Title
The Importance of Selectivity in Memory: The Influence of Value on Monitoring, Learning, and Cognitive Aging

Author
Friedman, Michael Charles

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

The Importance of Selectivity in Memory:

The Influence of Value on Monitoring, Learning, and Cognitive Aging

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Psychology

by

Michael Charles Friedman

2013

ABSTRACT OF THE DISSERTATION

The Importance of Selectivity in Memory:

The Influence of Value on Monitoring, Learning, and Cognitive Aging

by

Michael Charles Friedman

Doctor of Philosophy in Psychology

University of California, Los Angeles, 2013

Professor Alan D. Castel, Chair

Remembering information based on its importance or value can help facilitate a person's quality of life by having them selectively prioritize specific information at the cost of being less likely to remember other, less important, information. My dissertation examines the cognitive mechanisms that are involved with this "value-directed remembering" (VDR), as well as how effectively learning information based on its value or importance can be applied to using memory practically in everyday life. Specifically, I tested whether value influences category learning (Chapter 2; Experiments 1-3) and the impact of value on remembering medication side effects (Chapter 3; Experiments 4-6). Generally speaking, the results from the present research suggest that VDR can occur in these paradigms under certain conditions, and that value influences memory performance in both younger and older adults. The findings show that that goal-directed learning occurs in the presence of high value information and that the selective encoding and/or rehearsal of important information during learning is a cognitively active and effortful process. This research illustrates how the importance or value we place on information

can help us prioritize our learning and, in turn, facilitate our later retrieval for that information in both theoretically driven and practical memory tasks. The implications of this research range from assisting students to improve the quality of their learning and study habits, to helping doctors identify and classify tumors and broken bones, to improving overall quality of life in older adults.

The dissertation of Michael Charles Friedman is approved.

Douglas Bell

Robert Bjork

Barbara Knowlton

Alan Castel, Committee Chair

University of California, Los Angeles

2013

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Thank you first to Dr. Alan Castel. You have not only been an incredibly wonderful advisor to me, but you have also been a fantastic role model. You have been so patient, supportive, and helpful with the editing this dissertation, as well as every other manuscript I have written. It has been a sincere pleasure working with you for so many years, and I look forward to many more to come.

I would also like to thank Dr. Douglas Bell, Dr. Elizabeth Ligon Bjork, Dr. Robert Bjork, and Dr. Barbara Knowlton for your support, comments, and feedback at various stages of my dissertation work. You have all helped shape the way I think about things as a researcher, and I will try to apply what I have learned in my future undertakings to the best of my ability. Thank you to Dr. Daniel Oppenheimer, for teaching me that academics can have fun just like everyone else. Thanks to Dr. Benjamin Storm for teaching me the basics of experimental research and Dr. John Nestojko, who helped me find my post-doctoral fellowship.

Dr. Emily Barkley-Levinson, Dr. Monica Birnbaum, and Dr. Michael Vendetti, your friendship and support over the last five years have been indispensable for me. Thank you all for being such a wonderful group of friends. Courtney Clark, Mikey Garcia, and Veronica Yan, your expertise allowed me to complete my dissertation research on time, so thank you as well. I would also like to acknowledge the work of Victor Sungkhasettee, Monika Holser, Stephanie Knipprath, and Angel Wu. Without the efforts of dedicated research assistants like you, none of this research would be possible.

To my parents, Robert and Sheryl Friedman, thank you for being so supportive and understanding with me, I could not have accomplished this without your care and advice. To my wife Stefanie, you have been the engine pushing me forward for so long and your faith and trust in me has been impeccable, even when I do not have confidence in myself. I love you and I look forward to our next adventure in life together.

Last, I would like to thank Dr. Shannon McGillivray. You have become my older sister.

VITA

2007	B.A., Psychology University of California, Los Angeles Los Angeles, California
2005-2008	Research Assistant Department of Psychology University of California, Los Angeles
2008-2013	Graduate Student Research Assistant Department of Psychology University of California, Los Angeles
2009-2013	Teaching Assistant Department of Psychology University of California, Los Angeles
2009	M.A., Psychology University of California, Los Angeles

PUBLICATIONS

Bui, D. C., Friedman, M. C., McDonough, I. M., & Castel, A. D. (in press). False memory and importance: Can we prioritize without consequence? *Memory & Cognition.*

Friedman, M. C., & Castel, A. D. (in press). Memory, priority encoding, and overcoming high-value proactive interference in younger and older adults. *Aging, Neuropsychology and Cognition.*

Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S., & Link, I. (2013). Selecting valuable information to remember: Age-related differences and similarities in self-regulated learning. *Psychology and Aging, 28*, 232-242.

Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition, 48*, 28-35.

McGillivray, S., Friedman, M. C., & Castel, A. D. (2012). Impact of aging on thinking. In K. Holyoak & R. Morrison (Eds.), *The Oxford Handbook of Thinking and Reasoning* (pp. 650-672). New York: Oxford University Press.

Castel, A. D., McGillivray, S., & Friedman, M. C. (2012). Metamemory and memory efficiency in older adults: Learning about the benefits of priority processing and value-directed remembering. In M. Naveh-Benjamin & N. Ohta (Eds.), *Memory and Aging: Current Issues and Future Directions* (pp. 245-270). Philadelphia, PA: Psychology Press.

Friedman, M. C., & Castel, A.D. (2011). Are we aware of our ability to forget? Metacognitive predictions of directed forgetting. *Memory & Cognition*, 39, 1448-1456.

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review*, 18, 973-978.

CONFERENCE PRESENTATIONS

Friedman, M. C., Castel, A. D., & Noh, S. (November 2012). Goals can enhance inductive learning with interleaved study schedules. *Poster presented at the 53rd annual meeting of the Psychonomic Society*. Minneapolis, MN.

Miyatsu, T., Friedman, M. C., Castel, A. D., & Bjork, R. A. (November 2012). Are high-value items more or less vulnerable to retrieval-induced forgetting? *Poster presented at the 53rd annual meeting of the Psychonomic Society*. Minneapolis, MN.

Friedman, M. C., McGillivray, S., & Castel, A.D. (June 2012). Memory for medication side effects in younger and older adults: The role of perceived importance. *Poster presented at the 17th Annual UCLA Research Conference on Aging*. Los Angeles, CA.

Friedman, M. C., Sungkhasettee, V. W., & Castel, A. D. (November 2011). The misuse of inversion and font size when making judgments of learning. *Poster presented at the 52nd annual meeting of the Psychonomic Society.* Seattle, WA.

Bui, D. C., McDonough, I. M., Friedman, M. C., Castel, A. D., & Myerson, J. (November 2011). False memory and importance: *Can we prioritize without consequence? Poster presented at the 52nd annual meeting of the Psychonomic Society*. Seattle, WA.

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (April 2011). Memory and metamemory for inverted words: The unappreciated effects of desirable difficulties. *Poster presented at the 91st annual convention of the Western Psychological Association.* Los Angeles, CA.

Friedman, M. C., Castel, A. D., Nestojko, J. F., Bjork, R. A., & Halamish, V. (April 2011). Retrieval as a metacognitive modifier? Estimates of forgetting and retrieval-induced forgetting. *Poster presented at the 91st annual convention of the Western Psychological Association.* Los Angeles, CA.

Flores, C. C., McGillivray, S., Friedman, M. C., & Castel, A. D. (April 2011). Memory for prices and better buys in older and younger adults. *Poster presented at the 91st annual convention of the Western Psychological Association*. Los Angeles, CA.

Friedman, M. C., & Castel, A. D. (November 2010). Metacognitive control, serial position effects, and the effectiveness of transfer to self-paced study. *Paper presented at the 51st annual meeting of the Psychonomic Society.* St. Louis, MO.

Cohen, M. S., Castel, A. D., Friedman, M. C., Bjork, R. A. (November 2010). The interactive influences of level of processing and item value on recall and metamemory judgments. *Poster presented at the 51st annual meeting of the Psychonomic Society.* St. Louis, MO.

Friedman, M. C., Castel, A.D., McGillivray, S., & Flores, C.C. (June 2010). Associative memory for money and faces in young and old adults. *Poster presented at the 15th annual UCLA Research Conference on Aging.* Los Angeles, CA.

Friedman, M. C., Castel, A.D., McGillivray, S., & Flores, C.C. (April 2010). Associative memory for money and faces in young and old adults. *Poster presented at the Cognitive Aging Conference 2010*. Atlanta, GA.

Friedman, M. C., & Castel, A. D. (April 2010). Memory, aging, and interference in a value-based encoding task. *Poster presented at the Cognitive Aging Conference 2010*. Atlanta, GA.

Friedman, M. C., & Castel, A. D. (November 2009). Memory and interference in a value-based encoding task. *Poster presented at the 50th Annual Meeting of the Psychonomic Society.* Boston, MA.

Storm, B.C., Friedman, M. C., & Bjork, R.A. (November 2009). On the transfer of prior tests or study events to subsequent study. *Poster presented at the 50th Annual Meeting of the Psychonomic Society.* Boston, MA.

Friedman, M. C., & Castel, A. D. (June 2009). The effect of value on binding and face recognition. *Poster presented at the 14th annual UCLA Research Conference on Aging.* Los Angeles, CA.

Friedman, M. C., & Castel, A.D. (November 2008). Are we aware of our forgetting? Metacognitive ratings of directed forgetting. *Poster presented at the 49th Annual Meeting of the Psychonomic Society.* Chicago, IL.

Friedman, M. C., & Castel, A.D. (May 2008). Are we aware of our forgetting? Metacognitive ratings of directed forgetting. *Poster presented at the 17th Annual Psychology Undergraduate Research Conference*. Los Angeles, CA.

Storm, B.C., Friedman, M. C., Bjork, E.L., & Bjork, R.A. (April 2008). The role of context in retrieval-induced forgetting. *Poster presented at the Western Psychological Association's 88th Annual Convention.* Irvine, CA.

Storm, J.C., Storm, B.C., Friedman, M. C., & Bjork, R.A. (August 2007). When expanding test schedules succeed and fail to enhance learning. *Poster presented at the 115th Annual Convention of the American Psychological Association*. San Francisco, CA.

Friedman, M. C., Storm, B.C., Bjork, E.L, & Bjork, R.A. (May 2007). Is retrieval-induced forgetting context specific? *Poster presented at the 16th Annual Psychology Undergraduate Research Conference.* Los Angeles, CA.

Chapter 1: General Introduction

Why is it that certain memories, when retrieved, are clear and vivid, yet others seem vague and difficult to retrieve? Why is it that we can remember our children's favorite food, but not where we left our keys? These problems illustrate how the importance or value we place on specific memories can influence how well we can retrieve that information later. Our ability to prioritize our thoughts, whether intentional or not, is an astounding achievement of human cognition that we could not survive without. We update information in our memory stores based on that information's current importance or relevance to us. For example, remembering where you parked your car today is much more valuable to know than where you parked your car every day for the first week of this month. Without having such a system in place, we would have a very difficult time interacting with our environment effectively.

While people generally understand that importance or value can influence memory for the better, they do not understand the dynamics required for it to influence memory. In recent court case of fame (i.e., *United States v. Libby, 2006*), former executive branch advisor Lewis "Scooter" Libby was charged and convicted of perjury by a federal jury. He was charged with divulging the identity of a covert Central Intelligence Agency operative to a *New York Times* reporter, which he claimed he did not have memory of disclosing. Jurors could not believe that Mr. Libby would have no memory of such an important conversation. However, what the jurors failed to realize is that valuable memories are only more likely to be memorable if the importance placed on them occurs at encoding. The finding that the value we place on a memory must occur either before or during learning, and cannot occur after the fact, as well as this failure to understand the role of value in memory by third-party judges, has been investigated and replicated empirically (Kassam, Gilbert, Swencionis, & Wilson, 2009; Soderstrom & McCabe, 2011).

