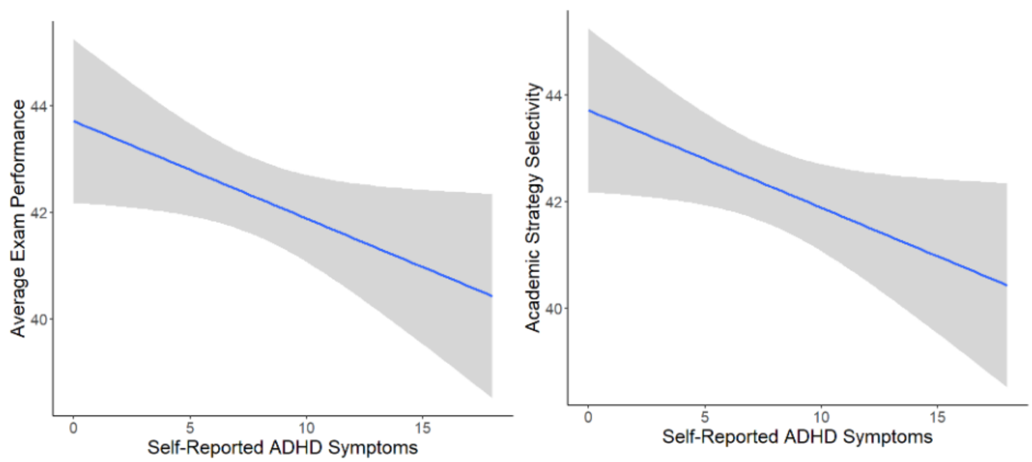


Figure 14. *Average exam performance as a function of self-reported ADHD symptoms in Experiment 5 (left). Academic Strategy Selectivity as a function of self-reported ADHD symptoms in Experiment 5 (right). Error bars represent the standard error of the mean.*

We also explored whether there was a relationship between scores on the VDL task and exam performance to better understand how the ability to use value cues to guide prediction of item importance related to real world academic outcomes in this sample. We conducted a Pearson correlation which revealed a significant positive relationship between VDL scores and average exam scores, [*r* = .21, *t*(508) = 4.81, *p* < .001].

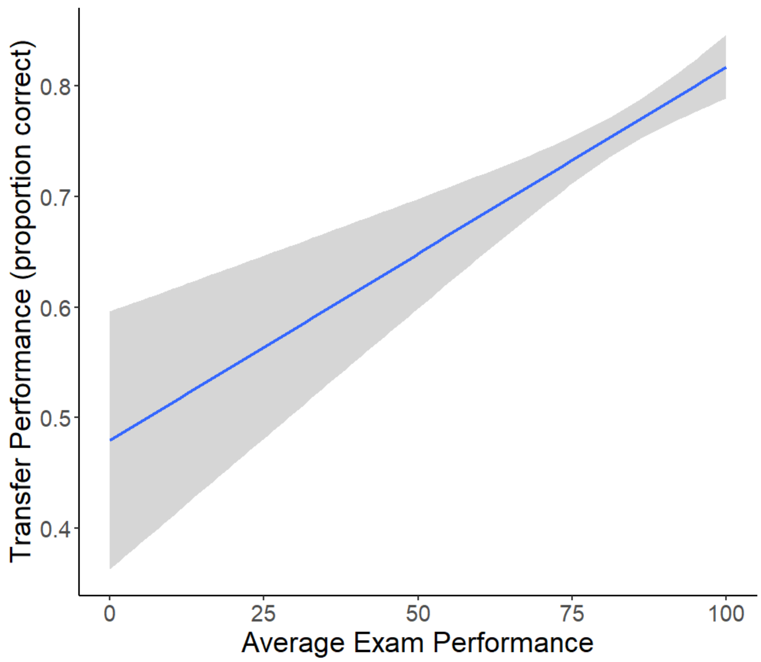


Figure 15. *Average transfer performance as a function of average exam performance in Experiment 5. Error bars represent the standard error of the mean.*

Experiment 5 Discussion

In Experiment 5, we further explored the relationship between study selectivity and classroom performance and how study selectivity relates to memory selectivity and the ability to predict the value of important information on our VDL task. We found that more selective studiers were not more selective on the VDR task. However, this was not a typical VDR task as there were only two value categories, students only had two lists to demonstrate selectivity in recall, and the items were semantically related. Future work should explore how memory selectivity relates to study selectivity in a standard VDR task with a larger sample size than we had in Experiment 4. We did find a relationship between transfer performance on our VDL task and study selectivity such that selective studiers were better at accurately pairing items with their associated values. Similarly, better performance on the VDL task was related to higher average exam scores. Therefore, the ability to use value cues to guide prediction of item performance on our VDL task seems to be positively related to real world classroom outcomes like study habits and exam performance.

We also found that exam performance was related to study selectivity such that more selective studiers performed better on the exams. Because students in this course took multiple-choice exams, future research should explore this relationship in other exam formats. It could be that the relationship observed here was related to skills required on multiple-choice tests or the content tested in Research Methods in Psychology. Furthermore, students with more symptoms of ADHD performed worse on exams and reported less selective study habits but performed just as well as other students on the VDL task. While this study is the first to look at how symptoms of ADHD in college students related to performance on a VDL task and selective study habits, these findings are supported by prior work showing that college students with symptoms of ADHD report using less effective study strategies and perform worse on their course exams. These findings inform efforts towards more inclusive teaching practices for neurodivergent students.

Chapter 3 Conclusions

Some prior work has begun to investigate the relationship between memory selectivity and selective study habits and how choosing more selective study habits relates to classroom performance (Patterson et al., unpublished manuscript). Further, the ability to prioritize high-value information in a lab setting is well-documented (Castel et al., 2002; see Knowlton & Castel, 2002 for a review) and participants report choosing more effective encoding strategies for high-value words in these value-directed remembering experiments (Hennessee et al., 2019). The ability to strategically encode and retrieve important information is likely an adaptive mechanism that relates to real-world outcomes such as preparing for an exam. The studies in this chapter contribute to our understanding of how strategic allocation of time, attention, and strategy use relate to classroom outcomes.

We observed some important differences between the relationships between interest in the course and memory selectivity between Experiments 4 and 5. There was a positive relationship between these variables in Experiment 4, but not a significant relationship in Experiment 5. Importantly, students in Experiment 4 participated for extra credit and may have been more interested in the course content than students in Experiment 5 who participated as a part of an active learning module. Further, the VDR task in Experiment 5 was not a standard VDR task, so it could be that students were not motivated to be selective in this task as they could have paid more attention to connections between the items as opposed to the items' values.

We also found that study selectivity was related to exam performance in Experiment 5, but not in Experiment 4. Students tested in these experiments were enrolled in completely different classes with different test formats and these details could impact the relationships we observed here. Students in both courses seemed to benefit from intrinsic motivation to remember (Experiment 4) and learn (Experiment 5). In Chapter 4 we explore the role of familiarity, age, and curiosity on memory for medical and trivia information.

Finally, we observed informative relationships between symptoms of ADHD and academic outcomes such as self-reported study habits and average exam scores. Specifically, reporting more ADHD symptoms predicted worse exam performance and was related to less selective study habits. Given that college classrooms often have students with ADHD who are undiagnosed or without formal accommodations, it is important to understand how students with ADHD symptoms perform on different types of learning activities and assessments. Here, we observed that students with more symptoms of ADHD performed similarly as other students on our lab-based learning task which was engaging and low-stakes. In contrast, these same students performed worse on high-stakes multiple-choice exams. Because less selective study habits were related to lower exam scores and higher ADHD symptomology, future work should explore whether study habits potentially mediate the relationship between ADHD symptoms and exam performance. This work contributes to our understanding of how individual differences in attention and impulsivity relate to classroom outcomes.

CHAPTER 4: INTRINSIC MOTIVATION IN LIFELONG LEARNING

Though we often think of value in terms of money, grades, or in other extrinsic ways, value can also be intrinsic. For example, people value their time, relationships, and learning about topics they are interested in or skills that may be useful to them. When thinking about older adult learners, these considerations may be especially important. Learning in complex domains often involves remembering the names of theories, formulas, and other concepts and their associated information. In many fields naming theories and other concepts after a person is commonplace, though research suggests that proper names are quite challenging to remember regardless of age (Cohen & Faulkner, 1986) and expertise (Waseem et al., 2005), though older adults may have a more difficult time recalling newly learned names compared to younger adults (Rendell et al., 2005). Using proper names with no real semantic connections to the theory or concept they are associated with combined with inconsistent use of these types of names in the literature and in practice, could put learners at a disadvantage and may be especially harmful to older adult learners. However, an individual's familiarity with these names or prior knowledge of materials could influence learning outcomes as well as curiosity or interest in learning new information.

Different types of curiosity have been examined in the literature including physical, intellectual, epistemic, state, trait, depth, breadth, and others, so it is important to distinguish between them (Grossnickle, 2016). Curiosity is different than interest, but often the two concepts are often used synonymously. Key characteristics of curiosity that may help in differentiating the construct from interest include the themes of the role of knowledge, goals and outcomes, and stability and malleability. McGillivray et al. (2015) investigated curiosity in older and younger and older adults who read trivia questions and rated how curious they were to learn the answer to each one. They also provided confidence and interest ratings. Then, they were presented with the answer to the questions and afterwards provided a judgment of learning (JOL). Later, they took a memory test on the answers to the questions, and there were no age-related differences in recall, and curiosity and interest ratings were related to JOLs. The participants were followed up with after one week and took another memory test on the items. At the follow-up test, there was a positive relationship between interest ratings and memory for older adults, suggesting that a person's subjective interest may enhance memory in older adulthood. In a similar study, younger and older adults read trivia questions and rated their curiosity to find out the answer for each one (Galli et al., 2018). Participants viewed faces between the question and answer, and then took a surprise memory test (free recall for answers to the trivia questions and a recognition test for the faces). Recall performance was related to higher ratings of curiosity, and recognition memory was better for faces that were presented near trivia questions that participants were more curious about. Therefore, curiosity may impact memory for unrelated information presented in proximity to high-curiosity items.

In Experiment 6, we investigated the impact of familiarity and the use of proper names on associative memory in older and younger adults. Here, we were interested in how using proper names that are less familiar (e.g., likelihood of prior knowledge of the material is low) may be a less inclusive way to present new information, especially for older adults, compared to using more descriptive names where there are semantic relationships between the name and associated information. We were also interested in how metacognitive monitoring relates to the types of materials used and how this may vary based on age. In Experiment 7, we investigated the impact of curiosity on learning and explored how learners' interactions with others during learning impact their memory for and curiosity about the material.

Experiment 6a

Aspirin, Broca’s Area, Down Syndrome and Alzheimer's disease are all examples of eponyms: discoveries, inventions, or any other object that is named after a person (Bartolucci et al., 2005). The medical field now has over 18,000 eponyms, each one denoting a specific medical condition, surgical technique, drug, or other phenomenon. While healthcare professionals are typically taught using eponyms, familiarity with these types of terms is not guaranteed and may present a challenge for the general population. Historically, proper names have been difficult to remember, suggesting that the use of eponyms, whether it be in the classroom or the hospital, can make remembering and deciphering them rather difficult. While eponyms can be useful and concise, they can still make learning and associating complex information challenging. Given these issues our study aims to explore memory for information associated with eponyms as compared to descriptive names, and how age and familiarity influences the impact of item presentation on memory.

Eponyms honor the hard work of the scientists who devoted their time and effort to making groundbreaking discoveries or innovations and can offer the advantage of simplifying complex scientific terminologies or procedures such as Trisomy 21, a genetic disorder caused by the presence of all or part of a third copy of chromosome 21, into concise terms like “Down Syndrome” (Whitworth, 2007). However, when work goes unrecognized or is plagiarized, glorifying a name can be ethically detrimental as opposed to beneficial. Some of the most researched cases of unethical eponyms are ones named after Nazi scientists. These scientists experimented on prisoners without their consent to discover various medical conditions which they promptly named after themselves (Strous & Edelman, 2007). Though in some cases eponyms are titled after those that did not deserve the accolades they received (see “Stigler’s

Law of Eponymy", Stigler, 1980), there have also been cases where people who did deserve this honor were deprived of their chance to have a discovery named after them. The infamous Watson and Crick model of DNA was named after James Watson and Francis Crick, and then it was later discovered that Rosalind Franklin had played a significant role in the discovery but was deprived of the same honor her peers received (Stasiak, 2001). While names like these can prove to be surprisingly controversial, the possible issues regarding eponyms extend beyond ethical concerns and into cognitive ones.

For many of us, young and old, it is a common occurrence to forget someone's name within seconds of being introduced to them. As such, one of the primary issues with the use of eponyms comes from the fact that proper names have been found to be one of the most difficult informational categories to remember (Cohen et al., 1986; Rendell et al., 2005) and this remains true irrespective of other factors like age (Cohen et al., 1986). In addition, a study showed that information associated with names can also fall victim to this deficit, with people making more errors in remembering information when it is associated with proper names as opposed to other categories like occupation with an exaggerated deficit in older adults (James, 2004).

While it may be difficult for the general populace to remember proper names used in eponyms, particularly medical ones, one would expect medical professionals to be able to do so. A study examining if surgeons and specialists in orthopedic surgery were able to correctly identify procedures named as eponyms showed that only about 10% were able to do so (Waseem, et. al., 2005). Eponym use is difficult when it comes to practical application even for professionals in a field. This deficit in remembering eponyms might lead to difficulties with information associated with these terms.

The alternative to the non-descriptive proper names used in eponyms would be descriptive names. Such an alternative could particularly assist older adults who have a deficit in retrieving eponyms (Fogler & James, 2007). Names like “irritable bowel syndrome” instead of Crohn’s disease are more salient and could enhance communication and understanding for patients as well as professionals who also find eponyms hard to use when communicating with others (Waseem, et al., 2005). While healthy aging is associated with impaired working and episodic memory (Cabeza et al., 2004; Salthouse & Babcock, 1991), familiarity, or prior experience or exposure to an item, can help older adults compensate for these deficits (Rendell et al., 2005). Though familiarity could lead to overconfidence in memory performance in some cases (Davidson & Glisky, 2002; Yonelinas & Levy, 2002). Therefore, we were interested in exploring potential benefits of using descriptive names when learning medical information compared to eponyms and how metacognitive predictions of performance may be influenced by familiarity with the studied items.