Based on the findings of Kassam et al. (2009), Soderstrom and McCabe (2011), as well as observations from everyday life, it is clear that the importance or value placed upon information has an impact on whether we can successfully retrieve that information later. However, it is relatively unknown whether people are aware of those benefits while they are learning valuable (or conversely, not valuable) information. Without such an awareness, it would be difficult for learners to take measures to ensure that we encode valuable information properly, be it with the help of internal devices (e.g., using memory aids like mnemonics) or external devices (e.g., writing a note to remind us later). Additionally, the influence of value in other domains aside from episodic memory is unknown. Can encoding information based on its value assist with teaching a child how to ride a bike or a novice how to hit a golf ball 200 yards? Lastly, does the format or type of value assigned to to-be-learned information (e.g., quantitative, qualitative, etc.) impact memory differentially, and if so, does it change across the adult lifespan? As people age, some of their cognitive faculties may begin to decline, but healthy older adults are quite capable of functioning in society independent of any assistance (Zacks & Hasher, 2006). Could value and selectively remembering only the most important information play a role in successful cognitive aging? The broad goal of this research is to begin to answer these questions.

Value-Directed Remembering (VDR)

One of the first studies to investigate learners' sensitivity to value, and its subsequent impact on recall, used a "selectivity" task (Castel, Benjamin, Craik, & Watkins, 2002; also see Watkins & Bloom, 1999; for a review, see Castel, 2008a). In a selectivity task, younger and older adults studied and were tested on multiple lists of words. Each word on a list was paired with a different numerical value ranging from 1 to 12 (e.g., table 5, uncle 9, apple 2, etc.; see Figure 1). Participants were instructed to remember each word based on its associated value, and if they were able to correctly recall that word at test they would be awarded that associated value to their point total. In this way, certain words were more important for the participants to remember than others. In order to encourage better performance on subsequent study-test lists, participants were given feedback on their point total for the previous list before studying the next list. Importantly, no tangible incentives (e.g., monetary reward) were offered to the participants in exchange for performing well on the task, as such incentives do not motivate participants to improve their behavioral performance (Nilsson, 1987). Recall as a function of point value is shown in Figure 1(b). Participants were able to selectively focus on and later recall the high value words (i.e., words worth 10, 11, and 12 points) better than the words paired with lower point values. Interestingly, older adults had equivalent performance compared to younger adults on the highest, most valuable items despite younger adults remembering more words overall. While this finding does illustrate that older adults recall fewer words, it also shows that they are also more selective than their younger counterparts. Older adults recalled fewer low value words (or forgot more words) than younger adults did, in favor of being able to ensure the recall of the highest value words. Since learners are sensitive to value, and high value information is better remembered than low value information, the finding has been referred to as value-directed remembering (VDR) or the VDR effect and has been replicated several times since its discovery (e.g., Bui, Friedman, McDonough, & Castel, in press; Castel, Farb, & Craik, 2007; Friedman & Castel, 2011; in press; McGillivray & Castel, 2011; Soderstrom & McCabe, 2011).

VDR requires participants to allocate attention and monitor their own memory effectively in order to ensure the successful encoding and retention of important information. VDR may also require the activation of cognitive reward systems during the encoding of high value information to increase the chances of successful consolidation (Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Shohamy & Adcock, 2010). However, the ability to selectively attend to high value information through attention and monitoring comes at the potential cost of learners being less likely to remember less important information. Indeed, certain clinical populations with attentional difficulties such as children with attention-deficit/hyperactivity disorder or older adults with mild forms of Alzheimer's disease are known to have lower performance on this task than age-matched controls (Castel, Balota, & McCabe, 2009; Castel, Lee, Humphreys, & Moore, 2011). Testing at-risk persons with a selectivity task may allow for the early detection of such attentional or memorial disorders.

Several components need to be considered with a selectivity task in order to ensure VDR will occur. These factors are outlined in Table 1. The first two factors deal with making the task sufficiently difficult, such that participants will be unable to remember all of the to-be-remembered words and will be forced to be selective. Factors that increase the difficulty of the selectivity task include using a large number of to-be-learned items in each studied list and also using a rapid presentation rate during study (1-2 s per item). The goal of the above factors is to make it apparent to the learner that they cannot hope to remember all of the items on any given list successfully, which will force them to selectively encode only specific items (i.e., the high value items). The second group of factors encourages learners to have better selectivity in future study-test lists through the use of multiple study-test lists (as opposed to just one) and providing immediate feedback on performance between study-test lists (i.e., telling the participant their point total for that list based on their recall performance). The last group of factors, while not confirmed empirically, may modify the magnitude of the VDR effect in the selectivity task.

These factors include test format (e.g., free recall vs. recognition) and the “scale” of the point values used in the task – categorical (1, 5, or 10) or continuous (1 – 20). For example, the VDR effect is reduced when a recognition test is used instead of a free recall test (Castel et al., 2007). Additionally, the incorporation of a temporal delay between study and test could also modify the magnitude of the VDR effect. In regards to the scale or range of values used in the selectivity task, negative values could be incorporated into the task, resulting in *penalties* to the participant’s point total if the associated words are recalled at test. Thus, the secondary goal of suppression or the forgetting of specific information can also be incorporated into the task, requiring the use of efficient inhibitory control (Castel et al., 2007; Friedman & Castel, 2011; in press). While some of the components described above cannot be modified for the VDR effect to occur (i.e., factors that increase task difficulty), other can be. For example, it is possible to find a reliable VDR effect with using only one study-test list and without providing feedback (Friedman & Castel, 2011). Importantly, the value assigned to an item must be presented shortly before, during, or immediately after studying the information (i.e., as opposed to during retrieval), as learners could not selectively prioritize and encode the high value or important information otherwise (Kassam et al., 2009; Soderstrom & McCabe, 2011).

Remembering information in the real-world requires the selective and/or prioritized encoding of to-be-learned information, which can influence the decisions or choices we make. For example, if a student knows that an upcoming exam is more likely to include material on Topic A than on Topic B, the student should realize that the material on Topic A is more valuable to learn. Given this knowledge, the student would likely devote additional time to studying Topic A and less time to Topic B, as it would increase their chances of getting a higher grade on the exam. The above example explains how the importance or value placed on information can cause us to prioritize that information and control the flow of our own learning.

Metacognition

Metacognition is broadly defined as thinking about your own thinking. Metacognitive research broadly investigates what learners are, or are not, aware of in their thought processes. Specifically, metamemory is one's assessment of his or her own learning and memory, and has been under investigation since the 1960s (Arbuckle & Cuddy, 1969; Flavell, 1979; Hart, 1965; 1966; 1967). In order to gauge a participant's accuracy in a metamemory task, their predicted performance is compared to their actual performance in one of two ways. Calibration, or absolute accuracy, is the overall difference between predicted and actual performance, and refers to the degree to which people over- or underestimate their abilities. Resolution, or relative accuracy, is when a person's metacognitive predictions accurately assess actual performance on an item-by-item basis, and is assessed via a Goodman-Kruskal gamma correlation (Nelson, 1984). The correlation represents how well a person's predicted performance can discriminate between items they will recall later compared to items they will not recall later. Based on work by Nelson and Narens (1990), metacognitive research can be broadly broken down into two categories: *monitoring* and *control*.

Metacognitive monitoring is a learner's assessments, theories, or understanding of how they think their memory works (Nelson & Narens, 1990). Metacognitive monitoring can be measured with predictions or estimates at encoding given by the participant on a numerical scale (e.g., judgments of learning, JOLs). A couple of theoretical explanations have been put forward to explain what factors influence monitoring. Fluency, or ease of processing, refers to the idea that information that is easy to learn or easy to retrieve from memory is more likely to be retrieved correctly later (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin & Bjork, 1996; Benjamin, Bjork, & Schwartz, 1998). While the logic of fluency seems valid, it functions as a heuristic (easily learned, easily remembered; see Miele, Finn, & Molden, 2011) - people fail to anticipate that the effort associated with retrieval can also increase the probability of successful retrieval later (Benjamin et al., 1998; Zechmeister & Shaughnessy, 1980). That is, if people can overcome their initial difficulties during a retrieval event, subsequent retrieval of that information will be that much easier.

Koriat (1997) posits that monitoring predictions are based on the weighting of cues. This cue-utilization approach explains that innate qualities of the information studied (i.e., intrinsic cues) serve as the basis for JOLs and typically outweigh experimental manipulations or conditions that facilitate the encoding and retrieval of memories (i.e., extrinsic cues). Using these two frameworks, it becomes apparent why there is considerable metacognitive research that focuses on the inaccuracies of metacognitive monitoring (Castel, 2008b; Castel, McCabe, & Roediger, 2007; Friedman & Castel, 2011; Koriat & Bjork, 2005; Koriat, Bjork, Sheffer, & Bar, 2004; Koriat, Sheffer, & Ma'ayan, 2002; Rhodes & Castel, 2008; 2009; Sungkhasettee, Friedman, & Castel, 2011; etc.). For example, Castel (2008b) illustrates that people typically fail to predict serial position effects, and are only capable of predicting primacy and recency effects under specific experimental conditions designed to make people more aware of serial position during study. Such work illustrates that our ability to monitor our learning accurately can be extremely biased and easily misled in many situations.

Metacognitive control is a learner's ability, through behavioral action (or continuation, or termination of action), to increase the strength of a given memory in order to increase its probability of successful retrieval later (Nelson & Narens, 1990). Broadly, research on metacognitive control investigates what strategies or actions people use when they know that they do not know specific to-be-learned information or when they know that they do know that information. Metacognitive control is evaluated through behavioral actions or choices that typically impact learning (e.g., self-paced study time, restudy choices, etc.). Two competing theories have been put forth to explain how people exert control. According to the region of proximal learning (Metcalfe, 2002; Metcalfe & Kornell, 2005), people opt to restudy information that is neither too easy nor difficult for them to learn, and instead choose to focus on information that they are on the verge of mastering (i.e., the *region* of proximal learning). In contrast, the agenda-based regulation model advances the idea that people control their learning based on their goals or agendas, which may not always coincide with trying to remember as much information as possible (Ariel, Dunlosky, & Bailey, 2009). For example, the goal one would have in a selectivity task would be to maximize their point total at test, and should therefore attempt to restudy or spend more time learning the most valuable information (Ariel et al., 2009).

Both metacognitive monitoring and control have implications for real-world learning and instruction. Bjork (1994) puts forth that intuition and "normal" practice are poor guides for optimal training, as both learners and instructors alike are easily fooled by fluency and other factors that may bias one's ability to accurately monitor their progress. By continually challenging the learner, the instructor can help that person reach their full potential, despite the benefit being misperceived as small or even counterproductive to progress (a concept referred to as desirable difficulties; Bjork, 1994; Bjork & Bjork, 2011). Additionally, Hacker and colleagues illustrated the importance of metacognition in education (Hacker, Bol, Horgan, & Rakow, 2000). In the aforementioned study, a psychology class was asked to estimate their score on an exam both before and after taking the test. While the students who scored the highest on the exam underestimated their actual performance, students that scored the lowest on the exam overestimated their actual performance. The results and discussion illuminated two factors that are likely harming the lowest performing students. The first factor is that the students who are doing the worst in the course are also the worst at monitoring their performance, meaning that they likely misunderstand the course material. The second factor is that this poor monitoring can lead to an inflated sense of confidence in those students which may lead to maladaptive study behaviors and cause them to stop studying prematurely (i.e., poor metacognitive control). The above study and findings highlight the critical nature of metacognition in learning, as accurate monitoring can help people make the best study decisions possible (a reflection of good metacognitive control), and improve what would been sub-par performance.

Cognitive Changes Across the Adult Lifespan

As people age and mature, their cognitive abilities change as well. While common stereotypes surrounding older adulthood would have us believe that advanced age causes declines in every cognitive faculty, this view is far from true (but see Salthouse, 1996). A review of cognitive function in later adulthood shows that a majority of the cognitive declines found in old age occur in episodic memory systems only, while healthy older adults have relatively intact short-term and semantic memory systems compared to younger controls (Zacks & Hasher, 2006). Despite declines in some areas, older adults may be able to draw upon a lifetime of experience and knowledge (i.e., schematic support) in order to optimize the encoding and retrieval of information (Castel, 2005; 2007; Miller, 2003; Morrow, Leirer, Altiteri, & Fitzsimmons, 1994). Additionally, other types of learning such as category induction and the benefits associated with certain study strategies found in younger adults (e.g., interleaved schedules of learning) may also remain intact across the adult lifespan (Wahlheim, Dunlosky, & Jacoby; 2011; Kornell, Castel, Eich, & Bjork, 2010).

In regards to the metacognitive abilities of older adults, findings from empirical works regarding the efficiency of older adults' monitoring and control is relatively mixed (for a recent review, see Castel, McGillivray, & Friedman, 2012). While healthy older adults are generally aware of their cognitive declines and are capable of accurately monitoring their learning (e.g., Halamish, McGillivray, & Castel, 2011; Hertzog & Hultsch, 2000; Hertzog, Kidder, Powell-Moman & Dunlosky 2002), other work shows that older adults can be overconfident in predicting their later recall performance (Connor, Dunlosky, & Hertzog, 1997). However, while this work is mixed with regards to metacognitive monitoring, older adults are as capable of exerting control over their own learning as well as younger adults can. Work by Dunlosky and Conner (1997) found that both younger and older adults chose to restudy information they deemed more difficult to remember (based upon reported JOLs) than information they deemed easier to remember. Additionally, older adults are able to maximize their performance on a selectivity task as well as younger adults can. In a unique task in which younger and older adults studied low and high value words, participants had to "bet" on whether or not they could recall those words later. Importantly, older adults achieved the same level of performance (in terms of point totals) as younger adults, further supporting the idea that older adults can effectively control their learning (McGillivray & Castel, 2011).

Given that the original selectivity task compared older to younger adult performance (Castel et al., 2002), there has been extensive amount of VDR research on cognitive aging. The importance of this work is in investigating what specific mechanisms associated with VDR and cognitive function change across the lifespan (Castel, 2008a; Castel, Humphreys, Lee, Galvan, Balota, & McCabe, 2011). Figure 2 shows the general trends of VDR across the human lifespan. From this figure, only the most extreme age groups (i.e., children, adolescents, and the older-old groups) show reduced VDR effects, or reduced sensitivity to value, on the selectivity task. Importantly though, older adults can do well on selectivity tasks despite known associative deficits (i.e., the reduced ability to pair information together in memory, such as word-value associations; see Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008). However, recent work has found that older adults are sensitive to value initially (and show typical VDR effects), but have severe declines in selectivity performance (and no longer show VDR effects) when established values are suddenly changed (Friedman & Castel, in press). Specifically, when the same word-value pairs were repeatedly studied and tested across multiple lists, but then the values paired with the words were suddenly switched on later lists, both younger and older adults showed immediate drops in recall performance, point totals, and VDR effects. Interestingly, younger adults were able to fully recover from the impairment caused by the value switch, while older adults could not; suggesting that the ability to update the value or importance associated with specific information is impaired in later adulthood.

Inhibitory deficits in older adulthood may play a role in the reduced capacity for selectivity following either a switch in the values associated with specific words (Friedman & Castel, in press), or when negative values are incorporated into the selectivity paradigm (Castel et al., 2007). Inhibitory deficits relate to a person's cognitive inability to effectively suppress or inhibit competing, interfering, or unimportant information, which has been found in older adulthood (Hasher & Zacks, 1988). The necessity of an optimal inhibitory system is obvious for the selectivity task. As previously mentioned, the accidental recall of words paired with negative values places a penalty on participants' point totals in a selectivity task (equivalent to the numerical value it was associated with). Castel et al. (2007; Experiment 2) shows that, while older adults are capable of suppressing the recall of words associated with negative values for a free recall test and show a typical VDR effect in a selectivity task with negative values, older adults identify disproportionately more words paired with negative values than younger adults do on a subsequent recognition task, thus illustrating age-related inhibitory deficits. Overall, the selectivity task is an ideal paradigm with which to investigate questions about cognitive aging, as it incorporates so many different cognitive components and systems.

Current Studies and Goals of Dissertation

The broad aim of my dissertation research is to explore the cognitive and metacognitive mechanisms associated with learners' sensitivity to value, whether VDR can facilitate the learning of categories, and whether or not younger and older adults are equally sensitive to value in a relatively practical selectivity task. Specifically, can attaching a value to to-be-learned information help guide metacognitive judgments? Can selective encoding based on point values enhance category learning or other types of learning in a similar fashion to how it can impact episodic recall or can VDR only occur in certain contexts of learning? Finally, how sensitive are younger and older adults to value when the type of value placed on information (either objectively defined by the experimenter or subjectively rated by the learner) is manipulated in a practical task that assesses memory, and are younger and older adults more sensitive to one type of value over another?

Within the following sections, two series of value-related experiments are reported in an attempt to address the questions described above. Chapter 2 (Experiments 1, 2, and 3) explores the mechanisms that take place during the selective prioritization and encoding processes associated with VDR. Specifically, can the effect of VDR (which is thought to be an active, effortful process) be applied towards successfully learning new categories based on their key features (i.e., inductive learning, a process thought to be relatively passive and sub-conscious; see Goldstone, Landy, & Brunel, 2011)? If so, are people able to accurately predict this benefit during learning and retrieval (Kornell & Bjork, 2008)? If so, the findings reported in Chapter 2 would have the potential to assist people with learning valuable and potentially life-saving information (e.g., identifying and classifying broken bones, tumors, edible plants, poisonous/dangerous animals in the wild, etc.). Chapter 2 (Experiments 2 and 3) also investigates whether value can function as an informative metacognitive cue, through the use of metacognitive monitoring measures (i.e., JOLs and confidence ratings). It is likely that people view value or importance as a goal-oriented cue (Ariel et al., 2009) and can use it to accurately predict later memory performance (Koriat, 1997; but see Koriat et al., 2004; Experiment 6B; Nilsson, 1987). If value does impact both metamemory and actual memory, then VDR could provide students with a means to accurately assess their own learning without being biased by other factors such as fluency or ease of processing.

Chapter 3 (Experiments 4, 5, and 6) examines whether or not the impact of value on memory can change depending on the type of value assigned to information in a practical memory task (i.e., memory for medication side effects). Specifically, Chapter 3 examines whether objective values (e.g., points) or more subjective values (e.g., "I feel this information is important") influence memory differently, as well as whether this influence changes across the adult lifespan. The findings from this study may help better our understanding of how we should convey important or valuable information to others (and the way to best frame importance to others), and how the type of value or importance used may need to be modified based on the target audience receiving that information (i.e., younger or older adults). Such knowledge would inform instructors and professionals alike and help their respective audiences successfully retain the critical information they are trying to convey. Chapter 3 also has implications for the distribution of medical information to patients, especially for older adults, and it can assist with ensuring that valuable information is not forgotten, and treatment plans are successfully adhered to.

Finally, Chapter 4 summarizes findings from the research in this dissertation, describes potential future directions for each of the studies, and provides a general conclusion for how the information gained from this work may impact student learning, monitoring, and successful aging.

Chapter 2: An Exploration of the Mechanisms Associated with Value-Directed Remembering in Inductive Learning

Based on prior research investigating value-directed remembering (VDR), encoding information based on its perceived value or importance can increase a learner's probability of retrieving that valuable information later (Castel et al., 2012). Importantly, the increased recall of high-value information illustrates that learning is impacted by value or importance (Castel et al., 2002; Castel, 2008a). To date, VDR and selectivity (i.e., selectively choosing to remember specific information identified as important or valuable) have only been investigated in the context of episodic memory – memory for the recall of prior events. In the context of episodic memory, remembering an event accurately would be considered successful if the retrieved information matches what was learned during the event in question.

However, not all learning and memory tasks require perfect retrieval in order for the participant to be considered successful in that task. For example, older adults, despite their cognitive declines, can recall the main or most important points of a story or event that was told to them, despite lacking perfect memory for the specific details of that story or event (Adams, 1991; Adams, Smith, Nyquist, & Perlmutter, 1997). Another example of successful retrieval that does not require flawless encoding of information would be learning a new category of information, or an induction task (e.g., Kornell & Bjork, 2008). Although attending to details is important, it is not absolutely necessary in category learning. It is possible to excel in an induction task without having specific memories of any of the previously studied exemplars. Furthermore, the ability to rehearse materials in category learning tasks is relatively difficult compared to rehearsing words for a memory test (e.g., the rote rehearsal of non-verbal stimuli such as paintings is difficult to do successfully; Kornell & Bjork, 2008; Kornell et al., 2010).

Based on the above argument, some have argued that inductive learning is a relatively passive or sub-conscious process that learners do not need to put effort into in order to be successful at learning new categories (Goldstone, Landy, & Brunel, 2011). Therefore, it is worthwhile to investigate whether the selective encoding processes associated with VDR would have any influence on category induction, as the encoding mechanisms associated with VDR are arguably effortful in nature. The aim of this chapter is to explore whether encoding information based on value or importance can facilitate other types of learning beyond episodic memory, focusing on the domain of category learning.

Broadly, induction tests allow researchers to examine how well people can learn novel categories, and investigate manipulations may make that may facilitate that process. Several recent works have shown that the benefits of interleaved practiced (i.e., mixing the study of several categories together) outweigh those of blocked practice (i.e., studying the same category sequentially before studying another category), even though people actually believe the opposite (Birnbaum, Kornell, Bjork, & Bjork, 2013; Kornell & Bjork, 2008; Kornell et al., 2010; Wahlheim, et al., 2011). The fact that the advantage of interleaved study in episodic memory transfers to other types of learning such as category induction gives support to the idea that other operations known to influence episodic recall (e.g., VDR) may also convey similar benefits to inductive learning. This idea can also be supported by our everyday interactions with our environment. Generally, people are more sensitive to important or valuable information around them (e.g., listening for a nearby car horn may prevent a traffic collision). However, does paying attention to valuable information in the environment mean that people will more effective at learning categories classified as more valuable?

There are reasons to believe the value or importance attached to information, while beneficial for episodic memory, may not influence category learning. To-be-learned categories can either be rule-based or information integration-based in nature. For rule-based categories, learners try to actively check rules of category exemplars in order to recognize what differentiates one category from another. The processing associated with rule-based category induction is therefore thought to be cognitively effortful in nature (as learners are guessing and checking rules while studying the categories; see Ashby, Maddox, & Bohil, 2002). For example, a concrete rule for identifying whether or not a given number is a member of the even category is to divide that number by 2 and check whether or not the answer is a whole number, and therefore learners can apply that rule when learning new exemplars which may or may not be members of the even category. In contrast, information integration category learning does not require any active rule-based checking by the learners to discover the features critical for category membership; in fact, features in such categories cannot typically be verbalized precisely. Therefore, integration-based categories are thought to be relatively more passive in terms of the cognitive effort needed to learn the categories (i.e., requires little, if any, working memory). For example, looking at samples of work by a given artist may not lead learners to identify any explicit rules associated with that artist's style, but may still help lead the learner to identify new works that were created by that artist. Since the current chapter aims to explore the relationship between VDR and information integration category learning, it seems possible that the effortful and selective encoding associated with VDR would impact the relatively passive processes of category induction associated with information integration.