We expected that older adults would be more familiar with medical conditions, including the use of medical eponyms, due to their personal experiences with such conditions and prior research in this area (Brown & Park 2002). On the other hand, they may show some deficits in associative memory, especially when it comes to information paired with proper names. In Experiment 6, we investigated the impact of familiarity and the use of proper names on associative memory in older and younger adults. Here, we were interested in how using proper names that are less familiar (e.g., likelihood of prior knowledge of the material is low) may be a less inclusive way to present new information, especially for older adults, compared to using more descriptive names where there are semantic relationships between the name and associated information. We were also interested in how metacognitive monitoring relates to the types of materials used and how this may vary based on age. We tested older and younger adults on information paired with the names of medical conditions either presented as eponyms or as descriptive names. We also manipulated the familiarity of each item with some items being well-known (e.g., “Alzheimer’s Disease”) and others being less commonly known (e.g., “Alibert’s Disease”).

Method

Participants

Participants were 39 undergraduate University of California, Los Angeles (UCLA) students recruited from the UCLA Psychology Subjects Pool. The mean age of the participants was 20.59 years (SD = 1.91). Of the total participants, 9 individuals self-identified as male, while 34 participants identified as female. The racial identities of the participants were as follows: 14 Asian, 12 Caucasian, 5 Arab, 4 Other, 3 Multiracial, and 1 preferred not to answer. All participants reported that they were fluent in the English language, with an average age of starting to speak English being 2.00 years old (*SD*=3.74). An additional 21 participants were excluded for admitting to engaging in other activities during the experiment on a post-experiment questionnaire.

Materials

Participants studied 16 medical conditions each paired with a cause, symptom, and treatment. For each participant, half of the conditions were presented as eponyms while the other half were presented as descriptive names. Further, half of the conditions chosen were well-known conditions while the other half were less familiar. We had a separate participant group norm these items by rating their familiarity with each one. Further, we collected individual measures of familiarity with each condition from the participants in the current study. Examples of the materials can be found in Figure 13 (see Appendix D for a full list of materials).

a) **ALZHEIMER'S DISEASE**

Cause	Symptom	Treatment
Brain Changes	Memory Impairment	Behavioral Intervention

b) **PRESENILE DEMENTIA**

Cause	Symptom	Treatment
Brain Changes	Memory Impairment	Behavioral Intervention

c) **CHIARI'S SYNDROME**

Cause	Symptom	Treatment
Small Skull	Neck Pain	Pain Medication

d) **CEREBELLAR MALFORMATION**

Cause	Symptom	Treatment
Small Skull	Neck Pain	Pain Medication

Figure 16. *Examples of materials used in Experiment 6 for each of four conditions: a) familiar eponyms, b) familiar descriptive names, c) unfamiliar descriptive eponyms, and d) unfamiliar descriptive names.*

Procedure

Participants participated remotely on their own devices. After clicking the experiment link, they were randomly assigned to a counterbalanced order of the four conditions. While familiarity of the conditions was manipulated between-items, each condition was randomly assigned to be presented as either an eponym or descriptive name for each participant to control for potential item effects. After being presented with instructions, participants began the first study trial. In each trial, the name of each condition appeared at the top of the screen in all capital letters paired with a table of information below it containing a cause, symptom, and treatment of the condition (see Figure 16). Participants had 30 seconds to study each condition. Then, they automatically advanced to the next page where they were asked to make a JOL for each fact associated with the just studied medical condition: "How likely are you to remember the cause of this medical condition from 0 (not at all likely) to 100 (I will definitely remember this information)?" Then, participants would repeat this procedure for the three other conditions in the trial block.

After studying and making JOLs for each medical condition, participants received test instructions. They were told they would be tested on the medical conditions they had just studied and would have 60 seconds to choose the best answer. For each medical condition, three questions appeared on the same page: one testing memory for the cause, symptom, and treatment that was studied. Participants had a total of 60 seconds to answer all three questions before they were automatically advanced to the next page. Then, they were prompted to make a confident judgment for each of their three responses: “How confident are you that the cause you selected is correct from 0 (not at all confident) to 100 (I definitely got this question right)?” After completing the study and test phase in one block (e.g., unfamiliar eponyms), they proceeded to follow the same procedure in a different experimental block (e.g., familiar descriptive names). See Figure 17 for a schematic of the procedure for Experiment 6.

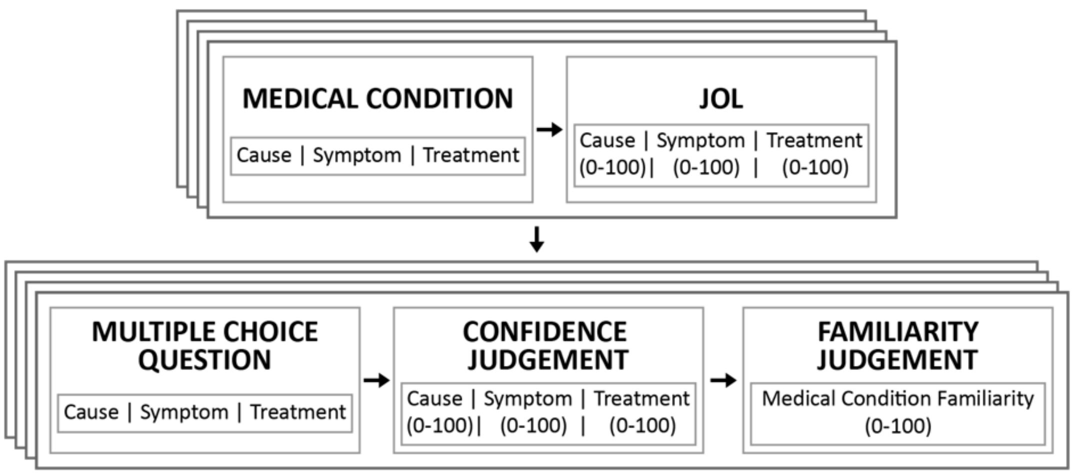


Figure 17. *Procedure for Experiment 6.*

After completing all four blocks, participants completed a post-experiment questionnaire. Participants were debriefed and told that they studied eponyms like “Guillain Barre Syndrome” which is a disorder named after a person and descriptive names where the disorder was presented using more descriptive language like “Acute Paralysis Disorder”. They were asked which of these two types of presentations felt easier to remember and could also indicate that they felt both were equally difficult. We also asked whether there was a particular category of information that felt most important to remember (i.e., causes, symptoms, or treatments) and participants could also indicate that they felt they were equally important. We asked participants what strategies they used during the task and whether their strategies were different based on whether they were studying causes, symptoms, or treatments. Finally, participants disclosed whether they cheated or were doing anything else while the experiment was going on for exclusion purposes.

Results

To examine differences in JOLs, accuracy on the multiple-choice test, and confidence judgments between trials presented as eponyms and descriptive names and between familiar and unfamiliar items, we conducted MLMs. Because accuracy for each item was binary (i.e., correct or incorrect), we conducted a logistic MLM for this analysis and report log ratios of correct recall along with 95% confidence intervals for each predictor variable.

Metacognitive Predictions. A linear mixed model analysis was conducted to investigate the effects of presentation format (eponym vs. descriptive name) and familiarity category (unfamiliar vs. familiar) on judgments of learning (JOLs), while accounting for the hierarchical structure of the data with items nested within participants (IDs). The model demonstrated convergence with a REML criterion of 17611.1. The analysis revealed significant main effects for familiarity category, b = 18.21, SE = 1.31, $t(1925.98) = 13.91$, $p < .001$, and presentation format, b = 10.50, SE =1.31, t =10.50, $t(1925.98) = 8.02$, $p < .001$on JOLs. Participants provided higher JOLs for familiar items compared to unfamiliar items, and JOLs were higher for items presented as descriptive names compared to eponyms. Furthermore, there was a significant interaction effect between presentation format and familiarity category, $b = -15.76$, SE =1.85, $t(1925.98) = -8.52$, $p < .001$ (see Figure 18). The adjusted ICC for the random intercept model was 0.46 indicating that approximately 46% of the total variance in JOLs can be attributed to between-participant differences.

A follow-up test of simple slopes was conducted to probe the interaction effect between presentation format (descriptive vs. eponym) and familiarity category (unfamiliar vs. familiar) on judgments of learning (JOLs). For both unfamiliar and familiar items, the slope representing the relationship between item presentation and JOLs was significant. Specifically, items presented as descriptive names were associated with an increase in JOLs compared to items presented as eponyms for unfamiliar items, b = 10.50, SE = 1.31, t(1926) = 8.02, p < .001. However, descriptive names were associated with lower JOLs compared to eponyms for familiar items, b = -5.26, SE = 1.31, t(1926) = -4.02, p < .001.

We conducted a linear mixed model analysis to investigate the effects of presentation format (eponym vs. descriptive name) and familiarity category (unfamiliar vs. familiar) on confidence judgments while accounting for the hierarchical structure of the data with items nested within participants (IDs). The model demonstrated convergence with a REML criterion of 18329.5. The analysis revealed significant main effects for familiarity category, b = 22.03, SE = 1.59, t(1926.07) = 13.88, p < .001, and presentation format, b = 17.00, t(1926.07) =10.72, SE =1.59, p < .001on JOLs. Participants provided higher confidence judgments for familiar items compared to unfamiliar items, and for items presented with descriptive names compared to eponyms. Furthermore, there was a significant interaction effect between presentation format and familiarity category, b = -17.24, SE = 2.24, t(1926.07) = -7.68, p < .001 (see Figure 18). The adjusted ICC for the random intercept model was 0.23 indicating that approximately 23% of the total variance in confidence judgments can be attributed to between-participant differences.

A follow-up test of simple slopes was conducted to probe the interaction effect between presentation format (descriptive vs. eponym) and familiarity category (unfamiliar vs. familiar) on confidence judgments. For unfamiliar items, the slope representing the relationship between item presentation and confidence judgments was significant. Specifically, for unfamiliar items, participants showed a significant increase in confidence judgments for items presented as descriptive names compared to eponyms, b =17.00, SE = 1.59, t(1926) = 10.72, p < .001. However, for familiar items, the slope representing the relationship between item presentation and confidence judgments was not significant, b = -0.24, SE = 1.59, t(1926) = -0.15, p = .881.

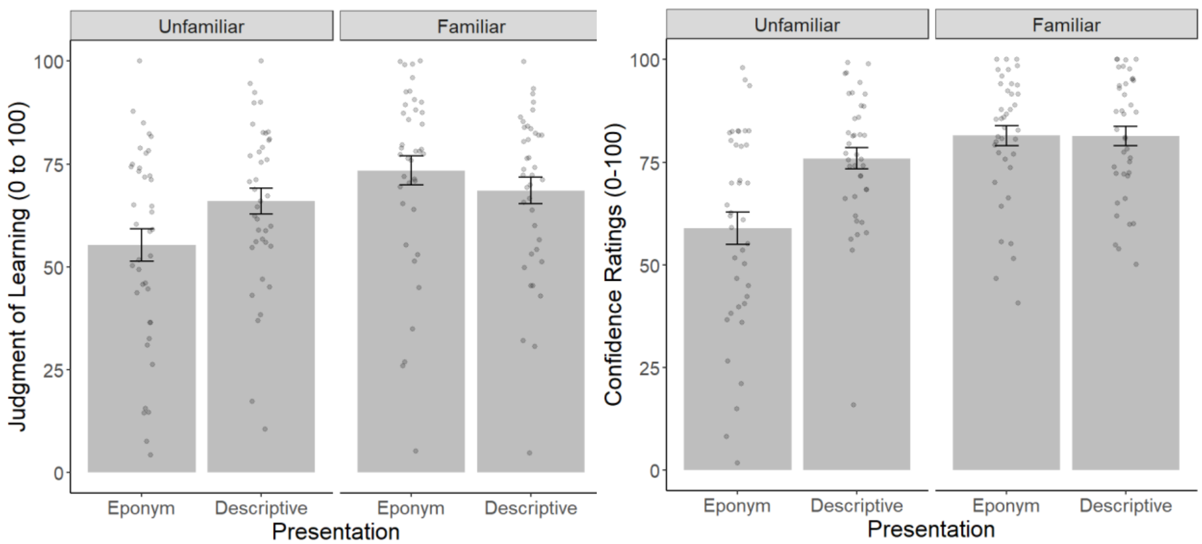


Figure 18. *Judgments of learning (left) and confidence judgments (right) as a function of item presentation and familiarity. Error bars represent the standard error of the mean.*

Test Performance. To examine the effects of presentation format (eponym vs. descriptive name) and familiarity category (unfamiliar vs. familiar) on accuracy on the multiple-choice test performance while accounting for the hierarchical structure of the data with items nested within participants, we conducted a generalized linear mixed effects model. The model was fitted using maximum likelihood estimation with Laplace Approximation. The model demonstrated a good fit to the data with an Akaike Information Criterion (AIC) of 1676.20 and a Bayesian Information Criterion (BIC) of 1704.1. The fixed effects of item presentation [e^B = 4.62, $CI_{95\%}$ = 3.25-6.57, z = 8.53, p < .001], familiarity [e^B = 3.55, $CI_{95\%}$ = 2.41-4.67, z = 7.15, p < .001], and the interaction between item presentation and familiarity [e^B = .25, $CI_{95\%}$ = 0.15-0.41, z = -5.42, p < .001] were significant predictors of test performance such that information associated with familiar items was more frequently identified compared to information paired with unfamiliar items (see Figure 19). Similarly, information associated with items presented as descriptive names was more frequently identified compared to information paired with items presented as eponyms. The random effects analysis revealed significant variability in intercepts across individual participants with a variance estimate of 1.84 and a standard deviation of 1.36. The adjusted ICC was 0.36, which indicates that approximately 36% of the total variance in test performance can be attributed to differences between participants.

A follow-up test of simple slopes was conducted to probe the interaction effect between presentation format (descriptive vs. eponym) and familiarity category (unfamiliar vs. familiar) on test performance. For unfamiliar items, the slope representing the relationship between item presentation and test performance was significant. Specifically, for unfamiliar items, participants had higher accuracy on multiple-choice items testing memory for information paired with descriptive names compared to eponyms, b = 1.53, z = 8.53, p < .001. However, for familiar items, the slope was not significant, b = 0.13, z = 0.67, p = .502.