Conversely, there is reason to believe that VDR may impede category induction. Work on verbal overshadowing has shown that specific types of effortful processing (e.g., processing that requires the retrieval of superfluous information) interferes with the processes at encoding and impairs subsequent recognition compared with material that did not receive such processing (Dodson, Johnson, & Schooler, 1997). Thus, it may be possible that the selective processing associated with high value information may impede category learning, and categories that are less valuable to know would be counter-intuitively easier to learn and identify later. The current experiments aim to identify which of these lines of reasoning is correct. Specifically, if the cognitively effortful processes associated with selectively encoding high-value information does benefit category learning (i.e., a task considered to be cognitively less effortful, Goldstone et al., 2011), then the finding would support the idea that attentional resources or effort do play a role in category induction. However, if the effortful processes associated with value-based encoding have no effect (or a detrimental effect) on category learning, then attentional resources or effort may not be a critical component of information integration category learning (or may even impede the process of learning categories).

Surprisingly, only a little work has been done exploring the role of value or importance in monitoring learning (Ariel et al., 2009; Friedman & Castel, 2011; Gerlach, 2008; Koriat et al., 2006; Soderstrom & McCabe, 2011), which is surprising given the role value or importance can have in memory function (Castel et al., 2012). Therefore, a secondary goal of this chapter is to explore how value can impact the role of monitoring, both during learning and during retrieval. Generally, people are sensitive to value, in the sense that value will cause people to give higher judgments of learning (JOLs) for valuable or important to-be-remembered information than less valuable information (Friedman & Castel, 2011; Gerlach 2008). However, value does not always have an impact on learning, even if learners or observers expect it to (Kassam et al., 2009; Soderstrom & McCabe, 2011). From this new body of work we can conclude that people are sensitive to value in terms of monitoring, yet are unaware of the dynamics in which value or importance may impact memory, since it is not always clear whether value will function as a salient cue for memory (i.e., a factor that will actually impact memory performance). This may apply especially to the current study, as the impact of value on category learning is currently unknown. Therefore, the latter portion of this chapter focuses on learners' ability to monitor their own performance during learning (Experiment 2) and during retrieval (Experiment 3).

The incorrect monitoring of cues that may benefit learning, however, is not restricted only to value. In the context of category learning, E.Z. Rothkopf said "spacing is the friend of recall, but the enemy of induction" (cited in Kornell & Bjork, 2008). After several studies disproved this claim (e.g., Kornell & Bjork, 2008; Kornell et al., 2010; Vlach, Sandhofer, & Kornell, 2008), most researchers currently consider spaced learning (i.e., learning events of the same category being temporally spaced apart from one another) to be better than massed learning (i.e., studying all exemplars of a given category simultaneously or sequentially before studying a different category) for successful category induction. However, what actually benefits category learning is not the temporal spacing of materials, but the opportunity to compare and discriminate categories from one another (Birnbaum et al., 2013; Kang & Pashler, 2012). Experimental procedures that allow for active comparison among different categories, known as interleaved study, allow learners to juxtapose different categories and highlight differences between categories better than does blocked learning (studying all exemplars of one category together before learning the next category). Recent work on category learning supports this discriminative contrast hypothesis as an explanation of how interleaved schedules of learning benefit category induction (Birnbaum et al., 2013; Kang & Pashler, 2012; Zulkiply & Burt, in press).

Crucial to the current work is the question of whether and to what extent VDR would be observed within blocked versus interleaved learning. For blocked study, learners may have better or exclusive focus on high-value categories during learning, and may preferentially rehearse those high-value categories during the study of low-value categories. However, learners may not have to attend to value under blocked learning conditions because they will not have to selectively prioritize during a block of valuable information (as everything would be important to learn), allowing the learners to better learn categories without being distracted by value. Conversely, VDR could be more pronounced in interleaved schedules of learning because, as stated earlier, interleaved study allows learners to recognize differences between categories more easily, and may cause learners to selectively focus on the differences associated with valuable categories. Recent work on episodic memory found that the influence of value on interleaved schedules of learning is more potent (i.e., learners show better recall for high than for medium or low-value conditions) than it is for blocked schedules (Bui et al., in press). Therefore, the current experiments may help provide further insight as to why value-directed remembering may or may not be present under certain contexts. Specifically, by exploring whether the influence of value is more robust during either blocked or interleaved presentation, we can better understand what parameters are essential for optimizing value-based encoding.

The basic experimental paradigm in this chapter was that participants studied several paintings by various artists with the goal of learning their "style" (and being able to successfully identify new paintings by those artists later), while manipulating two variables. The first variable manipulated was the study schedule in which the paintings by a given artist were shown – either blocked (i.e., all paintings by that artist shown together before the next artist's paintings) or interleaved (i.e., paintings by that artist shown in an order mixed with paintings by other artists).

The second variable manipulated was the point value assigned to a given artist – 1 point, 5 points, or 10 points. Participants were given instructions to pay attention to artists' paintings (i.e., categories) associated with higher values, thus making 10 point categories more valuable to learn than 1 or 5 point categories. After studying all the paintings associated with each of the to-be-learned artists, participants were given an induction test in which they were shown new paintings by the artists they had originally studied, and had to identify which of those artists painted the work. Experiments 2 and 3 also investigated the participants' ability to monitor their own learning in this paradigm, by assessing their JOLs during learning (Experiment 2) and by assessing their confidence at test for their responses (Experiment 3).

Experiment 1

To investigate whether or not VDR can impact category learning, an established and reliable induction task was used. By using the methodology outlined in Kornell and Bjork (2008), this experiment investigates the mechanisms associated with selective encoding in category induction. Specifically, if the sub-conscious processes associated with inductive learning (Goldstone et al., 2011) cannot be influenced by the active processes associated with the effortful encoding of high-value information (Castel et al., 2002), then VDR effects should not be found in this task. However, if the additional effort associated with selective encoding interferes with category induction as it does with recognition (Dodson et al., 1997), then *better* category induction should be found in the low value categories relative to the high value categories.

Methods

Participants. Seventy-two undergraduates from the University of California, Los Angeles were recruited for this task (19 males, mean age = 20.4). All participants self-reported they had little to no background in the artists that were studied prior to participating in the experiment. All participants were compensated with course credit for their participation.

Materials. All of the materials used in Experiment 1were identical to those of Kornell & Bjork (2008; Experiment 1A), with the exception of the incorporated selectivity components (i.e., point values). The materials were 10 paintings by each of 12 artists (Braque, Cross, Hawkins, Lewis, Mylrea, Pessani, Schlorff, Seurat, Stratulat, Wexler, and YieMei) for a total of 120 paintings presented during study and testing. Six paintings by each artist were presented during the study phase, and four during the testing phase. All of the paintings were of landscapes or skyscapes (for an example, see Figure 3). The paintings were cropped to remove any names or signatures. Additionally, each artists' paintings were paired with a value (1, 5, or 10 points), such that four artists had all six of their studied paintings paired with a value of 1 point, another four artists paired with a value 5 points, and the remaining four artists paired with a value of 10 points. The value assigned to each artist's paintings was counterbalanced across participants.

Procedure. The procedure was adapted from Kornell & Bjork (2008; Experiment 1A). Before beginning, participants were instructed about the nature of the study and test phases, as well as the selectivity directions. Participants were told that, unlike a "normal" memory test, they would need to learn the styles of the 12 artists by studying a set of their paintings. Additionally, some artists' styles would be more valuable to remember than others, as indicated by the point values paired with each artist's name. If, during the test phase, the participant could correctly identify the artist of a new painting that they did not study during the initial learning phase, they would be awarded its associated point value which would be added to their total. Participants were told their goal was to maximize their point total during the test phase by trying to remember the artists' styles as best they could, but particularly those of artists paired with the highest values (i.e., 10 points).

After participants went through and understood the instructions, they were then shown 6 paintings by each of the 12 artists, totaling to 72 paintings. A given painting was shown on the screen for 3 s, with the artist's last name, the associated point value for that artist, and the painting directly beneath (see Figure 3). The paintings from six of the artists were presented in a consecutive order (blocked), while the paintings from the other six artists were presented intermingled with one another (interleaved). Across participants, each artist was assigned to be presented in a blocked or interleaved order equally. Groups of six paintings, presented in either blocked presentation (B) or interleaved presentation (I) were displayed to participants in the order BIIB BIIB BIIB.

After the last painting was shown, participants engaged in a 15 s distractor task, during which the participant was instructed to count backwards by 3s from a random three-digit number. The test phase began immediately following this task. Participants were instructed that they would see new paintings by the 12 same artists, and it would be their job to identify which artist painted each work. If the participant correctly identified the artist, they were awarded the associated point value for that painting (i.e., 1, 5, or 10 points), which was added to their running total. However, participants were told that no point values would be shown during the test phase. During the test phase, a painting appeared on the screen with a list of the 12 artist's names (as well as a "Don't Know" response), which prompted the participant to verbally respond to the experimenter which artist they thought created the work (see Figure 4). Participants were encouraged to guess if they did not know the correct response and had as much time as they needed to make their guess. No feedback was given by the program or the experimenter as to whether or not the participant's responses were correct. Half of the participants, instead of being given a test with new paintings, were given a memory test with old paintings instead. The procedure was exactly the same as that of the test with the new paintings, except that the paintings participants saw in the memory test were a selection of 48 paintings from the learning phase (4 randomly selected paintings from each of the 12 artists).

There were 48 test trials divided into four blocks of 12 paintings. Each block consisted of one painting by each of the 12 artists, presented in a random order. After the test phase, the terms blocked and interleaved were explained to the participant, and they were asked which study schedule they thought was better for this task (i.e., blocked, interleaved, or about the same). Additionally, participants were asked whether or not they felt that the values shown during the learning phase helped them on this task, and also to identify which values were paired with which artist's paintings from the learning phase. Lastly, participants were asked if they were familiar with any of the artists from the task before the learning phase began.

Results and Discussion

Induction Test Performance

Figure 5 displays the data for the participants assigned to take the induction test. The top panel of Figure 5 (A) shows the participants' performance on the induction test (i.e., the proportion of new paintings the participants correctly categorized). There appears to be no overall effect of VDR on category learning, however a small effect of VDR was found when categories were studied in an interleaved pattern, but not when studied in a blocked pattern. A 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures analysis of variance (ANOVA) found a main effect of study schedule, $F(1,35) = 56.92$, $MSE = 0.12$, $p < .001$, $\eta_p^2 = .62$ such that interleaved study ($M = .57$, $SD = .24$) lead to better category learning than blocked study (*M* = .22, *SD* = .13). A marginal study schedule-point value interaction was also found, $F(2,70) = 2.05$, $MSE = 0.04$, $p = .14$, $\eta_p^2 = .06$, such that that was no apparent influence of value on blocked study ($F < 1$), but there was a slight influence of value on interleaved study, $F(2,70) = 2.72$, $MSE = 0.05$, $p = .07$, $\eta_p^2 = .07$. This marginal effect for interleaved study was in the hypothesized direction such that 10 point categories (*M* = .63, *SD* = .26) were correctly identified more often than 1 point categories (*M* = .52, *SD* = .32), $t(35) = 2.74$, $p = .01$, $d = .38$.

Panel B of Figure 5 displays the summed average number of incorrect responses the participants made during the induction test. For a given incorrect response on the test, the associated study schedule-point value condition associated with that incorrect response (i.e., artist) was coded, and then summed, for each participant. Two participants' incorrect response data were not collected, and were therefore not included in this analysis. Overall, it appears that value had an influence on the number of incorrect guesses during the induction test for blocked study, but not for interleaved study. The 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a main effect of value, $F(2,66) = 3.06$, $MSE = 6.30$, $p = .05$, $\eta_p^2 = .09$, such that more incorrect guesses were given by participants for artists associated with 10 point categories (*M* = 4.96, *SD* = 1.98) than with 1 point categories (*M* = 3.90, *SD* = 1.92), $t(33) = 2.12$, $p = .04$, $d = .54$. A marginal study schedule by point value interaction was also found, $F(2,66) = 2.02$, $MSE = 4.09$, $p = .14$, $\eta_p^2 = .06$. Subsequent ANOVAs revealed that point value had no apparent influence on incorrect guesses for interleaved study ($F< 1$), however value did influence the number of incorrect guesses participants made for blocked study, $F(2, 66) = 5.21$, $MSE = 4.92$, $p< .01$, $\eta_p^2 = .14$, such that 10 point categories (*M* = 5.15, *SD* = 2.80) had significantly more incorrect guesses than 1 point categories (*M* = 3.41, *SD* = 2.09), $t(33) = 2.92$, $p< .01$, $d = .70$.

Panel C of Figure 5 displays the proportion of values correctly assigned to the artists by participants after the induction test, during the post-test questionnaire. The proportion of values correctly assigned appeared to increase across point value conditions for interleaved study, yet decrease across point values for blocked study. The associated 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a marginal effect of study schedule, $F(1,35) = 2.79$, $MSE = 0.14$, $p = .10$, $\eta_p^2 = .07$, such that interleaved study ($M = .52$, $SD = .29$) lead to better memory for the artist-point value associations than blocked study ($M = .44$, $SD = .27$). A study schedule by point value interaction was also found $F(2,70) = 6.06$, $MSE = 0.09$, $p<.01$, $\eta_p^2 = .15$, such that there was a marginal effect of value for interleaved study $F(2,70) = 1.99$, $MSE = 0.11$, $p = .14$, $\eta_p^2 = .05$, and a strong effect of value for blocked study, $F(2,70) = 4.62$, $MSE = 0.40$, $p = .01$, $\eta_p^2 = .12$. For interleaved study, participants remembered the values of artists associated with 10 points ($M = .60$, $SD = .39$) better than artists associated with a value of 1 point ($M = .44$, $SD = .35$), $t(35) = 1.99$, $p = .05$, $d = .43$. However, for blocked study, the associations between artists and 10 points ($M = .32$, $SD = .32$) were remembered worse than artists associated with a value of 1 point ($M = .51$, $SD = .39$), $t(35) = 2.91$, $p< .01$, $d = .53$.

Following the test phase of the experiment, participants were asked two post-test questions: which study schedule they thought was the most beneficial for their learning, and which point value they thought was most beneficial for their learning. The results are plotted in Figure 6. The top panel of Figure 6 (A) displays the participants' responses to the study schedule question, based upon which schedule the participant judged to be the most effective and what schedule was actually the most effective for them on the induction test. Overall, a majority of participants felt that blocked study was more effective than interleaved learning (81%), yet most participants showed the opposite result for their actual performance on the induction test, in that interleaved study led to better category learning than blocked study (89%).

The lower panel of Figure 6 (B) shows the perceived and actual effectiveness of the different point value conditions. Overall, participants thought that the 10-point categories (53%) were the best learned, or that the point values did not influence their learning of the categories (42%). Interestingly, a small majority of participants actually learned the 1-point categories better (36%) than the 5 point (22%) or 10 point (28%) categories. It should be noted for this analysis, however, that if a participant had equivalent levels of performance in two or more of the point value conditions, their actual performance was scored in the same/did not matter classification, which made up a small portion of the data (14%). Therefore, even though a small majority of participants actually did perform best in the 1-point condition, this result should not be weighted too heavily.

Memory Test Performance

Figure 7 displays the data for the participants assigned to take the memory test. The top panel of Figure 7 (A), show the participants' performance on the memory test (i.e., the proportion of paintings initially studied that the participants correctly categorized). Overall, it appears that interleaved study, like on the induction test, had a positive benefit on helping participants remember the artists' paintings, as did point value. A 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a significant main effect of study schedule, $F(1,35) = 93.32$, $MSE = 0.09$, $p < .001$, $\eta_p^2 = .73$, such that interleaved study ($M = .71$, $SD = .20$) lead to better memory performance than blocked study ($M = .31$, $SD = .23$). A main effect of value was also found, $F(2,70) = 3.34$, $MSE = 0.05$, $p = .04$, $\eta_p^2 = .09$, such that both 10 point categories (*M* = .55, *SD* = .22) and 5 point categories (*M* = .53, *SD* = .19) were remembered better than 1 point categories (*M* = .46, *SD* = .23; *p*s = or < .05).

Panel B of Figure 7 displays the summed average number of incorrect responses participants made during the memory test. One participant's incorrect response data was not collected, and was therefore not included in this analysis. Other than a marginal effect of value, no other significant effects were found. The associated 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a marginally significant effect of value, $F(2,68) = 1.95$, $MSE = 4.37$, $p = .15$, $\eta_p^2 = .05$, such that participants gave fewer incorrect responses associated with artists that were paired with a value of 1 point (*M* = 3.27, *SD* = 1.85) compared to artists paired with a value of 5 points (*M* = 3.96, *SD* = 1.71), $t(34) = 2.03$, $p = .05$, $d = .39$. All other effects within this ANOVA were non-significant.

Panel C of Figure 7 displays the proportion of values correctly assigned to the artists by participants after the memory test, during the post-test questionnaire. One participant's response data for this set of questions was not collected, and was therefore not included in this analysis. Only a benefit of interleaved study was found in remembering the point values. The 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a main effect of study schedule $F(1, 34) = 10.18$, $MSE = 0.13$, $p < .01$, $\eta_p^2 = .23$, such that participants better remembered the point values associated with categories presented in an interleaved fashion (*M* = .55, *SD* = .21) than those in a blocked fashion (*M* = .40, *SD* = .28). No other significant effects were found in this analysis.

The results plotted in Figure 8 show participants judged and actual effectiveness of the different study schedule and point value conditions, based on their post-test responses and their performance during the memory test. The top panel of Figure 8 (A) displays the participants' responses to the study schedule question, based upon which schedule the participant judged to be the most effective and what schedule was actually the most effective for them on the memory test. Similar to the induction test, participants assigned to memory test overwhelming thought that blocked study schedules were better for category learning (81%), yet in reality participants did better at learning categories under interleaved study (92%).

The lower panel of Figure 8 (B) shows the perceived and actual effectiveness of the different point value conditions for participants in the memory test. Again, participants were split between thinking that 10 point categories were learned the best (44%) or that point values did not matter (47%). In reality, participant performance was relatively distributed. For some, higher point value categories such as 10 points (36%) or 5 points (28%) had the best performance on the memory test, while other participants did better at remembering categories paired with a value of 1 point (19%), or the point value manipulation lead to equivalent levels of performance across two or more of the value conditions (17%).

Interestingly, the benefit of interleaved practice over blocked practice on category induction (about 40%) was nearly twice as large as what was found in prior work (about 25%; Kornell & Bjork, 2008). The difference in magnitude of this effect between the current experiment and Kornell and Bjork (2008) is striking, given the materials and procedures used were nearly identical (i.e., same materials, same time durations, same number of items at test, etc.). The difference can only be attributed to value manipulation used in the current experiment. Interestingly, the relative difference between blocked and interleaved schedules is not driven by improvements in interleaved categories, but by impairments in blocked categories (about 20% and about 35% in this current experiment and Kornell & Bjork, 2008, respectively). A bifurcation model could potentially explain this relationship (Halamish & Bjork, 2011; Kornell, Bjork, & Garcia, 2011; Storm, Friedman, Murayama, & Bjork, in press). Specifically, 10-point categories in the interleaved condition can be selectively prioritized and learned during study, and those categories subsequently receive a larger benefit. However, for blocked categories it is likely more difficult for the learners to selectively focus on the 10-points categories because those categories cannot be prioritized as effectively (i.e., learners cannot selectively encode when all of the high value information is presented at the same time). Therefore, all of the blocked categories are learned to an equivalent degree, while the 10-point categories are learned relatively better than other categories in the interleaved condition. Due to interference and forgetting, a significant proportion of blocked categories are below a learning criterion (i.e., whether or not the category has been learned sufficiently for the induction test), while 10-point interleaved categories are still above that criterion. However, depending on the measure used to interpret this and the other reported effects, the results of the current and subsequent experiments could be interpreted differently[1].

Overall, these results indicate that VDR is present in category learning, but only under certain conditions. Specifically, value only has an influence on category learning when categories are studied in an interleaved fashion, and not in a blocked fashion. Interestingly, value may even carry a negative impact under blocked schedules of learning, in the sense that participants had the greatest difficulty correctly matching the value associated with high-value blocked categories. However, it appears that studying categories with value (regardless of the study schedule used) may come at a potential cost, in the sense that participants were more inclined to give incorrect responses during the test that were associated with higher value categories. Conversely, since participants were instructed to optimize their point total at test, the above result may reflect the use of an optimal strategy by participants (i.e., although participants may have given more incorrect responses associated with high value categories, they also correctly identified more paintings at test associated with high value categories than low value categories). A majority of these findings were also found when a memory test was given instead of an induction test (to differing degrees). Importantly, a large majority of participants reported that blocked study was more beneficial for category learning than interleaved study, despite the opposite being true for their actual performance (on both the induction test and the memory test). To further investigate this dissociation between perceived and actual performance, Experiment 2 investigated participants' abilities to monitor their own learning.

Experiment 2

Experiment 1 illustrated that, despite the positive benefit interleaved practice had on category learning, participants thought that blocked schedules of learning led to better category induction. To further explore learners' awareness, or lack thereof, Experiment 2 was designed to look at participants' ability to monitor their own learning during study, using JOLs. Specifically, would participants give higher JOLs to categories perceived as more fluent (i.e., blocked; see Zechmeister & Shaughnessy, 1980). Also would value, which had a smaller yet meaningful influence on category learning, also be viewed as relevant cue for participants to use in monitoring their own learning? Prior work has shown that learners use value as a cue when making their JOLs to varying degrees in episodic memory tasks (Friedman & Castel, 2011; Gerlach, 2008; Soderstrom & McCabe, 2011), but will learners consider value to be a useful predictor of learning in a category induction task?

Methods

Participants. Sixty-two undergraduates from the University of California, Los Angeles were recruited for this task (17 males, mean age = 20.8). All participants self-reported they had little to no background in the artists that were studied prior to participation. All participants were compensated with course credit for their participation. The data from two participants were removed from all analyses for failure to properly follow directions.

Materials. The materials used for Experiment 2 were identical to that of Experiment 1.

Procedure. The procedure for Experiment 2 was identical to that of Experiment 1 with the following exceptions. The experiment was administered through Collector – an open-source data collection software. After entering basic demographical information (age, gender, etc.), participants were given the instructions about the experiment, informing them about learning artists' styles, how the test would be administered, and to pay particular attention to artists associated with higher point values. Additionally participants were instructed, after studying each painting, to rate how well they felt they knew that artist's "style" on a scale from 0 to 100. A rating of 0 indicated that the participant did not know the artist's style at all, while a rating of 100 indicated that the participant knew the artist's style extremely well. Participants had as much time as they wanted to enter their ratings, and were encouraged to use the entire scale. Participants entered their ratings by typing their responses in on a keypad.