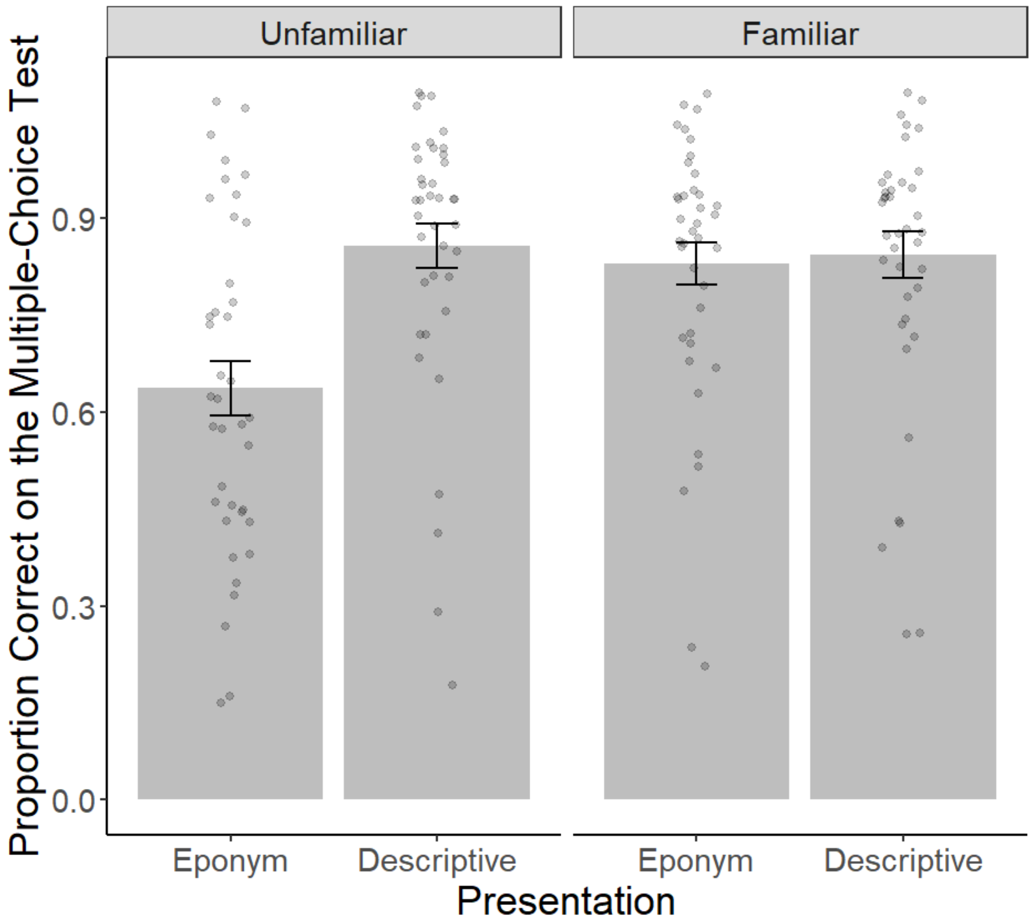


Figure 19. *Test performance as a function of item presentation and familiarity. Error bars represent the standard error of the mean.*

Experiment 6a Discussion

In Experiment 6a we find evidence that presentation format and familiarity impact metacognitive monitoring where more familiar and more descriptive items are rated as more likely to be remembered and this same trend was found for confidence in answers to the multiple-choice questions. However, familiar eponyms were rated as more likely to be remembered than familiar descriptive names while there were no differences in confidence judgments between familiar eponyms and descriptive names. This suggests that familiar descriptive names may impact metacognitive monitoring during encoding, but not after being tested on the information. This could be due to participants being metacognitively accurate in this case as performance on the multiple-choice test showed a similar trend as the confidence judgments. Specifically, familiar items and descriptive items were better remembered, and unfamiliar descriptive names were better remembered than unfamiliar eponyms. However, there were no differences between memory for information paired with familiar eponyms and familiar descriptive names. Next, we test these effects in a sample including older adults.

Experiment 6b

Experiment 6b used the same procedure and materials as Experiment 6a but had a larger sample size that included older adults. We expected older adults to benefit from descriptive names but be more familiar with these conditions overall, but to provide lower metacognitive judgments as they may be more aware of their memory limitations.

Method

Participants

Participants were 64 older adults (Age: M = 66.92, SD = 5.62) recruited from Cloud Research Prime Panels (Chandler et al., 2019) and 64 younger adults (Age: M = 20.20, SD = 1.67) recruited from the UCLA Psychology Human Subjects Pool. On average, participants learned English before the age of 2 years old (M = 1.20, SD = 3.36). Participants reported their gender and racial identities and educational background on a post-experiment questionnaire. There were 56 younger and 38 older adults who identified as female and 8 younger and 26 older adults who identified as male. For racial identities, 4 younger and 1 older adults identified as Arab or Middle Eastern, 20 younger and 2 older adults as Asian, 22 younger and 52 older as white or Caucasian, 7 older adults identified as Black or African American, 8 younger and 1 older as multi-racial, and 8 younger adults identified as other, and 2 younger and 1 older adults preferred not to report their race. Educational background for the sample was as follows: some high school (1 younger, 1 older), high school graduate or equivalent (10 younger, 13 older), some college no degree (38 younger, 13 older), associate's degree (14 younger, 10 older), bachelor's degree (1 younger, 14 older), and 13 older adults reported having graduate degrees. An additional 25 participants were tested but excluded from analyses based on their responses to a post-experiment survey: 1 older adult participant admitted to cheating (i.e., writing down answers) and 24 (18 younger and 6 older adults) admitted to engaging in other activities during the experiment.

Materials and Procedure

The materials and procedure used in Experiment 6b were identical to those used in Experiment 6a (see Figures 16 and 17).

Results

Analyses conducted in Experiment 6b were similar to Experiment 6a except that age was used as a categorical predictor in the model where older adults were coded as 1 and younger adults were coded as 0.

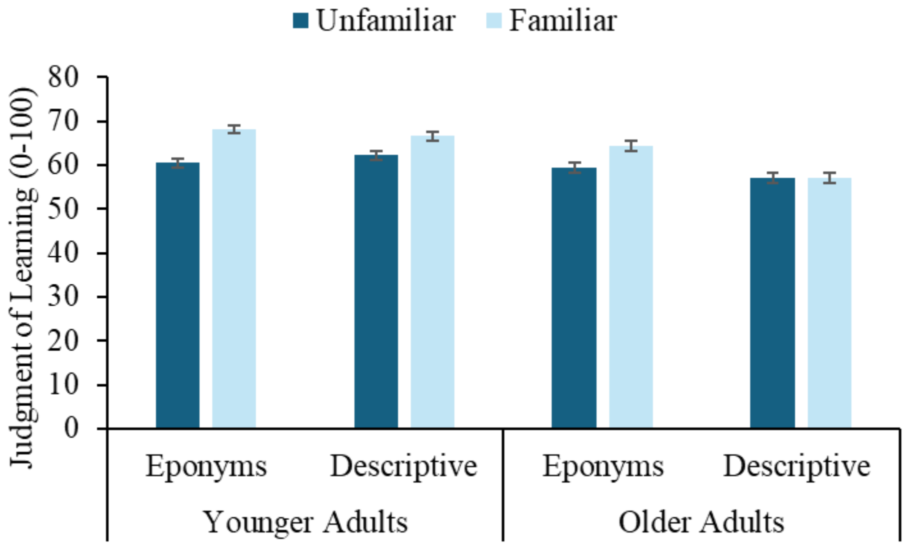
Metacognitive Predictions. A linear mixed model analysis was conducted to investigate the effects of presentation format (eponym vs. descriptive name) and familiarity category (unfamiliar vs. familiar) on judgments of learning (JOLs), while accounting for the hierarchical structure of the data with items nested within participants. The model demonstrated convergence with an REML criterion of 53886.4. The analysis revealed a significant main effect for familiarity category on JOLs, b = 7.64, SE = 0.95, t(6001.65) = 8.06, p < .001, such that participants provided higher JOLs for familiar items compared to unfamiliar items. Furthermore, there was a significant interaction effect between presentation format and familiarity category, b = -3.33, SE = 1.34, t(6001.65) = -2.48, p = .013, and between presentation format and age, b = -

4.04, *SE* = 1.34, *t*(6001.65) = -3.02, *p* = .003. The adjusted ICC for the random intercept model was 0.60 indicating that approximately 60% of the total variance in JOLs can be attributed to between-participant differences. However, no other predictors were significant (all *p*s > .05).

As shown in Figure 20, a follow-up test of simple slopes was conducted to probe the interaction effect between presentation format (descriptive vs. eponym) and familiarity category (unfamiliar vs. familiar) on JOLs. For familiar items, the slope representing the relationship between item presentation and JOLs was significant, *b* = -4.52, *SE* = 0.67, *t*(6002) = -6.74, *p* < .001, such that items presented as descriptive names were associated with lower JOLs compared to items presented as eponyms. However, the simple slope of item presentation for unfamiliar items was not significant, *b* = -0.27, *SE* = 0.67, *t*(6002) = -0.42, *p* = .682. Next, I conducted a follow-up test of simple slopes to investigate the interaction between presentation format and age, which revealed a significant slope of item presentation for older adults, *b* = -4.87, *SE* = 0.67, *t*(6002) = -7.27, *p* < .001, but not for younger adults, *b* = 0.08, *SE* = 0.67, *t*(6002) = 0.12, *p* = .901, such that older adults provided lower JOLs for descriptive names compared to eponyms whereas younger adults provided similar JOLs regardless of how the item was presented.

I conducted a linear mixed model analysis to investigate the effects of presentation format (eponym vs. descriptive name) and familiarity category (unfamiliar vs. familiar) on confidence judgments while accounting for the hierarchical structure of the data with items nested within participants. The model demonstrated convergence with a REML criterion of 56362.2. The analysis revealed significant main effects for familiarity category, *b* = 10.48, *SE* = 1.17, *t*(6001.96) = 8.99, *p* < .001, and presentation format, *b* = 6.87, *SE* = 1.17, *t*(6001.96) = 5.90, *p* < .001 on confidence judgments. Participants provided higher confidence judgments for familiar items compared to unfamiliar items, and for items presented with descriptive names compared to eponyms. Furthermore, there was a significant interaction effect between item presentation and familiarity, $b = -11.81$, $SE = 1.65$, $t(6001.96) = -7.16$ $p < .001$. There was also a significant main effect of age, $b = -14.84$, $SE = 4.03$, $t(157.96) = -3.69$, $p < .001$, such that older adults were significantly less confident in their responses to the multiple-choice questions compared to younger adults. There was a significant interaction between familiarity and age, $b = 5.66$, $SE = 1.65$, $t(6001.96) = 3.43$, $p = .001$. The adjusted ICC for the random intercept model was 0.49 indicating that approximately 49% of the total variance in JOLs can be attributed to between-participant differences.

As shown in Figure 20, a follow-up test of simple slopes was conducted to probe the interaction effect between presentation format (descriptive vs. eponym) and familiarity category (unfamiliar vs. familiar) on confidence judgments. For unfamiliar items, the slope representing the relationship between item presentation and confidence judgments was significant for both familiar and unfamiliar items. Specifically, for unfamiliar items, participants showed a significant increase in confidence judgments for items presented as descriptive names compared to eponyms, $b = 6.47$, $SE = 0.82$, $t(6002) = 7.84$, $p < .001$. However, for familiar items, confidence judgments were lower for descriptive names compared to eponyms, $b = -6.87$, $SE = 0.82$, $t(6002) = -8.33$, $p < .001$. Next, I probed the interaction between age and item familiarity using a test of simple slopes which revealed a significant effect of familiarity on confidence judgments for both older [$b = 8.71$, $SE = 0.82$, $t(6002) = 10.56$, $p < .001$] and younger adults [$b = 4.57$, $SE = 0.82$, $t(6002) = 5.55$, $p < .001$], such that participants were more confident in their answers for questions about familiar items compared to unfamiliar items.



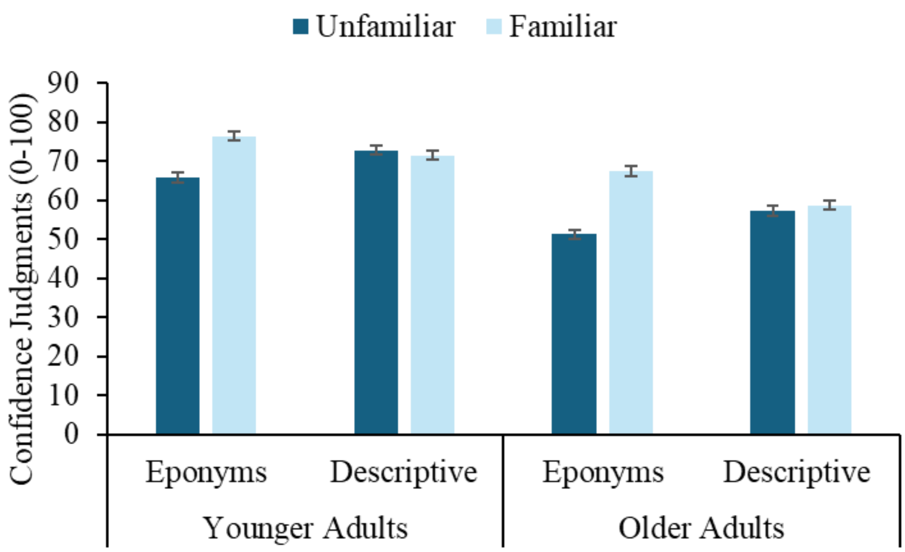


Figure 20. *Judgments of learning (top) and confidence judgments (bottom) as a function of item presentation, familiarity, and age. Error bars represent the standard error of the mean.*

Test Performance. To examine the effects of presentation format (eponym vs. descriptive name) and familiarity category (unfamiliar vs. familiar) on accuracy on the multiple-choice test while accounting for the hierarchical structure of the data with items nested within participants, we conducted a generalized linear mixed effects model. The model was fitted using maximum likelihood estimation with Laplace Approximation. The model demonstrated a good fit to the data with an Akaike Information Criterion (AIC) of 6129.8 and a Bayesian Information Criterion (BIC) of 6190.3. The fixed effects of item presentation [e^B = 31.13, $CI_{95\%}$ = 23.10-41.95, z = 22.59, p < .001], familiarity [e^B = 32.94, $CI_{95\%}$ = 23.47-44.50, z = 22.76, p < .001], and the three-way interaction between item presentation, familiarity, and age [e^B = 2.76, $CI_{95\%}$ = 1.62-4.72, z = 3.72, p < .001] were significant predictors of test performance such that information associated with familiar items was more frequently identified compared to information paired with unfamiliar items. Similarly, information associated with items presented as descriptive names was more frequently identified compared to information paired with items presented as eponyms. The random effects analysis revealed significant variability in intercepts across individual participants with a variance estimate of 1.19 and a standard deviation of 1.09. The adjusted ICC was 0.27, which indicates that approximately 27% of the total variance in test performance can be attributed to differences between participants.

As seen in Figure 21, a follow-up test of simple slopes was conducted to probe the three-way interaction effect between presentation format (descriptive vs. eponym), familiarity category (unfamiliar vs. familiar), and age (old vs. young) on test performance. Younger adults better recalled information paired with descriptive names for unfamiliar, b = 3.44, SE = 0.15, z = 22.58, p < .001, but not familiar, b = -0.15, SE = 0.15, z = -1.04, p = .30, items. On the other hand, older adults benefitted from studying descriptive names compared to eponyms for unfamiliar items, b

= 2.21, *SE* = 0.12, *z* = 17.52, *p* < .001. However, they performed better when studying eponyms compared to descriptive names for familiar items, *b* = -0.37, *SE* = 0.12, *z* = -3.19, *p* < .001.

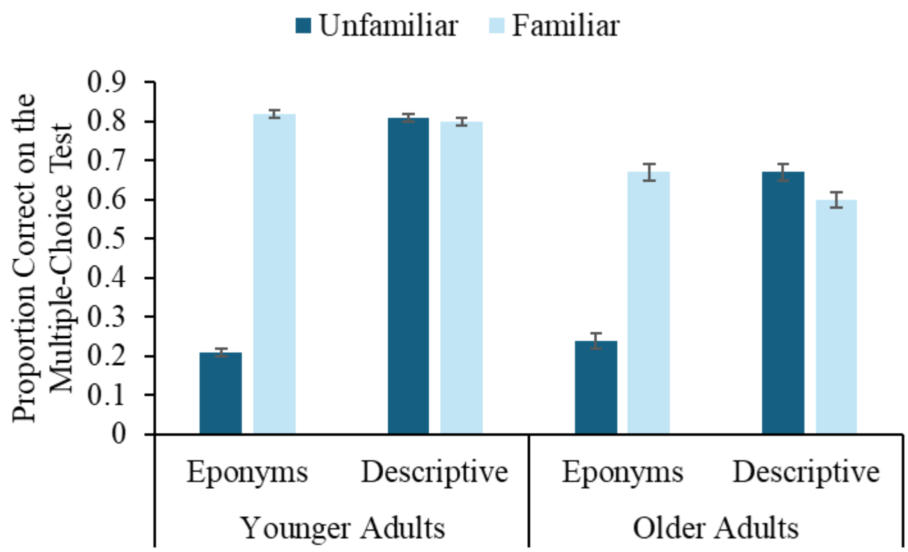


Figure 21. *Test performance as a function of item presentation, familiarity, and age. Error bars represent the standard error of the mean.*

Experiment 6b Discussion

Higher JOLs were provided for familiar items, but there were no differences in JOLs between eponyms and descriptive names. This differed from our findings in Experiment 6a. Familiar eponyms were given higher JOLs than familiar descriptive names. This may be driven in part by older adult ratings as older adults provided lower JOLs for descriptive names compared to eponyms. Participants were more confident in their answers to questions about descriptive items compared to eponyms and in their answers to questions about familiar compared to unfamiliar items. Interestingly, confidence judgments were higher for unfamiliar descriptive names compared to unfamiliar eponyms but lower for familiar descriptive names than familiar eponyms. Younger and older adults benefitted from descriptive names when studying unfamiliar items. However, there was no difference in test performance between familiar descriptive names and eponyms for younger adults, but older adults performed worse for familiar descriptive names compared to familiar eponyms.

Experiment 7

Goal-directed behavior which is influenced by both intrinsic and extrinsic motivation (Duan et al., 2020; Ryan & Deci, 2000). Curiosity is a type of intrinsic motivation that can enhance memory and can activate similar brain regions as anticipated rewards (Kang et al., 2009). However, some work has suggested that the mechanisms by which extrinsic and intrinsic rewards enhance memory may be different. For example, curiosity may allocate attentional resources towards the target to-be-remembered information while monetary rewards may suppress task-irrelevant information (Duan et al., 2020). Therefore, curiosity is a type of intrinsic motivation to learn that can sustain attention and influence memory.

In a recent study, participants were presented with short readings about a variety of topics followed by a set of trivia questions related to the topics they read (Reichardt et al., 2023). Importantly, the answers to the questions were not in the readings. Participants subjectively rated their prior knowledge of the answers to the questions and their curiosity to know the answer after completing the readings but before being presented with the answers to the questions. Then, participants took a memory test where they had to recall answers to the questions. Results showed that curiosity was similar whether questions were related to the readings or not; however, curiosity enhanced memory for questions related to the readings. Further, participants overestimated their prior knowledge of the questions related to the readings. Other work has found that pretesting can enhance curiosity to learn the answers to trivia questions presented multiple times (Chen et al., manuscript under review). Chen and colleagues presented participants with trivia questions, asked them to provide curiosity ratings, and then presented them with the same questions again and asked participants to guess the answer before rating their curiosity again. This procedure resulted in enhanced curiosity for the same questions between trials, suggesting that guessing the answer influenced curiosity. Therefore, it seems that curiosity enhances memory, and this effect may be strengthened by repeated exposure to the questions or related information. On the other hand, prior knowledge of the studied material may lead to metacognitive overconfidence in memory for the information.

States of curiosity can even enhance memory for information presented temporally near information that elicited curiosity. Memory was worse for scholastic information presented after high-curiosity questions compared to low-curiosity questions (Keller et al., 2024) contradicting prior work showing the opposite trend (Fandacova & Gruber, 2021; Murphy et al., 2021; Stare et al., 2018). Therefore, states of curiosity may impact memory for unrelated information when learning. Curiosity is related to learning, but the mechanisms underlying these two constructs are different (Wade & Kidd, 2019). Curiosity is better predicted by metacognitive ratings of prior knowledge of the answer rather than actual prior knowledge whereas learning is better predicted by actual prior knowledge of the answer and curiosity to know the answer. Some work has proposed that curiosity may be a type of metacognition of how close someone is to closing a knowledge gap (Metcalfe et al., 2020). However, once a knowledge gap is closed, curiosity often declines as the quest for knowledge has been satiated. Further engagement with the material may or may not develop to form a deeper interest in the content (Hidi & Renninger, 2019).

Group work and collaborative learning has become much more common in classrooms over the years and is one of the most studied and effective practices in promoting student growth (Parr & Townsend, 2002). Though curiosity is known to enhance learning and memory and educators often design lessons in a way to foster curiosity in their students, most research on curiosity focuses on independent learning (Sinha et al., 2017). Some work has shown that collaborative learning may enhance curiosity through interpersonal knowledge identification and acquisition (e.g., asking each other questions, sharing information, evaluating each other's ideas). In Experiment 7 we were interested in exploring curiosity as an intrinsic motivator to remember. Additionally, we investigated the impact of collaborative learning on curiosity and memory for trivia facts. Though we are interested in these effects in participants across the lifespan, here we first explore these factors in an undergraduate sample.

Method

Participants

Participants were 100 UCLA undergraduate students who participated through the Psychology Human subjects pool and receive partial course credit for their participation (Age: 18-38, M = 20.62, SD = 3.26). On average, participants learned English before the age of 3 years old (M = 2.09, SD = 4.17). Participants shared their racial and gender identifies on a post-task questionnaire: 38 students identified as Asian/Pacific Islander, 26 as White, 18 as Latinx, 9 as Multiracial, 3 as Black, and 6 as Other, and 70 identified as female, 28 as male, and 2 as nonbinary. An additional 3 participants were tested but excluded from analyses for admitting to cheating (i.e., looking up the answers) on a post-experiment survey.

Materials

Materials used in Experiment 7 were 60 trivia questions from the McGillivray et al. (2015) database and adapted from Chen and colleagues (under review). The trivia questions were split into two lists of 30 questions where half of the questions on each list were categorized as eliciting low curiosity and half as eliciting high curiosity based on prior work by Chen and colleagues. See Appendix D. for a complete list of the trivia questions used in Experiment 7.

Procedure

All participants were tested in person at UCLA in one of three testing rooms. Each room had two desks each with a computer and chair. Participants were greeted by a research assistant and randomly assigned to one of two conditions: collaborative learning or independent learning. Participants in the independent learning condition sat in a room by themselves and studied 60 trivia questions. After each question was presented, they made a guess for the answer, rated their curiosity from 1 (*not at all curious*) to 10 (*extremely curious*), and then they received the answer to only half of the questions. Next, they viewed all 60 questions again, but this time they only typed in an answer to the questions they didn't receive answers to during the first study phase. Then, they rated their curiosity for those questions. They also received answers to only the 30 questions they did not receive the answers to on the first study phase. Finally, they were tested immediately afterwards on all 60 questions. They typed in their answers, rated their curiosity, and then received answers to all 60 questions. Finally, they completed a post-experiment survey to provide information about whether they cheated during the task, their demographic information, and whether they experienced any technical difficulties.

Participants in the collaborative condition participated in the same initial study phase and final test phase as participants in the independent condition, but their second study phase was with a second participant. Two participants sat in a room together each at separate desks. During the initial study phase, one partner received answers to questions on list 1 while the other partner received answers to questions on list 2. During the second study phase, participants only saw the questions they did not receive answers to and had to ask their partners for the answers. They would type in the answer their partner suggested, rate their curiosity, and then be shown the answer. For both conditions, the question lists were counterbalanced across participants. See

Figure 22 for a visual schematic of the procedure in Experiment 7.

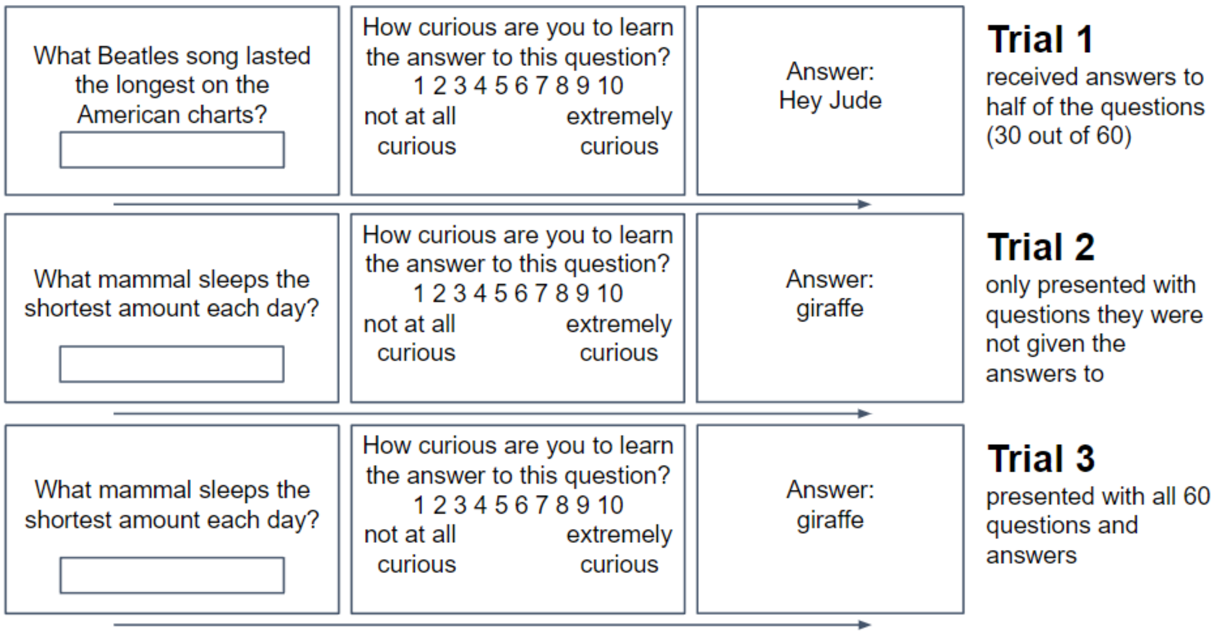


Figure 22. *Procedure for the trivia paradigm in Experiment 7.*

Results

Prior Knowledge. First, we wanted to test prior knowledge of the answers to the trivia questions to ensure that there were no differences in prior knowledge of the answers between conditions. We also included curiosity level in the model and the interaction between curiosity level and condition to examine any potential differences in prior knowledge by curiosity level. Because accuracy was a binary variable, we conducted a generalized linear mixed effects model. The model was fitted using maximum likelihood estimation with Laplace Approximation. The model demonstrated a good fit to the data with an Akaike Information Criterion (AIC) of 3886 and a Bayesian Information Criterion (BIC) of 3916. There were no differences in prior knowledge between the independent and collaborative group [e^B = 0.86, $CI_{95\%}$ = 0.57-1.30, z = -0.70, p = .483], between high and low curiosity questions [e^B = 1.02, $CI_{95\%}$ = 0.83-1.25, z = 0.16, p = .873], and there was no significant interaction between condition and curiosity level

[e^B = 1.26, $CI_{95\%}$ = 0.92-1.71, z = 1.43, p = .151]. The random effects analysis revealed significant variability in intercepts across individual participants with a variance estimate of 0.76 and a standard deviation of 0.87. The adjusted ICC was 0.19, which indicates that approximately 19% of the total variance in test performance can be attributed to differences between participants. Therefore, there were no differences in prior knowledge between conditions or curiosity level. However, we use prior knowledge as a predictor of performance and curiosity in subsequent analyses to understand how this relates to these outcomes as prior knowledge did vary between individual participants.