Following the study and rating of the last painting during the learning phase, participants played 45 s of Tetris before beginning the induction test. Like in Experiment 1, participants were instructed that if they could correctly identify the artist who painted the work, they would be awarded that artist's associated point value which would be added to their total. Again, participants had as much time as they wanted to enter their responses, by clicking on the associated artist's name using the mouse (as opposed to verbally giving their responses like in Experiment 1). After completing the induction test, all participants were given instructions for a memory test. Participants were told they would see paintings they had initially studied from the learning phase of the experiment, and they would have to identify (or remember) which artist painted the work. For this memory test, 48 paintings from the initial learning phase were selected (a random subset of 4 from each of the 12 artists). All 48 paintings were shown in a block randomized order such that, across each block of 12 paintings, 1 painting from each artist was shown. Like in the induction test, participants had as much time as they wanted to respond by clicking the artist's name on the screen as a given painting was shown. Following the completion of the memory test, participants responded to the same post-test questions used in Experiment 1.

Results and Discussion

Judgments of Learning

Figure 9 displays the average JOL value assigned by participants across the learning phase of Experiment 2. Overall, JOLs got progressively higher as participants studied additional paintings from each artist and10 point categories were given higher JOLs than lower point categories. Additionally, blocked categories were given lower JOLs than interleaved categories during the first study occurrence, but participants gave equivalent JOLs to blocked and interleaved categories by the final study occurrence. A 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) by 6 (Occurrence during Study: 1, 2, 3, 4, 5, 6) repeated measures ANOVA found a marginal effect of schedule, $F(1, 59) = 2.24$, $MSE = 810.24$, $p = .14$, $\eta_p^2 = .04$, such that categories studied in an interleaved fashion ($M = 47.90$, $SD = 19.09$) were assigned slightly higher JOLs by participants than were categories studied in a blocked fashion ($M = 46.08$, $SD = 18.62$). A main effect of value was also found, $F(2, 118) = 17.98$, $MSE = 831.97$, $p < .001$, $\eta_p^2 = .23$, such that categories assigned a value of 10 points ($M = 52.25$, $SD = 18.56$) were given higher JOLs than categories assigned a value of 5 points ($M = 44.61$, $SD = 18.69$), $t(59) = 5.74$, $p < .001$, $d = .41$, or 1 point ($M = 44.12$, $SD = 21.06$), $t(59) = 4.48$, $p < .001$, $d = .41$. The above main effect is relatively important, as it supports the claim that people think that valuable information is more memorable than less valuable information, and thus value or importance may be one cue that people attend to during study. A main effect of occurrence was also found, $F(5, 295) = 39.10$, $MSE = 230.50$, $p < .001$, $\eta_p^2 = .40$, such that JOLs got progressively higher as participants saw additional paintings from each category. Lastly, a schedule by study occurrence interaction was found, $F(5, 295) = 4.54$, $MSE = 104.84$, $p = .001$, $\eta_p^2 = .07$, such that participants gave higher JOLs on the first study occurrence for interleaved categories ($M = 43.50$, $SD = 20.01$) than blocked categories ($M = 37.47$, $SD = 19.53$), $t(59) = 5.37$, $p < .001$, $d = .81$. However, by the last study occurrence interleaved categories ($M = 53.70$, $SD = 22.88$) and blocked categories ($M = 52.82$, $SD = 19.53$) were given equivalent JOL values, $t(59) = 0.47$, $p = .64$. This interaction provides support for the idea that fluent information – in this case, categories shown in a blocked fashion – are perceived as easier to process and therefore the JOLs participants assigned to those categories got higher at a faster rate than information perceived as less easy to process (i.e., interleaved categories).

Induction Test Performance

Performance on the induction test is plotted in Figure 10. The top panel of Figure 10 (A) displays average proportion of new paintings correctly identified by participants. For interleaved categories, the proportion of correct responses did not significantly change with value, but the proportion of correct responses decreased across value for blocked categories. The 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a main effect of value, $F(1, 59) = 67.89$, $MSE = 0.09$, $p < .001$, $\eta_p^2 = .54$, such that participants correctly identified more new paintings when those categories were studied in an interleaved fashion ($M = .58$, $SD = .19$) than in a blocked fashion ($M = .32$, $SD = .19$). A study schedule by point value interaction was also found, *F*(2, 118) = 6.00, *MSE* = 0.06, *p*< .01, η_p^2 = .09, such that categories studied with an interleaved schedule did not change with point value, *F*(2, 118) = 2.00, *MSE* = 0.05, *p* = . 14, η_p^2 = .03 (although the results appear to be trending towards greater proportion correct with higher values), but a very different pattern was found for categories studied under blocked schedules. Specifically, the associated ANOVA on value for blocked learning only, *F*(2, 118) = 5.58, *MSE* = 0.05, *p*< . 01, η_p^2 = .09 revealed that categories assigned a value of 1 point (*M* = .40, *SD* = .26) were better learned than categories assigned a value of 5 points (*M* = .28, *SD* = .28) or 10 points (*M* = .27, *SD* = .27; *t*s > 2.5)

The lower panel of Figure 10 (B) displays the summed average of incorrect responses participants gave during the induction test for each study schedule-point value condition. As point values increased, the number of incorrect responses decreases for interleaved categories, but there was no apparent difference was found for blocked categories. The associated 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a significant interaction, *F*(2, 118) = 4.83, *MSE* = 5.07, *p*= .01, η_p^2 = .08, such that participants gave more incorrect responses associated with interleaved-1 point categories (*M* = 5.03, *SD* = 2.92) than blocked-1 point categories (*M* = 3.93, *SD* = 2.77), *t*(59) = 2.43, *p* = .02, *d* = .39. However, this trend reversed for the10 point categories, in that participants gave marginally more incorrect responses associated with blocked-10 point categories (*M* = 4.55, *SD* = 2.46) than interleaved -10 point categories (*M* = 3.87, *SD* = 2.63), *t*(59) = 1.76, *p* = .08, *d* = .27.

Memory Test Performance

Performance on the memory test is plotted in Figure 11. The top panel of Figure 11 (A) displays the average proportion of paintings studied during the learning phase of the experiment which participants correctly identified. The results reported here are relatively similar to those reported in the induction test, in that performance decreases as value increases for categories studied in a blocked manner, however, this pattern reverses for interleaved schedules of learning in that the proportion of correct identifications increases with value. The 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a main effect of schedule, $F(1, 59) = 42.62$, $MSE = 0.12$, $p < .001$, $\eta_p^2 = .42$, such that paintings associated with interleaved categories ($M = .59$, $SD = .21$) were correctly identified more often than paintings associated with blocked categories ($M = .35$, $SD = .22$). Additionally, a significant study schedule-point value interaction was found, $F(2, 118) = 9.87$, $MSE = 0.07$, $p < .001$, $\eta_p^2 = .14$, such that there were no apparent differences in participants correctly identifying paintings associated with either blocked or interleaved study schedules for 1 point categories, $t(59) = 1.16$, $p = .25$, yet the proportion of paintings correctly identified was significantly higher for interleaved study than blocked study when the value assigned to the categories was 5 points, $t(59) = 5.32$, $p < .001$, $d = .96$, or 10 points, $t(59) = 6.96$, $p < .001$, $d = 1.31$.

The lower panel of Figure 11 (B) displays the summed average of incorrect responses that participants gave for each study schedule-point value condition on the memory test. As the point values associated with the categories increased, the number of incorrect responses participants gave decreased for interleaved categories, however no such influence of value was found for blocked categories. The associated 2(Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a marginal interaction between study schedule and value, $F(2, 118) = 2.49$, $MSE = 6.24$, $p = .09$, $\eta_p^2 = .04$, such that, as values increased for interleaved categories, the number of incorrect responses participants gave associated with those categories decreased, $F(2, 118) = 3.17$, $MSE = 6.78$, $p = .05$, $\eta_p^2 = .05$, while no apparent differences associated with value were found for blocked categories ($F < 1$).

Post-Test Questions

Like in Experiment 1, three post-test questions followed the completion of the memory test. Figure 12 displays the proportion of values correctly assigned to the artists by participants post-test, based on those artists' (i.e., category) associated study schedule and point value conditions. Overall, point values were correctly assigned to their artists equivalently across the various conditions, with the exception of the blocked-10 point artists which were correctly assigned their appropriate value by participants significantly less often. The 2(Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a main effect of schedule, $F(1, 59) = 12.25$, $MSE = 0.10$, $p < .001$, $\eta_p^2 = .17$ such that values associated with interleaved schedules ($M = .53$, $SD = .24$) were correctly assigned more often by participants than values associated with blocked schedules ($M = .42$, $SD = .22$). A main effect of value was also found, $F(2, 118) = 7.95$, $MSE = 0.12$, $p = .001$, $\eta_p^2 = .12$, such that categories associated with a value of 10 points ($M = .37$, $SD = .25$) were correctly assigned their value by participants less often than for categories associated with a value of 5 points ($M = .54$, $SD = .30$) or 1 point ($M = .51$, $SD = .29$, ts> 3). However, both main effects need to be qualified by the significant study schedule-point value interaction, $F(2, 118) = 5.69$, $MSE = 0.10$, $p < .01$, $\eta_p^2 = .09$, such that there were no significant differences in participants correctly assigning values post-test between blocked and interleaved study schedules for 1 point and 5 point conditions (ts < 1), yet blocked categories assigned a value of 10 points ($M = .23$, $SD = .30$) were correctly assigned their value less often by participants than interleaved categories ($M = .51$, $SD = .40$), $t(59) = 4.27$, $p < .001$, $d = .79$. Therefore, it appears as though the above interaction is being driven by participants' relatively poor performance in remembering the values associated with the blocked-10 point condition.

Following the post-test assignment of values to their artists, participants were asked whether they thought it was easier to learn the categories when the paintings by a given artists were shown all together (blocked) or shuffled together with paintings from other artists (interleaved). Like in Experiment 1, a substantial proportion of participants (75%) thought that blocked study led to the best learning of the categories, while a minority thought that interleaved practice led to better category learning (8%), or that both learning schedules were equally good (8%), or that the different schedules did not matter for learning the categories (8%). However, when compared to the participants' actual performance on the induction test, a majority of participants (83%) actually showed a clear benefit for interleaved study schedules, while a small minority showed a benefit for blocked practice (12%) or equivalent levels of performance for blocked and interleaved categories (5%).

Participants were also asked whether or not they thought the point values were beneficial to learning the categories, and if so, which point value category led to the best learning. A large majority of participants reported that point values did not help them learn the categories any better (70%), despite the fact that participants gave higher ratings to 10 point categories in their JOLs. A smaller proportion reported they thought that categories associated with a value of 10 points were best learned (28%). In reality, however, performance on the induction test was relatively more distributed. A small majority of participants actually did best at learning categories associated with a value of 1 point (37%), while the remainder of subjects actually had equivalent levels of performance in two or three of the point value conditions (25%), then categories assigned a value of 10 points (20%), and last categories assigned a value of 5 points (18%).

Overall, participants' JOLs were relatively equivalent across both schedules of study, in the sense that JOLs for blocked and interleaved schedules equaled one another by the last studied exemplar during the learning phase. However, across occurrences during learning, participants gave relatively higher JOLs during each subsequent occurrence for blocked schedules than interleaved schedules, indicating some degree of fluency within the former. That is, participants likely thought that the blocked categories were easier to learn (and therefore gave progressively higher JOLs) because those categories were presented in sequence and seemed easier to process. By comparison, participants rated their JOLs more conservatively for interleaved categories (as indicated by the slopes of the interleaved lines in Figure 10), yet performance was significantly higher on both the induction and memory tests for interleaved categories.