Initial Curiosity Ratings. To follow up on the categorization of trivia questions into low and high curiosity categories by Chen and colleagues (under review), we wanted to examine initial curiosity ratings by curiosity level and prior knowledge and the interaction between these two predictors. We conducted a linear mixed model analysis to investigate the effects of prior knowledge, curiosity level, and their interaction on initial curiosity ratings while accounting for the hierarchical structure of the data with items nested within participants. We also added condition in as a covariate to test whether initial curiosity ratings varied between the independent and collaborative groups. The model demonstrated convergence with a REML criterion of 14192.1. The analysis revealed a significant main effect of curiosity level [b = -0.53, SE = 0.13, t(2904.48) = -4.13, p < .001] such that items classified as low-curiosity questions received significantly lower curiosity ratings than high-curiosity judgments. No other predictors were significant (all ps > .05). Therefore, prior knowledge did not significantly influence curiosity ratings, and this did not vary by curiosity level of each item. Furthermore, curiosity ratings on the first study trial did not vary by condition. This is an important finding as we test for differences in curiosity ratings during the second phase which we did expect to vary by condition. The adjusted ICC for the random intercept model was 0.26 indicating that approximately 26% of the total variance in curiosity ratings can be attributed to between-participant differences.

Curiosity Ratings during the Intervention Phase. We conducted a linear mixed model analysis to investigate the effects of prior curiosity level, condition, and their interaction on curiosity ratings during the second study phase while accounting for the hierarchical structure of the data with items nested within participants. We also added prior knowledge and initial curiosity ratings in as covariates. The model demonstrated convergence with a REML criterion of 12882. The analysis revealed a significant main effect of condition on curiosity ratings [b = 0.85, SE = 0.35, t(105.87) = 2.41, p = .018] such that participants collaborating with a partner provided significantly higher curiosity ratings compared to participants working alone despite both groups having knowledge gaps on the questions that were rated. Additionally, curiosity ratings on the initial trial significantly predicted curiosity on the second trial [b = 0.26, SE = 0.01, t(2970.10) = 17.71, p < .001] such that higher curiosity ratings on trial 1 predicted higher curiosity ratings on trial 2. No other predictors were significant (all ps > .05). Therefore, prior knowledge, curiosity level, and the interaction between curiosity level and condition did not significantly influence curiosity ratings. The adjusted ICC for the random intercept model was 0.42 indicating that approximately 42% of the total variance in curiosity ratings can be attributed to between-participant differences.

Memory for Trivia Answers on the Final Test. We conducted a generalized linear mixed effects model to explore the influence of condition, curiosity level, prior knowledge, and curiosity ratings on memory for trivia answers on the final test. We also tested for an interaction between curiosity ratings on the second study trial and condition as these ratings varied by condition during the intervention phase. The model was fitted using maximum likelihood estimation with Laplace Approximation. The model demonstrated a good fit to the data with an Akaike Information Criterion (AIC) of 2792.3 and a Bayesian Information Criterion (BIC) of 2834.4 There were no differences in memory performance between the independent and collaborative group [e^B = 0.67, $CI_{95\%}$ = 0.36-1.25, *z* = -1.25, *p* = .210], suggesting that both groups performed similarly on the final memory test. There was a significant effect of curiosity level [e^B = 0.63, $CI_{95\%}$ = 0.52-0.76, *z* = -4.74, *p* < .001], curiosity ratings during the second study trial [e^B = 1.06, $CI_{95\%}$ = 1.01-1.12, *z* = 2.22, *p* = .026], and prior knowledge [e^B = 2.19, $CI_{95\%}$ = 1.78-2.69, *z* = 7.41, *p* < .001]. Therefore, participants more accurate in their responses to high curiosity questions compared to low curiosity questions, having higher prior knowledge of the questions during the initial phase predicted better accuracy on the final test, and higher curiosity rating during the second trial predicted performance on the final test. Despite the collaborative condition having higher curiosity ratings during the second trial, there was no interaction between these ratings and condition on performance on the final test [e^B = 1.02, $CI_{95\%}$ = 0.94-1.11, *z* = 0.43, *p* = .670]. The random effects analysis revealed significant variability in intercepts across individual participants with a variance estimate of 0.57 and a standard deviation of 0.76. The adjusted ICC was 0.15, which indicates that approximately 15% of the total variance in test performance can be attributed to differences between participants.

Change in Curiosity Ratings. We conducted a linear mixed model analysis to examine how curiosity changed with each trial and how this effect may have varied by condition. We added curiosity level of each item and accuracy on the memory test as covariates in the analyses. The model demonstrated convergence with a REML criterion of 42780.8. The analysis revealed a significant main effect of trial on curiosity ratings [*b* = -1.17, *SE* = 0.05, *t*(8899.11) = -25.98, *p*

$< .001$] such that curiosity ratings decreased with each subsequent trial. Curiosity level [$b = -0.17$, $SE = 0.05$, $t(8896.03) = -3.06$, $p = .002$] and memory accuracy [$b = -0.36$, $SE = 0.06$, $t(8952.86) = -5.72$, $p < .001$] significantly predicted curiosity ratings such that curiosity was lower for low curiosity questions and better memory performance predicted lower curiosity ratings. The analysis revealed a significant interaction effect between trial and condition on curiosity ratings [$b = 0.16$, $SE = 0.07$, $t(8896.02) = 2.33$, $p = .020$] such that curiosity ratings decreased with each subsequent trial but this effect depended on condition, but there was no main effect of condition on curiosity ratings [$b = 0.31$, $SE = 0.32$, $t(141.67) = 0.97$, $p = .333$]. The adjusted ICC for the random intercept model was 0.24 indicating that approximately 24% of the total variance in curiosity ratings can be attributed to between-participant differences.

A follow-up test of simple slopes was conducted to probe the interaction effect between trial and condition on curiosity ratings. The slope of trial was significant and negative for both the independent [$b = -1.23$, $SE = 0.05$, $t(8996) = 24.28$, $p < .001$] and collaborative [$b = -1.07$, $SE = 0.06$, $t(8996) = 18.71$, $p < .001$]. We conducted a contrast analysis to examine whether the effect of slope varied between the two conditions and found that the slope of trial was significantly more negative for the independent compared to the collaborative condition [$b = 0.16$, $SE = 0.07$, $z = 2.33$, $p < .020$], suggesting that curiosity ratings decreased with each trial at a greater rate in the independent condition (see Figure 23).

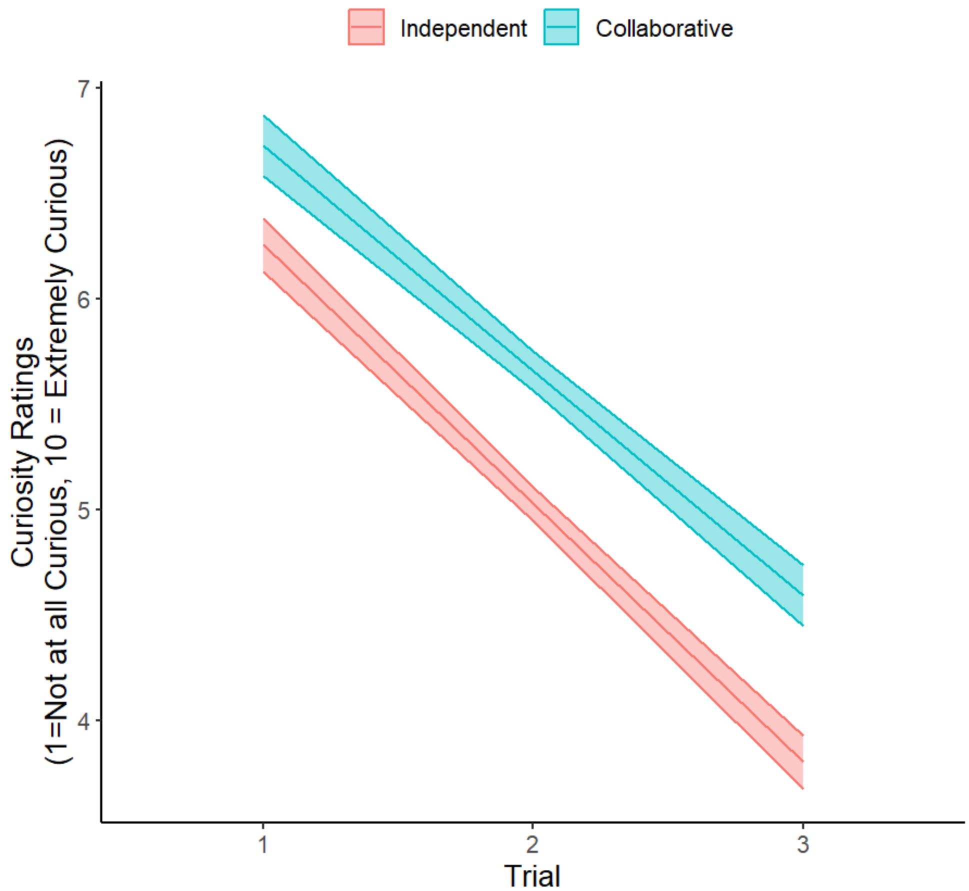


Figure 23. *Curiosity ratings as a function of trial and condition. The legend shows the Independent condition is represented by the red line while the Collaborative condition is represented by the blue line. Confidence bands represent 95% confidence intervals.*

Experiment 7 Discussion

In Experiment 7 we found that collaborating on the answers to trivia questions resulted in higher curiosity for the answers compared to working alone, though collaboration did not impact memory for the answers. Additionally, we found that curiosity decreased with repetition which fits with prior work highlighting the importance of novelty for curiosity (Kashdan & Silva, 2009), but contrasts with other work showing that curiosity can increase across repetitions of trivia questions (Chen et al., under review). Curiosity did predict memory for trivia questions, which also aligns with prior research.

According to the literature, the relationship between intellectual curiosity and age is usually negative. Chu et al. (2020) found a negative association between age and intellectual curiosity even after controlling for sex, culture, and education level. They conducted a moderated serial multiple mediation model and found an indirect effect of age on curiosity through future time perspective and importance of curiosity, but no direct effect of age on curiosity. The authors suggest that their results have implications for understanding the relationship between intellectual curiosity and age, such that as future time becomes more limited with advancing age, curiosity becomes less valuable.

There are some situations in which older adults may be just as intellectually curious as younger adults. Importantly, curiosity may be important in older adulthood as it has been associated with better physical, psychological, cognitive, and social well-being. Swan and Carmelli (1996) found that older adults who were still living after a five-year longitudinal study, previously reported higher levels of state and trait curiosity than those who were deceased. Curiosity is associated with life satisfaction, positive effect, and meaning of life (Wang & Li, 2015). Older adults show less intellectual curiosity than middle-aged adults (Zimprich et al., 2009). One explanation for why this type of curiosity declines in old age is the socioemotional selectivity theory (Carstensen, 1995). If time is perceived as more limited, a person is less likely to make it a priority to meet new people or learn new skills. However, if an older adult wants to maximize their social interactions, they may prioritize information that they could later share with important people in their lives like friends, neighbors, and grandchildren. Future work should explore the effect of collaboration during learning on curiosity and memory for trivia facts in a lifespan sample.

Chapter 4 Conclusions

Experiment 6 provides additional evidence of the influence of familiarity on memory and metacognitive awareness of learning. Our findings suggest that it may be important to consider how information associated with proper names is presented when teaching in a complex domain, especially to promote more age-inclusive instructional practices. Age may be an important consideration when learning medical information as descriptive names were beneficial for participants when items were unfamiliar, but harmful for older adults when items were familiar yet they had no effect for younger adults when items were familiar. Older adults seem to be aware of this effect as their metacognitive ratings reflect a similar trend where familiar eponyms receive higher ratings compared to familiar descriptive names.

Experiment 7 explored how other variables such as curiosity and social connection motivated learners to remember information and curiosity to learn more. We found that curiosity promotes memory for trivia facts, replicating this finding from the literature (Chen et al., under review; Gruber et al., 2014; McGillivray et al., 2015). Though curiosity decreased with each trial, it decreased less rapidly when participants collaborated with a partner. Thus, collaborative learning may influence curiosity when learning and should be a consideration for future research.

CHAPTER 5: CONCLUSIONS

Learning goals may vary from context to context as well as across the lifespan. Often in the classroom, teachers determine what is important to remember and learn and students try their best to pay attention to cues and feedback from their instructors to succeed. However, even in classroom learning students come in with prior knowledge and motivation to learn. Some students will quickly learn what types of questions get asked on exams or have developed helpful study strategies to support regulate their learning when they find new information difficult to grasp. Other learners may not have as well-developed study habits and may even enter the classroom with cognitive deficits that make it hard for them to pay attention, regulate their time, or remember information.

In Experiments 1, 2, and 3, rewards combined with existing schemas were effective in supporting participants in learning the schematic reward structures present in the word lists. Further, metacognitive monitoring and control seems to play an important role in this process. Experiment 3 provided more information about how this type of strategy supports learning and memory for older adults. Students may not always understand why certain assignments, or a certain question are worth more points than others. In our laboratory-based tasks, we find that when points are meaningfully tied to semantically related information, learners are successful in predicting what information is most important to remember. Interestingly, we also found that older adults prioritized high-value items even when they did not study words paired with rewards but did not observe this effect in younger adults. It could be that older adults are motivated to remember items through their category membership in addition to the awards associated with each category whereas younger adults may benefit from having the rewards visibly present during encoding. We are interested in following up on these findings in an additional sample of older adults compared to a sample of younger adults from UCLA to see if we replicate these findings. We were surprised not to observe an age-related deficit in transfer performance and are interested in seeing how older adults perform in comparison to an undergraduate student sample who may be more motivated to perform well on the task compared to our participants recruited through Prime Panels in Experiment 3. We did not test older adults in a delayed condition as we did younger adults in Experiment 1b, nor did we test whether they could adapt to multiple schematic reward structures like in Experiment 2. Therefore, future work should explore how age relates to performance on this paradigm in these additional contexts.

Experiment 4 investigated learners in an upper division psychology course to better understand how being more selective with one's time and study habits may relate to selective memory abilities and exam performance. We did not observe a relationship between study selectivity and memory selectivity in Experiment 4, though we did find that interest in the course predicted higher memory selectivity. Therefore, our participants seemed to be motivated to perform well on the VDR task not only due to earning points during the task or extra credit in the class, but also because they are more interested in Cognitive Neuroscience. Students with higher GPAs were more underconfident in their exam performance despite performing better on exams while lower achieving students were overconfident. Metacognitive skillfulness continues to be an important predictor of memory and learning in both laboratory and classroom settings.

It could be that some learners are not selective with their study habits because they are not able to identify what is important to remember or could benefit from extra information about why some information is more important than others. Therefore, Experiment 5 examined how value-directed learning, or being able to use rewards to determine what makes something important to remember, may be a helpful strategy for identifying connections between studied items. Understanding relationships between different concepts can support predictions of what new information will be important to remember.

Students in Experiment 5 were enrolled in a required Research Methods course where they took two exams. Throughout the course, students were provided with opportunities to practice old exam questions from previous quarters. The types of questions asked on the exams always highlight the same important concepts. Examining the relationship between students' performance on our value-directed learning task and exam performance allowed us to better understand whether the ability to use rewards to guide learning of what is important relates to real-world academic outcomes. We found that more selective studiers performed better on our VDL task and on their Research Methods in Psychology exams. Though we did not find a similar relationship between study selectivity and exam performance in Experiment 4, there were some major differences between these two samples. Therefore, our small sample size in Experiment 4 compared to our larger sample size in Experiment 5 could be one explanation for the differences in results we observed. The exams in Cognitive Neuroscience were short answer exams whereas the exams in Research Methods were multiple-choice. Therefore, it could be more beneficial to be selective on a multiple-choice exam where you must distinguish between correct and incorrect answer choices compared to a short answer exam where you are constructing a written response to a prompt.

Additionally, we found that motivation to do coursework was related to exam performance in both courses though the direction of the effect of motivation on exam performance varied between experiments. In Cognitive Neuroscience, students who were more motivated performed better on exams while in Research Methods, students who were more motivated performed worse on exams. We are unsure of the explanation for this difference, but future work should explore how motivation to do coursework relates to exam performance on exams of different formats and in different subject areas.

However, some students may not be well-motivated by extrinsic incentives or may not have enough metacognitive skillfulness, executive functioning skills, or interest in the material to identify what is important to learn. In Experiment 5, we found that students reporting more symptoms of inattention and hyperactivity performed worse on exams even though there was no effect of these symptoms on transfer performance on our VDL task. Further, learning often happens outside of the classroom. While there are increasingly more older adults enrolled in college courses and programs, we continue to learn across our lifespans whether we continue our formal education or not. It is important to understand what motivates learning across the lifespan when there are not grades to earn.

Curiosity and metacognitive awareness may be important motivators to learn. Older adults who are metacognitively aware of their limited cognitive abilities as they grow older may be interested in stimulating their minds to remain healthy in older age. In Experiment 6 we found that older adults were aware that they were less likely to remember information paired with familiar descriptive names compared to familiar eponyms. In contrast, younger adults showed no difference in memory between familiar eponyms and descriptive names. This could be due to older adults having more familiarity of these familiar medical eponyms.

An important part of survival is social connection and is a motivator of many choices to learn. Learning information with the goal of sharing it with someone else can be a powerful motivator for learning. In Experiment 7 we found that curiosity influenced memory for trivia facts while collaboration enhanced curiosity while learning. Future work should explore curiosity while collaborating in older adults and how this impacts their memory for information.

Understanding motivations to learn across the lifespan is important for instructors teaching in diverse classrooms, doctors and other care providers who educate older adults in healthcare contexts, and for individuals as they age who want to empower themselves and self-regulate their learning. Taken together, these studies provide insight on how well both intrinsic and extrinsic motivational strategies support learning and memory in the classroom and across the lifespan.

Appendix A. Items used in the word lists in Experiments 1a, 1b, and 3

Birds	Fish	Mammals
bluejay	anchovy	aardvark
dove	angelfish	alpaca
eagle	barracuda	baboon
emu	bass	badger
falcon	carp	buffalo
ferret	catfish	camel
flamingo	clownfish	dingo
goose	cod	donkey
gull	dory	elephant
hawk	eel	elk
kookaburra	flounder	gazelle
loon	guppy	giraffe
macaw	halibut	gopher
mockingbird	koi	hedgehog
owl	marlin	human
parrot	minnow	hyena
partridge	perch	jaguar
peacock	roughy	koala
pelican	salmon	lion
penguin	sardine	mongoose
pigeon	seahorse	muskrat
quail	shark	otter
raven	snapper	puma
rooster	sole	raccoon
sparrow	stingray	reindeer
stork	sturgeon	squirrel
swallow	sunfish	turkey
swan	swordfish	whale
vulture	trout	wolf
woodpecker	tuna	zebra

Appendix B. Items used in the word lists in Experiment 2

Birds	Fish	Mammals
parrot	shark	aardvark
flamingo	tuna	baboon
penguin	carp	whale
emu	eel	camel
hawk	flounder	donkey
owl	halibut	badger
dove	cod	muskrat
eagle	trout	mongoose
raven	perch	lion
vulture	bass	koala

Fruit	Vegetables	Meat
apples	corn	salami
oranges	broccoli	chicken
lemons	celery	turkey
grapes	peas	pork
bananas	spinach	ham
pineapples	beets	veal
plums	kale	pepperoni
cherries	cabbage	beef
apricots	onions	quail
peaches	carrots	bologna

Clothing	Shoes	Jewelry
pants	flip-flops	earrings
dress	sneakers	necklace
coat	boots	ring
skirt	loafers	bracelet
shorts	socks	anklet
jacket	sandals	choker
shirt	slippers	brooch
jumpsuit	moccasins	pins
jumper	espadrilles	cufflink
jeans	heels	armlet

Bedroom	Bathroom	Kitchen
bed	toilet	fork
drawers	shower	strainer
wardrobe	bathtub	tupperware
armchair	toothbrush	plate
sheets	shampoo	saucepan
quilt	razor	kettle
hanger	urinal	cooker
nightstand	floss	toaster
pillow	loofah	fridge
mattress	mouthwash	mixer

Air	Land	Water
spaceship	truck	jetski
rocket	van	cruise
plane	wagon	lifeboat
hovercraft	bus	canoe
jet	excavator	boat
balloon	pickup	kayak
blimp	tractor	ferry
helicopter	taxi	sailboat
drone	sedan	yacht
glider	motorcycle	submarine

Herbs	Trees	Flowers
cilantro	palm	tulip
parsley	cedar	orchid
dill	birch	peony
chive	sequoia	violet
rosemary	redwood	rose
sage	pine	poppy
thyme	oak	lily
oregano	maple	iris
basil	elm	lilac
mint	bamboo	daisy

Appendix C. Items used in the word lists in Experiment 5.

Themes							
Animals		***Plants***		***Fashion***		***Vehicles***	
Categories							
Fish	***Birds***	***Trees***	***Flowers***	***Clothing***	***Shoes***	***Land***	***Air***
shark	parrot	palm	tulip	pants	flip-flops	truck	spaceship
tuna	flamingo	cedar	orchid	dress	sneakers	van	rocket
carp	penguin	birch	peony	coat	boots	wagon	plane
eel	emu	sequoia	violet	skirt	loafers	bus	hovercraft
flounder	hawk	redwood	rose	shorts	socks	excavator	jet
halibut	owl	pine	poppy	jacket	sandals	pickup	balloon
cod	dove	oak	lily	shirt	slippers	tractor	blimp
trout	eagle	maple	iris	jumpsuit	moccasins	taxi	helicopter
perch	raven	elm	lilac	jumper	espadrilles	sedan	drone
bass	vulture	bamboo	daisy	jeans	heels	motorcycle	glider

Appendix D. Complete list of materials used in Experiments 6a and 6b.

Eponym	Descriptive Name	Facts	
		Unfamiliar Conditions	
Addison's Disease	Adrenal Insufficiency	Cause	protein buildup
		Symptom	weight loss
		Treatment	hydrocortisone pills
Gardner's Syndrome	Intestinal Polyposis	Cause	cell growth
		Symptom	adrenal masses
		Treatment	tumor removal
Fanconi's Disease	Hypoplastic Anemia	Cause	autosomal inheritance
		Symptom	excessive thirst
		Treatment	blood transfusions
Gilbert's Syndrome	Familial Jaundice	Cause	enzyme dysfunction
		Symptom	yellowish skin
		Treatment	stress management
Alibert's Disease	T-Cell Lymphoma	Cause	abnormal cells
		Symptom	skin lesions
		Treatment	light therapy
Chiari's Syndrome	Cerebellar Malformation	Cause	small skull
		Symptom	neck pain
		Treatment	pain medication
Marshall's Syndrome	Saddle Nose	Cause	collagen abnormalities
		Symptom	hearing loss
		Treatment	annual check-ups
Anton's Syndrome	Visual Anosognosia	Cause	brain damage
		Symptom	cortical blindness
		Treatment	risk management

Eponym	Descriptive Name	Facts	
	Familiar Conditions		
Down’s Syndrome	Trisomy 21	Cause	extra chromosome
		Symptom	developmental delays
		Treatment	self-care skills
Alzheimer’s Disease	Presenile Dementia	Cause	brain changes
		Symptom	memory impairment
		Treatment	behavioral intervention
Tourette’s Syndrome	Tic Disorder	Cause	genetic factors
		Symptom	repeated movements
		Treatment	stimulant medication
Parkinson’s Disease	Shaking Palsy	Cause	impaired cells
		Symptom	slowed movement
		Treatment	brain stimulation
Crohn’s Disease	Inflammatory Bowels	Cause	autoimmune reaction
		Symptom	abdominal pain
		Treatment	anti-diarrheal medication
Huntington’s Disease	Progressive Chorea	Cause	genetic mutation
		Symptom	involuntary movements
		Treatment	occupational therapy
Asperger’s Syndrome	High-Functioning Autism	Cause	preconception mutations
		Symptom	flat speech
		Treatment	behavior analysis
Salmonella’s Disease	Food Poisoning	Cause	ingesting bacteria
		Symptom	severe diarrhea
		Treatment	rehydration liquids

Appendix E. Complete list of trivia questions used in Experiment 7.

Question	Answer	A-Priori Curiosity	List
What is the name of the island country that lies off the southeast coast of India?	Sri Lanka	low	1
What vegetable did ancient Egyptians place in their right hand when taking an oath?	onion	low	1
There are five halogen elements including Fluorine, Chlorine, Bromine, and Astatine. What is the name of the fifth?	iodine	low	1
What gas forms almost 80% of Earth's atmosphere?	nitrogen	low	1
What was a gladiator armed with in addition to a dagger and spear?	net	low	1
What unit of measurement is used for fuel wood?	cord	low	1
Who was the first person to use the V sign as a victory sign?	Winston Churchill	low	1
What 17th century artist painted more than 60 self-portraits?	Rembrandt	low	1
What organ of the buffalo did Plains Indians use to make yellow paint?	gallbladder	low	1
In parts of India, the older brother must marry first. If he cannot find a wife, what can choose to marry?	a tree	low	1
What product is second, only to oil, in terms of the largest trade volumes in the world?	coffee	low	1
What is the name of the scientific scale used for measuring the hardness of rocks?	Moh's scale	low	1
What was the first animated film to be nominated for an Oscar for best picture?	Beauty and the Beast	low	1
What was the first nation to give women the right to vote?	New Zealand	low	1
What note do most American car horns beep in?	F	low	1
What organ destroys old red blood cells?	spleen	high	1
What is the oldest written code of law in history?	Hammurabi's code	high	1
What industry used 20% of China's harvested plants?	medicine	high	1
What handicap did Thomas Edison suffer from?	deafness	high	1
The Gold Coast is now known as what country?	Ghana	high	1
What is the slowest swimming fish in the world?	seahorse	high	1
What was Dr. Frankenstein's first name?	Victor	high	1
What mammal sleeps the shortest amount each day?	giraffe	high	1

What was the only type of product ever promoted by Elvis Presley in a television commercial?	donuts	high	1
What is the longest common English word without any vowels?	rhythm(s)	high	1
What did girls in medieval Spain put in their mouths to avoid unwanted kisses?	toothpicks	high	1
What part of a woman’s body were ancient Chinese artists forbidden to paint?	foot	high	1
What trade was Greek philosopher Socrates trained for?	stonecutting	high	1
What is the only consumable food that won’t spoil?	honey	high	1
Before the barometer, what animal did German meteorologists use to predict air pressure changes?	frog	high	1
What novel contains the longest sentence in literature with 832 words?	Les Miserables	low	2
What is the monetary unit of Korea?	won	low	2
What is the only type of bird that has nostrils at the tip of its beak?	kiwi	low	2
What is the name of the instrument used to measure wind speed?	anemometer	low	2
Which scientist was the first to receive the Nobel Prize twice?	Marie Curie	low	2
What Beatles song lasted the longest on the American charts?	Hey Jude	low	2
What city has the only drive thru post office in the world?	Chicago	low	2
What is the biggest constellation in the sky?	hydra	low	2
With what product did the term “brand name” originate?	whiskey	low	2
What country has the highest population density?	Monaco	low	2
What world capital city has the fewest cinemas in relation to its population?	Cairo, Egypt	low	2
In what country is Angel falls, the tallest waterfall, located?	Venezuela	low	2
What reptile, according to ancient legend, was able to live in fire?	salamander	low	2
What fish produces more than 200 million eggs at a time?	sunfish	low	2
What American novel was the first to sell over 1 million copies?	Uncle Tom’s Cabin	low	2
What was the first product to have a bar code?	Wrigley’s gum	high	2
Which metal is the best conductor of electricity?	silver	high	2
What city has the shortest name in the world?	Y (France)	high	2

What is the only country in the world that has a bible on its flag?	Dominican Republic	high	2
What is the most common first name in the world?	Mohammed	high	2
What city is referred to as the Pittsburgh of the South?	Birmingham, Alabama	high	2
What instrument was invented to sound like a human singing?	violin	high	2
What animal’s excrements are consumed as a luxury food?	bats	high	2
What food will make a drug test show up positive?	poppy seeds	high	2
What was the name of Smokey the Bear’s mate?	Goldie	high	2
What is the only planet in our solar system that rotates clockwise?	Venus	high	2
What is the hardest natural substance known?	diamond	high	2
What snack food can be used as an ingredient in the explosive dynamite?	peanuts	high	2
Who was the first Christian Emperor of Rome?	Constantine	high	2
Setting a world record, how many days can a human stay awake?	11	high	2

REFERENCES

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum. https://doi.org/10.4324/9780203771730

Badham, S. P. (2024). The older population is more cognitively able than in the past and age-related deficits in cognition are diminishing over time. *Developmental Review, 72*, 101124. https://doi.org/10.1016/j.dr.2024.101124

Balota, D. A., Yap, M. J., & Hutchison, K. A. (2007). The English Lexicon Project. *Behavior Research Methods 39*, 445-459. https://doi.org/10.3758/BF03193014

Barr, J. (2016). The plus 50 population in higher education: Challenges and triumphs. The *Journal of Continuing Higher Education, 64*(1), 51-55. https://doi.org/10.1080/07377363.2016.1130570

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge, England: Cambridge University Press.