Interestingly, participants did use value as a basis for their JOLs, yet value had a fascinating impact on test performance. Namely, value had a positive impact on performance with interleaved schedules of learning (with better performance associated with higher values), yet value had the opposite impact with blocked schedules of learning (with *worse* performance associated with higher values). These "ironic" effects of value are theoretically interesting, as they may help explain why VDR may not be found (or even trend in the opposite direction) based on the study schedule employed. Specifically, blocked study schedules may make selectively rehearsing high-value information too difficult, as it may become overly-challenging to recognize differences across categories (Birnbaum et al., 2013; Kang & Pashler, 2012). The additional effort or rehearsal participants were putting into learning the high-value categories may have contributed to their relatively lower performance (Dodson et al., 1997), which is why the blocked 1-point categories were correctly identified more often. The above finding can almost be viewed as an undesirable difficulty – if everything is perceived to be important (like it would be in high-value blocked study), learners would be unable to selectively encode or focus on specific information over other information, leading to a perceived insensitivity to value on performance.

Experiment 3

Experiment 2 explored participants' ability to accurately monitor their own learning during study, using JOLs. The results show that participants were not sensitive to the different study schedules (i.e., blocked or interleaved), yet were relatively more sensitive to value. The goal of Experiment 3 was to further explore learners' sensitivity to these conditions by assessing their monitoring at test, using confidence ratings. Additionally, to further explore the costs associated with learning categories of different values, a recognition test was used, similar to that used in Kornell and Bjork (2008; Experiment 2). The goal of this change in the test format was to explore whether participants would be more or less likely to misidentify a stylistically similar painting as belonging to a category based on its study schedule-point value condition. High-value information, while remembered better overall by learners, comes at the cost of being more susceptible to errors of commission with semantically related information, relative to low-value information (Bui et al., in press). Therefore, the test procedure of the current experiment was changed such that participants were asked at test whether the paintings they were studying at test (some painted by the initially studied artists, others painted by stylistically similar artists, see Figure 13) was created by an artist they had initially studied (i.e., old), or by an artist the learner had not previously studied (i.e., new). Based on the findings from Bui et al.(in press), it seems probable that participants would be more likely to incorrectly identify a stylistically similar painting as old if that initially studied category was relatively higher in value than lower in value. Additionally, the learning procedure was modified to include a pre-training phase in which learners could effectively study just the artist-point value associations. In that way, learners would not be overly distracted on the artist-value associations during the actual learning of the categories, which may have partially affected the results of Experiments 1 and 2.

Methods

Participants. Sixty-three people were recruited using Amazon Mechanical Turk for this experiment (37 males, mean age = 32.6). All participants self-reported they had little to no background in the artists that were studied prior to participation, and were fluent in English. Participants were compensated $1 each for their participation. Of the 63 participants, 73% were currently living in the United States at the time of participation, 24% were living in India, and 4% from other countries. The data from two participants were removed from all analyses for failure to complete the experiment, while another participant's data were removed for taking a long pause during the experiment (to eat lunch).

Materials. The materials used for Experiment 3 were identical to those in Experiments 1 and 2; however an additional 48 paintings were also used (see Kornell & Bjork, 2008, Experiment 2). These additional paintings were selected based on their stylistic similarity to that of each of the original 12 artists. There were four paintings selected for each of the 12 artists studied, and each of those four paintings was created by a different artist (see Figure 13).

Procedure. The procedure to Experiment 3 was similar to that of Experiment 2 except that, before learning, a pre-training phase was introduced to familiarize learners with the artist-value pairings, ratings during learning (i.e., JOLs) were removed and ratings during test (i.e., confidence ratings) were used, and the goal at test changed from identifying the artist of the work to identifying whether the artist who painted the work was previously studied or not.

Like in Experiments 1 and 2, participants were given instructions that they would see multiple paintings by various artists, and it would be their job to learn the styles of each artist in order to recognize new paintings by those artists, and that the styles of certain artists would be more valuable to remember than others based on the point value assigned to that artist. Before beginning the learning phase, participants were told they would go through a pre-training phase to get them acquainted with the artists' names and the point value associated with each artist. The goal of this pre-training was to make sure participants would not become severely distracted by either the names of the artist or the associated point values during learning, and allow them to focus more of their cognitive efforts on learning the styles of the 12 artists. During this pre-training phase, participants saw an artist's name with his or her associated point value on the screen. The participants studied each artist-value pairing as long as they wanted to, and advanced to the next artist-value pairing by clicking a button labeled "Next" directly below the pair. The 12 artist-value pairings were shown three times in blocked randomized order, such that each artist-value pairing was studied once within each block. After participants studied the last artist-value pairing in the pre-training phase, participants were reminded of the instructions again, and then began the learning phase of the experiment, similar to that used in Experiment 2. Importantly, participants were not prompted to give JOLs during the learning phase of the experiment.

After completing the 45 s Tetris distractor, participants began the test phase of the experiment. The participants were told that they would see paintings they did not see earlier and it would be their job to identify whether the artist who painted the work was one that was previously studied (i.e., old) or one that they did not study during the learning phase of the experiment (i.e., new). During the test phase, participants saw a painting on the screen with two buttons directly beneath it, each labeled “OLD” or “NEW” and had to identify whether the painting shown was created by an artist previously studied or not. Participants had as much time as they wanted to respond for each painting, and the program automatically advanced as soon as the participant made their decision for the current painting on the screen. After making their response, participants were asked to rate how confident they were in their previous old/new decision on a scale from 0 to 100, where 0 indicated the participant was not confident at all in their previous response, and 100 indicated the participant was very confident in their previous response. Participants were encouraged to use the entire scale for their ratings, and had as much time as they wanted to give their rating responses. Participants entered their responses with their keyboards or keypads when they were prompted to give their ratings. After making their old/new decision for a given painting, participants saw the question “How confident are you in your previous response?” appear on the screen, with a textbox beneath it to enter their numerical response. After entering their response, the program advanced to the next painting.

During the test phase of the experiment, a total of 96 paintings were shown to the participant. Forty-eight of the paintings shown, like in Experiments 1 & 2, were paintings created by the 12 artists studied during the learning phase of the experiment but were not studied during that phase of the experiment (four paintings for each of the 12 artists). The remaining 48 paintings were stylistically similar to those of the 12 artists initially studied, but were painted by different artists, and classified as distractor, or “lure”, paintings (for examples, see Figure 13). For each of the 12 artists, four paintings, each from a different artist not studied during the learning phase, were selected as distractors. During the test phase of the experiment, the 96 paintings were shown in four randomized blocks of 24 paintings. Within each block, one painting from each of the 12 artists was included, as well as one distractor painting associated with each of the 12 artists. After giving their responses and ratings to all 96 paintings, the participants responded to the same post-test questions used in Experiments 1 and 2 before being debriefed and dispersed their payment for participation.

Results and Discussion

Recognition Test Performance

Figure 14 displays the performance of participants on the recognition test. The black and white bars in that figure represent the proportion of new paintings from previously studied artists that participants correctly identified (i.e., hits). Overall, the proportion of hits did not increase with value for blocked categories, but hits did increase with value for interleaved categories. The associated 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA for hits found a significant main effect of schedule, $F(1, 59) = 9.69$, $MSE = 0.03$, $p < .01$, $\eta_p^2 = .14$, such that interleaved categories ($M = .66$, $SD = .21$) had more hits than blocked categories ($M = .60$, $SD = .22$). A marginal main effect of value was also found, $F(2, 118) = 2.78$, $MSE = 0.04$, $p < .07$, $\eta_p^2 = .05$, such that categories associated with a value of 10 points ($M = .66$, $SD = .24$) had a greater proportion of hits than categories associated with a value of 1 point ($M = .60$, $SD = .22$), $t(59) = 2.47$, $p = .02$, $d = .26$. A study schedule by point value interaction was also found, $F(2, 118) = 3.41$, $MSE = 0.04$, $p = .04$, $\eta_p^2 = .06$, such that there was no apparent effect of value on the proportion of hits for blocked categories ($F < 1$), yet the number of hits increased with value for interleaved categories, $F(2, 118) = 5.49$, $MSE = 0.04$, $p < .01$, $\eta_p^2 = .09$.

The two sets of gray bars in Figure 14 represent the proportion of new paintings from previously *unstudied* artists (yet similar in artistic style to artist categories initially studied during learning) that participants incorrectly identified as old (i.e., false alarms). The overall frequency of false alarms was relatively consistent across conditions, with the exception that there were significantly fewer false alarms for the blocked-10 point categories. The associated 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA for false alarms found a marginal effect of schedule, $F(1, 59) = 2.65$, $MSE = 0.02$, $p = .11$, $\eta_p^2 = .04$, such that categories learned under blocked study ($M = .46$, $SD = .20$) had slightly fewer false alarms than categories learned under interleaved study ($M = .48$, $SD = .21$). Interestingly, a main effect of value was not found ($F < 1$), contrasting with the findings of Bui et al. (2013), which found that errors of commission of semantically-related lures increased with value. Lastly, a marginally significant interaction was found, $F(2, 118) = 2.29$, $MSE = 0.04$, $p = .11$, $\eta_p^2 = .04$, such that participants had significantly fewer false alarms for categories assigned a value of 10 points studied in a blocked fashion ($M = .42$, $SD = .24$) compared to categories studied in an interleaved fashion ($M = .49$, $SD = .25$), $t(59) = 2.49$, $p = .02$, $d = .29$, while no such differences in false alarm rates were found between blocked and interleaved study for categories associated with values of 1 point or 5 points ($ts < 1$).

Confidence Ratings

The average confidence rating values reported by participants during the recognition test are plotted in Figure 15. A small number of participants' data were not included in this analysis, as those participants did not have misses or correct rejections, and therefore had no associated confidence ratings for those conditions. In the initial 4 (Response Type: hits, misses, false alarms, correct rejections) by 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA for confidence ratings, the only significant term was the interaction between response type and value, $F(6, 108) = 2.31$, $MSE = 103.85$, $p = .04$, $\eta_p^2 = .11$, so to further explore this interaction the study schedule independent variable was removed from the analysis (and collapsed across for Figure 15).Overall, value did not influence participants' confidence ratings across the different response types, with the exception of hits – categories assigned a value of 10 points were given higher confidence ratings by participants than categories assigned a values of 5 points or 1 point. The 4 (Response Type: hits, misses, false alarms, correct rejections) by 3 (Value: 1, 5, 10) repeated measures ANOVA found a significant effect of response type, $F(3, 144) = 12.05$, $MSE = 260.04$, $p < .001$, $\eta_p^2 = .20$, such that the confidence ratings were significantly higher for hits ($M = 73.95$, $SD = 15.31$) than for correct rejections, misses, or false alarms (all *t*s > 3). A marginally significant interaction between response type and point value was also found, $F(6, 288) = 70.13$, $MSE = 70.13$, $p = .08$, $\eta_p^2 = .20$, such that, for hits, participants' confidence ratings were significantly higher for 10 point categories than for 5 or 1 point categories, $t(58) = 2.39$, $p = .02$, d = .17, while no such differences were found across the different point value conditions for the other response types (*F*s < 2).

Post-Test Performance

Following the recognition test, participants were asked three post-test questions. Figure 16 displays the proportion of values correctly assigned to the artists by participants based on those artists' (i.e., category) associated study schedule and point value conditions. Overall, value had an influence on remembering the correct value associated with a given category such that the value associated with 10 point categories was remembered best, then 1 point categories, then 5 point categories. The 2 (Schedule: blocked, interleaved) by 3 (Value: 1, 5, 10) repeated measures ANOVA found only a main effect of value, $F(2, 118) = 14.49$, $MSE = 0.06$, $p < .001$, $\eta_p^2 = .20$, such that participants better remembered the values associated with 10 point categories ($M = .88$, $SD = .18$), than 5 point ($M = .72$, $SD = .27$), or 1 point ($M = .83$, $SD = .21$) categories (all *t*s> 2). Additionally, when this dependent measure in the current experiment was compared to the previous two experiments, performance was markedly higher, $F(1, 189) = 117.67$, $MSE = 0.23$,

$p < .001$, $\eta_p^2 = .38$ such that the average proportion of values correctly assigned in Experiment 3 ($M = .81$, $SD = .19$) was significantly greater than the overall proportion of values correctly assigned in Experiments 1 and 2 combined ($M = .48$, $SD = .19$). The above finding illustrates, quite clearly, that the pre-training used in Experiment 3 greatly improved the participants' ability to learn and remember the value-category associations.