Bartolucci, S. L., Stedman, T. L., & Forbis, P. (2005). Stedman's medical eponyms (2nd ed.). *Lippincott Williams & Wilkins*.

Bijleveld, E., Custers, R., & Aarts, H. (2012). Human reward pursuit: From rudimentary to higher-level functions. *Current Directions in Psychological Science, 21*(3), 194-199. https://psycnet.apa.org/doi/10.1177/0963721412438463

Boedecker, J., Lampe, T., & Riedmiller, M. (2013). Modeling effects of intrinsic and extrinsic rewards on the competition between striatal learning systems. *Frontiers in Psychology, 16*(4), Article 739. https://doi.org/10.3389/fgsyg.2013.00739

Bowen, L., Montepare, J., Silverstein, N., Whitbourne, S., & Beaulieu, C. (2022). Making the case for age inclusivity in higher education. *Innovation in Aging*, *6*(1),

268. https://doi.org/10.1093/geroni/igac059.1062

Bransford, J. D., & Schwartz, D. L. (1999). Chapter 3: Rethinking transfer: A simple proposal with multiple implica-tions. *Review of Research in Education*, *24*, 61-100. https://doi.org/10.2307/1167267

Brewer, W. F., & Lichtenstein, E. H. (1981). Event schemas, story schemas, and story grammars. In J. Long & A. Baddely (Eds.), *Attention and performance IX* (pp. 363-379). Hillsdale, NJ: Lawrence Erlbaum.

Brewer, W. F., & Lichtenstein, E. H. (1982). Stories are to entertain: A structural-affect theory of stories. *Journal of Pragmatics*, *6*, 473-486. https://doi.org/10.1016/0378-2166(82)90021-2

Brown, S. C., & Park, D. C. (2002). Roles of age and familiarity in learning health information. *Educational Gerontology*, *28*(8), 695-710. https://doi.org/10.1080/03601270290099723

Cabeza, R., Daselaar, S. M., Dolcos, F., Prince, S. E., Budde, M., & Nyberg, L. (2004). Task-independent and task-specific age effects on brain activity during working memory, visual attention and episodic retrieval. *Cerebral cortex*, *14*(4), 364-375. https://psycnet.apa.org/doi/10.1093/cercor/bhg133

Carlson, C. L., Booth, J. E., Shin, M., & Canu, W. H. (2002). Parent-, teacher-, and self-rated motivational styles in ADHD subtypes. *Journal of Learning Disabilities*, *35*(2), 104-113. https://doi.org/10.1177/002221940203500202

Carstensen, L. L. (1995). Evidence for a life-span theory of socioemotional selectivity. *Current Directions in Psychological Science*, *4*(5), 151-156. https://doi.org/10.1111/1467-8721.ep11512261

Cassady, J. C. (2004). The influence of cognitive test anxiety across the learning–testing cycle. *Learning and Instruction*, *14*(6), 569-592. https://doi.org/10.1016/j.learninstruc.2004.09.002

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: the role of schematic support. *Psychology and Aging*, *20*(4), 718-721. https://doi.org/10.1037/0882-7974.20.4.718

Castel A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition*, *30*, 1078-1085. https://doi.org/10.3758/BF03194325

Castel, A. D., Farb, N. A., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition*, *35*, 689-700. https://doi.org/10.3758/BF03193307

Castel, A. D., Lee, S. S., Humphreys, K. L., & Moore, A. N. (2011). Memory capacity, selective control, and value-directed remembering in children with and without attention-deficit/hyperactivity disorder. *Neuropsychology*, *25*(1), 15-24. https://psycnet.apa.org/doi/10.1037/a0020298

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. https://psycnet.apa.org/doi/10.1037/h0046743

Chandler, J., Rosenzweig, C., Moss, A. J., Robinson, J., & Litman, L. (2019). Online panels in social science research: Expanding sampling methods beyond Mechanical Turk. *Behavior Research Methods*, *51*(5), 2022-2038. https://doi.org/10.3758/s13428-019-01273-7

Chelazzi, L., Perlato, A., Santandrea, E., & Libera, C. D. (2013). Rewards teach visual selective attention. *Vision Research, 85*, 58-72. https://doi.org/10.1016/j.visres.2012.12.005

Chen, A., Whatley, M. C., Halamish, V., & Castel, A. D. (under review). Does repetition or novelty enhance curiosity? Implications for memory and motivated learning.

Chu, L., Tsai, J. L, & Fung, H. H. (2020). Association between age and intellectual curiosity: The mediating roles of future time perspective and importance of curiosity. *European Journal of Ageing, 18,* 1-9. https://doi.org/10.1007/s10433-020-00567-6

Cohen, G., & Faulkner, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology*, *4*(2), 187-197. https://doi.org/10.1111/j.2044-835X.1986.tb01010.x

Cohen, M. S., Rissman, J., Suthana, N. A., Castel, A. D., & Knowlton, B. J. (2014). Value-based modulation of memory encoding involves strategic engagement of fronto-temporal semantic processing regions. *Cognitive, Affective, & Behavioral Neuroscience, 14*, 578-592. https://doi.org/10.3758/s13415-014-0275-x

Craik, F. I. M., & Bosman, B. A. (1992). Age-related changes in memory and learning. In H. Bouma & J. A. M. Graafmans (Eds.), *Gerontechnology* (pp. 79–92). Amsterdam: IOS Press.

Dauenhauer, J., Hazzan, A., Heffernan, K., & Milliner, C. M. (2022). Faculty perceptions of engaging older adults in higher education: The need for intergenerational pedagogy. *Gerontology & Geriatrics Education, 43*(4), 499-519. https://doi.org/10.1080/02701960.2021.1910506

Davidson, P. S. R., & Glisky, E. L. (2002). Neuropsychological correlates of recollection and familiarity in normal aging. *Cognitive, Affective & Behavioral Neuroscience, 2*(2), 174–186. https://doi.org/10.3758/CABN.2.2.174

DeLozier, S., & Rhodes, M. G. (2015). The impact of value-directed remembering on the own-race bias. *Acta Psychologica*, *154*, 62-68. https://psycnet.apa.org/doi/10.1016/j.actpsy.2014.11.009

Duan, H., Fernández, G., van Dongen, E., & Kohn, N. (2020). The effect of intrinsic and extrinsic motivation on memory formation: insight from behavioral and imaging study. *Brain Structure and Function*, *225*, 1561-1574. https://doi.org/10.1007/s00429-020-02074-x

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Thousand Oaks, CA: Sage.

Dunlosky, J., & Mueller, M. (2016). Recommendations for exploring the disfluency hypothesis toward establishing whether perceptually degrading materials impacts performance. *Metacognition and Learning*, *11*(1), 123-131. https://doi.org/10.1007/s11409-016-9155-9

Eich, T. S., & Castel, A. D. (2016). The cognitive control of emotional versus value-based information in younger and older adults. *Psychology and Aging*, *31*(5), 503.

Fandakova, Y. & Gruber, M. J. (2021). States of curiosity and interest enhance memory differently in adolescents and in children. *Developmental Science*, *24*(1), e13005. https://doi.org/10.1111/desc.13005

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Fogler, K. A., & James, L. E. (2007). Charlie Brown versus Snow White: The effects of descriptiveness on young and older adults' retrieval of proper names. *The Journals of*

Gerontology Series B: Psychological Sciences and Social Sciences, *62*(4), 201-207. https://doi.org/10.1093/geronb/62.4.P201

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(2), 234-257. https://psycnet.apa.org/doi/10.1037/0278-7393.10.2.234

Fries, L., Son, J. Y., Givvin, K. B., & Stigler, J. W. (2020). Practicing connections: A framework to guide instructional design for developing understanding in complex domains. *Educational Psychology Review*, *33*(2), 739-762. https://psycnet.apa.org/doi/10.1007/s10648-020-09561-x

Galli, G., Sirota, M., Gruber, M. J, Ivanof, B. E., Ganesh, J., Materassi, M., Thorpe, A., Loaiza, V., Cappelletti, M., & Craik, F. I. M. (2018). Learning facts during aging: the benefits of curiosity. *Experimental Aging Research*, *4:4*, 311-328. https://doi.org/10.1080/0361073X.2018.1477355

Garcia, M. A., & Kornell, N. (2014). Collector [Software]. Retrieved from https://github.com/gikeymarcia/Collector

Givvin, K. B., Stigler, J. W., & Thompson, B. J. (2011). What community college developmental mathematics students understand about mathematics, Part II: The interviews. The MathAMATYC *Educator*, *2*(3), 4–18.

Graessner, A. C., & Nakamura, G. V. (1982). The impact of schema on comprehension and memory. In G. H. Bower (Ed.), The psychology of learning and motivation (Vol. 16. pp. 59-109). New York: Academic Press.

Greeno, J. G., Moore, J. L., & Smith, D. R. (1993). Transfer of situated learning. In D. K. Detterman & R. J. Sternberg (Eds.), *Transfer on trial: Intelligence, cognition, and* *instruction* (pp. 99–167). Ablex Publishing.

Griffin, M. L., Benjamin, A. S., Sahakyan, L., & Stanley, S. E. (2019). A matter of priorities: High working memory enables (slightly) superior value-directed remembering. *Journal of Memory and Language*, *108*, Article 104032. https://doi.org/10.1016/j.jml.2019.104032

Grossnickle, E. M. (2016). Disentangling curiosity: Dimensionality, definitions, and distinctions from interest in educational contexts. *Educational Psychology Review*, *28*(1), 23-60. https://psycnet.apa.org/doi/10.1007/s10648-014-9294-y

Hall, J. C. (1957). Correlation of a modified form of Raven's progressive matrices (1938) with the Wechsler adult intelligence scale. *Journal of Consulting Psychology*, *21*(1), 23-26. https://psycnet.apa.org/doi/10.1037/h0041794

Hayes, M. G., Kelly, A. J., & Smith, A. D. (2013). Working memory and the strategic control of attention in older and younger adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *68*(2), 176-183. https://psycnet.apa.org/doi/10.1093/geronb/gbs057

Hennessee, J. P., Patterson, T. K., Castel, A. D., & Knowlton, B. J. (2019). Forget me not: Encoding processes in value-directed remembering. *Journal of Memory and Language*, *106*, 29-39. https://psycnet.apa.org/doi/10.1016/j.jml.2019.02.001

Hidi, S. E., & Renninger, K. A (2019). Interest development and its relation to curiosity: Needed neuroscientific research. *Educational Psychology Review*, *31*, 833–852. https://doi.org/10.1007/s10648-019-09491-3

James, L. E. (2004). Meeting Mr. Farmer versus meeting a farmer: Specific effects of aging on learning proper names. *Psychology and Aging*, *19*(3), 515-522.

https://doi.org/10.1037/0882-7974.19.3.515

Jarosz, A. F., Raden, M. J., & Wiley, J. (2019). Working memory capacity and strategy use on the RAPM. *Intelligence, 77*, Article101387. https://psycnet.apa.org/doi/10.1016/j.intell.2019.101387

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T. Y., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological science*, *20*(8), 963-973. https://doi.org/10.1111/j.1467-9280.2009.02402.x

Karaca, M., Geraci, L., Kurpad, N., Lithander, M. P. G., & Balsis, S. (2023). Low-performing students confidently overpredict their grade performance throughout the semester. *Journal of Intelligence*, *11*(10), 188. https://doi.org/10.3390/jintelligence11100188

Karpicke, J. D., & Aue, W. R. (2015). The testing effect is alive and well with complex materials. *Educational Psychology Review*, *27*(2), 317-326. https://psycnet.apa.org/doi/10.1007/s10648-015-9309-3

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, *331*(6018), 772-775. https://psycnet.apa.org/doi/10.1126/science.1199327

Kavé, G., & Halamish, V. (2015). Doubly blessed: older adults know more vocabulary and know better what they know. *Psychology and aging*, *30*(1), 68-73. https://doi.org/10.1037/a0038669

Keller, N. E., Salvi, C., Leiker, E. K., Gruber, M. J., & Dunsmoor, J. E. (2024). States of epistemic curiosity interfere with memory for incidental scholastic facts. *npj Science of*

Learning, *9*(1), 22. https://doi.org/10.1038/s41539-024-00234-w

Kerr, M. M., & Frese, K. M. (2017). Reading to learn or learning to read? Engaging college students in course readings. *College Teaching*, *65*(1), 28-31. https://doi.org/10.1080/87567555.2016.1222577

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine*, *35*(2), 245-256. https://doi.org/10.1017/s0033291704002892

Kim, A., & Merriam, S. B. (2004). Motivations for learning among older adults in a learning in retirement institute. *Educational Gerontology*, *30*(6), 441-455. https://doi.org/10.1080/03601270490445069

Knobloch, S., Patzig, G., Mende, A.-M., & Hastall, M. (2004). Affective news: Effects of discourse structure in narratives on suspense, curiosity, and enjoyment while reading news and novels. *Communication Research*, *31*(3), 259-287. https://doi.org/10.1177/0093650203261517

Knowlton, B. J., & Castel, A. D. (2022). Memory and reward-based learning: A value-directed learning perspective. *Annual Review of Psychology*, *73*, 25-52. https://doi.org/10.1146/annurev-psych-032921-050951

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131*(2), 147-162. https://doi.org/10.1037/0096-3445.131.2.147

Kuhns, J. M., & Touron, D. R. (2020). Schematic support increases memory strategy use in young and older adults. *Psychology and Aging*, *35*(3), 397-410. https://psycnet.apa.org/doi/10.1037/pag0000433

Lamnin, M., & Chase, C. C. (2019). Developing a thirst for knowledge: How uncertainty in the classroom influences curiosity, affect, learning, and transfer. *Contemporary Educational Psychology*, *59*, Article 101785. https://doi.org/10.1016/j.cedpsych.2019.101785

Lee, H. S., & Ha, H. (2019). Metacognitive judgments of prior material facilitate the learning of new material: The forward effect of metacognitive judgments in inductive learning. *Journal of Educational Psychology*, *111*(7), 1189-1201. https://psycnet.apa.org/doi/10.1037/edu0000339

Lei, S. A. (2010). Intrinsic and extrinsic motivation: Evaluating benefits and drawbacks from college instructors' perspectives. *Journal of Instructional Psychology*, *37*(2), 153-160.

Lujan, H. L., & DiCarlo, S. E. (2006). Too much teaching, not enough learning: What is the solution? *Advanced Physiological Education*, *30*(1), 17-22. https://doi.org/10.1152/advan.00061.2005

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203-208. https://psycnet.apa.org/doi/10.3758/BF03204766

Markman, A., & Ross, B. (2003). Category use and category learning. *Psychological Bulletin*, *129*(4), *592*-613. https://psycnet.apa.org/doi/10.1037/0033-2909.129.4.592

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European journal of cognitive psychology*, *19*(4-5), 494-513. https://doi.org/10.1080/09541440701326154

McGillivray, S., & Castel, A. D. (2017). Older and younger adults' strategic control of metacognitive monitoring: The role of consequences, task experience, and prior knowledge. *Experimental Aging Research*, *43*(3), 233-256. https://doi.org/10.1080/0361073X.2017.1298956

McGillivray, S., Murayama, K., & Castel, A. D. (2015). Thirst for knowledge: The effects of curiosity and interest on memory in younger and older adults. *Psychology and Aging, 30(4)*, 835-841. https://psycnet.apa.org/doi/10.1037/a0039801

McMinn, M. R., Tabor, A., Trihub, B. L., Taylor, L., & Dominguez, A. W. (2009). Reading in graduate school: A survey of doctoral students in clinical psychology. *Training and Education in Professional Psychology, 3(4),* 233-239. https://doi.org/10.1037/a0016405

Merton, R.K. (1965). *On the shoulders of giants. A Shandean postscript* (pp. 218–219). New York: Harcourt Brace & World.

Metcalfe, J., Schwartz, B. L., & Eich, T. S. (2020). Epistemic curiosity and the region of proximal learning. *Current opinion in behavioral sciences*, *35*, 40-47. https://doi.org/10.1016/j.cobeha.2020.06.007

Middlebrooks, C. D., & Castel, A. D. (2018). Self-regulated learning of important information under sequential and simultaneous encoding conditions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44(5)*, 779-792. https://doi.org/10.1037/xlm0000480

Murayama, K., & Kitagami, S. (2014). Consolidation power of extrinsic rewards: Reward cues enhance long-term memory for irrelevant past events. *Journal of Experimental Psychology: General*, *143*(1), 15-20. https://doi.org/10.1037/a0031992

Murayama, K., Kitagami, S., Tanaka, A., & Raw, J. A. (2016). People's naiveté about how extrinsic rewards influence intrinsic motivation. *Motivation Science*, *2*(3), 138-142. https://doi.org/10.1037/mot0000040

Murphy, C., Dehmelt, V., Yonelinas, A. P., Ranganath, C. & Gruber, M. J. (2021). Temporal proximity to the elicitation of curiosity is key for enhancing memory for incidental information. *Learning & Memory*, *28*(2), 34-39. https://doi.org/10.1101/lm.052241.120

Murphy, D. H., Agadzhanyan, K., Whatley, M. C., & Castel, A. D. (2021). Metacognition and fluid intelligence in value-directed remembering. *Metacognition and Learning*, *16*(3), 685-709. https://doi.org/10.1007/s11409-021-09265-9

Murphy, D. H., & Castel, A. D. (2020). Responsible remembering: How metacognition impacts adaptive selective memory. *Zeitschrift für Psychologie*, *228*(4), 301-303. https://psycnet.apa.org/doi/10.1027/2151-2604/a000428

Murphy, D. H., & Castel, A. D. (2021). Responsible remembering and forgetting as contributors to memory for important information. *Memory & Cognition*, *49*(5), 895-911. https://psycnet.apa.org/doi/10.3758/s13421-021-01139-4

Murphy, D. H., Hoover, K. M., Agadzhanyan, K., Kuehn, J. C., & Castel, A. D. (2022). Learning in double time: The effect of lecture video speed on immediate and delayed comprehension. *Applied Cognitive Psychology*, *36*(1), 69-82. https://doi.org/10.1002/acp.3899

Naveh-Benjamin, M. (1991). A comparison of training programs intended for different types of test-anxious students: Further support for an information-processing model. *Journal of Educational Psychology*, *83*(1), 134-139. https://psycnet.apa.org/doi/10.1037/0022-0663.83.1.134

Naveh-Benjamin, M. (2000). Adult-age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170-1187. https://psycnet.apa.org/doi/10.1037/0278-7393.26.5.1170

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 125-173). New York: Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

Nolen, S.B. (1988). Reasons for studying: Motivational orientations and study strategies. *Cognition and Instruction*, *5*, 269-287. https://doi.org/10.1207/s1532690xci0504_2

Norman, D. A., & Bobrow, D. G. (1976). On the analysis of performance operating characteristics. *Psychological Review, 83*(6), 508–510. https://doi.org/10.1037/0033-295X.83.6.508

Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *American Psychological Association*, *144*(7), 710-756. https://doi.org/10.1037/bul0000151

Parr, J. M., & Townsend, M. A. R. (2002). Environments, processes, and mechanisms in peer learning. *International Journal of Educational Research, 37*(5), 403-423. https://doi.org/10.1016/S0883-0355(03)00013-2.

Patterson, T. K., Gao, F., & Knowlton, B. J. (unpublished manuscript). Value-directed remembering in the classroom: Is strategic control of memory related to more selective academic strategies?

Perkins, D. N., & Salomon, G. (2012). Knowledge to go: A motivational and dispositional view of transfer. *Educational Psychologist, 47*(3), 248-258. https://doi.org/10.1080/00461520.2012.693354

Rayner, K., Schotter, E. R., Masson, M. E. J., Potter, M. C., & Treiman, R. (2016). So much to read, so little time: How do we read and can speed reading help? *Psychological Science in the Public Interest*, *17*(1), 4-34. https://doi.org/10.1177/1529100615623267

Reichardt, R., Polner, B., & Simor, P. (2023). Influencing prior knowledge through a short reading impacts curiosity and learning. *Applied Cognitive Psychology*, *37*(2), 458-464.

Rendell, P. G., Castel, A. D., & Craik, F. I. (2005). Memory for proper names in old age: A disproportionate impairment?. *The Quarterly Journal of Experimental Psychology Section A*, *58*(1), 54-71. https://doi.org/10.1080/02724980443000

Renkl, A., Mandl, H., & Gruber, H. (1996). Inert knowledge: Analyses and remedies. *Educational Psychologist*, *31*(2), 115-121. https://doi.org/10.1207/s15326985ep3102_3

Rhodes, M. G. (2016). Judgments of learning. In J. Dunlosky & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory* (pp. 65–80). New York: Oxford University Press.

Robison, M. K., & Unsworth, N. (2017). Working memory capacity, strategic allocation of study time, and value-directed remembering. *Journal of Memory and Language*, *93*, 231-244.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (ed.), *Cognitive Development and the Acquisition of Language*. Academic. pp. 111-144.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54-67.

Salomon, G., & Perkins, D. N. (1989). Rocky roads to transfer: Rethinking mechanisms of a neglected phenomenon. *Educational Psychologist*, *24*(2), 113-142. https://doi.org/10.1207/s15326985ep2402_1

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–776. https://doi.org/10.1037/0012-1649.27.5.763

Schwartz, S. T., Siegel, A. L. M., Eich, T. S., & Castel, A. D. (2023). Value-directed memory selectivity relies on goal-directed knowledge of value structure prior to encoding in young and older adults. *Psychology and Aging*, *38*(1), 30–48. https://doi.org/10.1037/pag0000720

Siegel, A. L., & Castel, A. D. (2018). Memory for important item-location associations in younger and older adults. *Psychology and Aging*, *33*(1), 30-45. https://doi.org/10.1037/pag0000209

Silaj, K. M., Agadzhanyan, K., & Castel, A. D. (2023). Value-directed learning: Schematic reward structure supports value prediction. *Memory & Cognition*, *51*, 1527-1546.

Silaj, K. M., Schwartz, S. T., Castel, A. D., & McDonough, I. M. (2021). Is the future bright or bleak? Assessing past and future outlooks across the adult lifespan. *Gerontology & Geriatric Medicine*, *7*, 1-10. https://doi.org/10.1177/23337214211046080

Silaj, K. M., Schwartz, S. T., Siegel, A. L., & Castel, A. D. (2021). Test anxiety and metacognitive performance in the classroom. *Educational Psychology Review*, *33*(4), 1809-1834. https://doi.org/10.1007/s10648-021-09598-6

Silvia, P. J., & Kashdan, T. B. (2009). Interesting things and curious people: Exploration and engagement as transient states and enduring strengths. *Social and Personality Psychology Compass, 3*(5), 785-797. https://doi.org/10.1111/j.1751-9004.2009.00210.x

Simon-Dack, S. L., Rodriguez, P., & Marcum, G. (2016). Study habits, motives, and strategies of college students with symptoms of ADHD. *Journal of Attention Disorders, 20*(9), 775-781. https://doi.org/10.1177/108705471454336

Sinha, T., Bai, Z., & Cassell, J. (2017). A new theoretical framework for curiosity for learning in social contexts. In *Data Driven Approaches in Digital Education: 12th European Conference on Technology Enhanced Learning, EC-TEL 2017, Tallinn, Estonia, September 12–15, 2017, Proceedings 12* (pp. 254-269). Springer International Publishing.

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 204-221. https://psycnet.apa.org/doi/10.1037/0278-7393.26.1.204

Staff, R. T., Hogan, M. J., & Whalley, L. J. (2014). Aging trajectories of fluid intelligence in late life: The influence of age, practice and childhood IQ on Raven's Progressive Matrices. *Intelligence*, *47*, 194-201. https://doi.org/10.1016/j.intell.2014.09.013

Stare, C. J., Gruber, M. J., Nadel, L., Ranganath, C., & Gómez, R. L. (2018). Curiosity-driven memory enhancement persists over time but does not benefit from post-learning sleep. *Cognitive Neuroscience*, *9*(3-4), 100–115. https://doi.org/10.1080/17588928.2018.1513399

Stasiak A. (2003). The first lady of DNA. *EMBO Reports*, *4*(1), 14. https://doi.org/10.1038/sj.embor.embor723

Stefanidi, A., Ellis, D. M., & Brewer, G. A. (2018). Free recall dynamics in value-directed remembering. *Journal of Memory and Language, 100*, 18-31. https://doi.org/10.1016/j.jml.2017.11.004

Stigler, S. M. (1980). Stigler's law of eponymy. *Transactions of the New York academy of sciences*, *39*(1 Series II), 147-157. https://doi.org/10.1111/j.2164-0947.1980.tb02775.x

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*(2), 244-253. https://doi.org/10.3758/MC.38.2.244

Strous, R. D., & Edelman, M. C. (2007). Eponyms and the Nazi era: Time to remember and time for change. *Medicine and Holocaust*, *9*, 207-214.

Swan, G. E., & Carmelli, D. (1996). Curiosity and mortality in aging adults: A 5-year follow-up of the Western Collaborative Group Study. *Psychology and Aging*, *11*(3), 449-453. https://doi.org/10.1037/0882-7974.11.3.449

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 1024-1037. https://doi.org/10.1037/0278-7393.25.4.1024

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. (2007). Schemas and memory consolidation. *Science, 316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(6), 1020-1034. https://psycnet.apa.org/doi/10.1037/0278-7393.33.6.1020

Unsworth, N. (2016). Working memory capacity and recall from long-term memory: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(1), 50-61. https://doi.org/10.1037/xlm0000148

Wade, S., & Kidd, C (2019). The role of prior knowledge and curiosity in learning. *Psychonomic Bulletin & Review*, *26*, 1377–1387. https://doi.org/10.3758/s13423-019-01598-6

Wang H., Li, J. (2015). How trait curiosity influences psychological well-being and emotional exhaustion: the mediating role of personal initiative. *Personality and Individual Differences, 75*(3), 135-140. https://doi.org/10.1016/j.paid.2014.11.020

Waseem, M., Khan, M., Hussain, N., Giannoudis, P. V., Fischer, J., & Smith, R. M. (2005). Eponyms: errors in clinical practice and scientific writing. *Acta Orthop Belg*, *71*(1), 1-8.

Watkins, M. J., & Bloom, L. C (1999). Selectivity in memory: An exploration of willful control over the remembering process. Unpublished manuscript.

Webb, C. E., & Dennis, N. A. (2019). Differentiating true and false schematic memories in older adults. *The Journals of Gerontology: Series B*, *74*(7), 1111-1120. https://doi.org/10.1093/geronb/gby011

Whatley, M. C., & Castel, A. D. (2022). The role of metacognition and schematic support in younger and older adults' episodic memory. *Memory & Cognition*, *50*(3), 601-616. https://doi.org/10.3758/s13421-021-01169-y

Whitworth J. A. (2007). Should eponyms be abandoned? No. *BMJ (Clinical research ed.)*, *335*(7617), 425. https://doi.org/10.1136/bmj.39308.380567.AD

Zentall, S. S., & Shaw, J. H. (1980). Effects of classroom noise on performance and activity of second-grade hyperactive and control children. *Journal of Educational Psychology*, *72*(6), 830-840. https://psycnet.apa.org/doi/10.1037/0022-0663.72.6.830

Yonelinas, A. P., & Levy, B. J. (2002). Dissociating familiarity from recollection in human recognition memory: Different rates of forgetting over short retention intervals. *Psychonomic Bulletin & Review 9*, 575–58. https://doi.org/10.3758/BF03196315

Zimprich, D., Allemand, M., & Dellenbach, M (2009). Openness to experience, fluid intelligence, and crystallized intelligence in middle-aged and old adults. *Journal of Research in Personality*, *4*(3), 444-454. https://doi.org/10.1016/j.jrp.2009.01.018