Following the post-test assignment of values to their artists, participants were asked whether they thought it was easier to learn the categories when the paintings by a given artist were shown in a blocked or interleaved pattern of study. A majority of participants thought that blocked study schedules were better for learning the categories (52%), while the remainder of participants thought that interleaved study was better for learning the categories (18%), or that both learning schedules were equally good for learning the categories (15%), or that the different schedules did not influence their ability to learn the categories (15%). In reality, however, when the mean differences between hits and false alarms on the recognition test were compared for participants, a majority of participants learned categories studied under interleaved schedules better (58%) compared to categories learned under blocked schedules (33%), while a small minority of participants had equivalent levels of performance across the two study schedule conditions (8%).

Last, participants were asked whether not they thought the point values were beneficial to learning the categories, and if so, which point value category led to the best learning. A majority of participants (62%) thought that the categories associated with a value of 10 points were learned best, while a smaller proportion of participants (33%) thought that the different values made no impact on their learning of the categories. When the mean differences between hits and false alarms on the recognition test were compared across the different point value categories for participants, the results were relatively more distributed. A small majority of participants (38%)

performed best at discriminating category members when those categories were associated with a value of 10 points. The remaining participants were best at discriminating category members when those categories were associated with a value of 5 points (25%), 1 point (20%), or performed equally across two or three of the different value conditions best (17%).

Again, a consistent pattern of VDR was found in interleaved schedules of learning, but not for blocked schedules, illustrating that VDR can only occur (or is more probable to occur) under certain learning conditions. Interestingly, there does not seem to be an associated cost to participants' sensitivity to value in interleaved schedules of learning, as the false alarm frequency does not change across value conditions. This contrasts with the findings of Bui et al. (in press), which did find an increase in the false recall of semantically-related information as point values increased. While speculative, the differences between Bui et al. (2013) and the results reported in this chapter may be due to the nature of the materials that were used (semantically-related words and non-verbal materials, respectively). Semantically-related verbal material may facilitate the ease of encoding high-value material in selectivity paradigms by helping learners activate a semantic network of related information, thus easing the recall of high-value information at the cost of increased rates of false recall of semantically-related information. The materials used in these reported experiments, while semantically-related to one another in a different sense (stylistically), may not lead to the same degree of semantic activation that verbal materials do, which is why the associated cost (i.e., increased false alarms with increased values) may not have changed across value conditions.

Interestingly, confidence ratings that participants gave were relatively consistent across the different response types and value conditions. The one exception to this consistency, however, was the hit responses, in the sense that the associated confidence ratings increased as value increased (i.e., paintings at test associated with 10 points were given higher confidence ratings by participants than those associated with 1 or 5 points). This VDR finding in confidence ratings for hits (and lack of VDR in confidence ratings for false alarms) is somewhat encouraging because it reflects the relative accuracy participants had in estimating their performance on this recognition task. Overall, the results of this experiment reflect the sensitivity learners have for value in interleaved study schedules of category learning (i.e., hit rate increasing with value) without showing an associated cost (i.e., no increase in false alarm rates with value), and learners being able to monitor their performance relatively accurately at test by using value as the basis for confidence ratings (all the more impressive since no value cues were given at test).

General Discussion

Across three experiments, learners' sensitivity to value was investigated in the context of category learning. Importantly, all three experiments found similar results: value-directed remembering - an active cognitive process - can lead to a benefit in category induction - a relatively passive form of learning - if categories are learned under interleaved study schedules. The above result is remarkable, given that participants only had 3 s to encode the artist's name, the associated point value, and the painting itself before the next one was shown. Additionally, the benefit of interleaved practice in category learning was consistently found as well, reproducing the same result found in other work on category induction (Birnbaum et al., 2013; Kang & Pashler, 2012; Kornell & Bjork, 2008). Importantly, the current work explored category induction with materials more closely associated with information integration category learning. Future work should explore learners' sensitivity to value in category learning by using rule-based categorization materials (Ashby et al., 2002). Since learning rule-based categories is a relatively more active process (which requires the use of working memory) compared to learning information integration categories, the attention and selective rehearsal processes required for VDR may be interfered with, reducing the influence of value on category learning.

The theoretical impact of these findings is relatively informative in explaining the relationship between value and memory for future research on selectivity tasks. Since the impact of value on learning appears to take place only during interleaved schedules of study (i.e., when high and low-value information are studied together), researchers could make predictions as to whether VDR would or would not be found under certain schedules of study. With interleaved schedules of study, learners are able to successfully select for and rehearse high-value information as opposed to rehearsing low-value information. Conversely, blocked learning schedules have all of the high-value information “clustered” together, which makes it difficult for learners to selectively encode important information because everything is valuable to remember during that time period of learning. Ultimately, the results found in these experiments emphasize that important information needs to be interleaved with less-valuable information to allow learners to become successful at selective retrieval. Future work should investigate these findings using more traditional selectivity paradigms (e.g., using word lists), to confirm whether or not blocked and interleaved schedules of learning always do (or do not) lead to VDR.

The metacognitive findings in relation to value in these experiments were remarkable as well. In both Experiments 2 and 3, participants accurately predicted that value would have a positive impact on their performance (i.e., as value increased, so did performance on the induction test). Although the effect of value on actual performance was relatively small compared to the benefit of interleaved learning, value had a reliable effect on performance and, importantly, learners did use value as an informative cue when making their metacognitive ratings. Theoretically, these metacognitive results support the view that value may be an extrinsic cue (Koriat, 1997), in that learners can use value as the basis of their metacognitive judgments, and it will accurately reflect their actual performance. Future research should attempt to replicate these metacognitive findings of learners' sensitivity to value with other manipulations known to influence metacognitive monitoring (e.g., manipulations of perceptual fluency) or actual performance (e.g., testing effects, generation, etc.).

Importantly, this is also the first work (to the author's knowledge) that investigated metacognitive monitoring for blocked and interleaved schedules of learning in the context of category induction. Overall, Experiment 2 showed that learners had a stronger sense of fluency for categories learned under blocked than interleaved study schedules (in that JOLs increased at a faster rate), while Experiment 3showed no apparent impact of the study schedules on participants' confidence ratings at test. These findings are surprising in the sense that actual performance in all three experiments reflects the positive benefit of interleaved schedules of learning; however, the post-test questions regarding blocked and interleaved study asked in the experiments reported and in related work on category induction (e.g., Kornell & Bjork, 2008) corroborate the metacognitive monitoring results. Namely, learners generally feel that blocked schedules of study are superior for category learning. Future research on category induction should attempt to make the benefits of interleaved study more apparent to learners and, in effect, make them more accurate at monitoring their own learning.

The implications of this research are very applicable to both formal and recreational learning. Specifically, since learners are sensitive to value and it can impact their learning of categories, specific categories could be emphasized over others to facilitate their learning. For example, medical interns could be trained to identify and classify hairline fractures in bones or malignant tumors (i.e., high-value) with greater accuracy over rashes or acne (relatively less valuable) if all these categories were interleaved together. This concept could also assist in recreational learning as well, such as being able to identify birds or trees native to your local area over those of from foreign countries, or being able to distinguish edible, nonedible, and poisonous plants and animals from one another. Based on the materials used in these experiments, the applications of this work to art history and appreciation are obvious. Ultimately, manipulating the learning of categories by interleaving examples of multiple categories together (some important, others less so) reflects learners' ability to prioritize what information they need to learn, and effectively exert control over their learning based on their desired goals (Ariel et al., 2009).

To conclude, the influence of value on learning, and the specific mechanisms associated with learners' sensitivity to value, still need further exploration. Across all three experiments reported, the importance of interleaved learning appears to be critical for both successful category induction and value-directed remembering. Conversely, studying information in blocked schedules appears to impede both category learning and learners' sensitivity to value, and can even lead to ironic effects of value (with participants performing better in low-value than high-value conditions).

Chapter 3: The Influence of Objective and Subjective Values on Memory for Medication Side Effects in Younger and Older Adults

Remembering important information can be necessary and critical for the efficient use of memory in a variety of settings. Based on Castel et al. (2002), it has been established that both younger and older adults can remember information based on its objective importance or value (value-directed remembering, VDR). However, it is unknown whether or not encoding information based on its value or importance can be (or is) influential on memory when VDR is applied to relatively practical situations. For example, while shopping at a grocery store, people may forget to purchase the occasional item, but do they always remember to buy the items they need in order to cook dinner that evening? Of interest to research on VDR, are the values placed on such items always objective or subjective in nature? That is, in order for VDR to be achieved, does there need to be a clearly defined and specific value associated with the item (e.g., 10 points), or can the value associated with the to-be-remembered item be more vague and/or assigned by the learner (e.g., "this is extremely important for me to remember")? Is it possible that this distinction between objective and subjective value may be irrelevant, as the learner may convert any objective value into a subjective value? If such a value-conversion process exists (i.e., from objective to subjective), it may explain why recall of the highest value items tend to be similar to one another, or at a possible virtual ceiling. For example, a process of converting objective values to subjective importance may explain why recall of 10-point words matches that of 12-point words in a selectivity task (see Castel et al., 2002), as all of those words were potentially just assigned into a generic, "important" category by the learner. Critically, older adults remember high-value information as well as younger adults (Castel et al., 2002; Friedman & Castel, in press), so it is worthwhile investigate whether or not the distinction between objective and subjective values' impact on memory changes across the adult lifespan. Perhaps older adults are more sensitive to subjective values as they need to use such measures of value to guide attention, given age-related limitations in memory?

There is strong reason to believe that, despite cognitive declines, older adults will perform as well as younger adults in a more realistic or practical selectivity task (Castel, 2008a; Zacks & Hasher, 2006). Although younger adults (college undergraduates in particular) might be viewed "memory experts", in that they are always studying for an upcoming exam or assignment (Castel, 2008a), healthy older adults are not oblivious to their declines in memory (e.g., Hertzog & Dunlosky, 2011; McGillivray & Castel, 2011) and are capable of taking steps to ensure their optimal performance on a given task (Baltes & Baltes, 1990). This can range from using mnemonics or strategies, writing down notes to themselves for later, or even requesting reminders from individuals they interact with (Hertzog & Dunlosky, 2011). Furthermore, older adults can draw upon their years of experience with a particular task (i.e., schematic support) to compensate for memory deficits. For example, Castel (2005) showed that older adults could remember the prices of grocery items just as well as younger adults could. However, older adults could only match the performance of younger adults when the prices assigned to the grocery items were realistic and not when the prices were unrealistic (i.e., $2.49 vs. $11.49 for a gallon of milk). Additionally, Miller (2003) showed that older adults with high levels of knowledge in the context of cooking could remember as much cooking-related text information as younger adults with high levels of cooking knowledge, but could not remember as much information for domain-unrelated text passages about biology. The above findings illustrate that older adults can rely on their prior knowledge to assist with a given task, and may reflect the idea that the cognitive declines found in laboratory studies do not truly showcase the abilities of older adults (Zacks & Hasher, 2006). The claim that older adults may use their prior knowledge to compensate for age-related declines in memory applies to the work in this chapter, as older adults tend to have experience (if not expertise) with information relating to medication and their side effects, and domain expertise is generally not impaired in later adulthood (Castel, 2007; Miller, 2003; Morrow et al., 1994). Many experimental designs that require the use of memory do not accurately reflect real-world scenarios and may find poorer performance for older adults than they would otherwise in a more "everyday" setting, as such studies may impede older adults from relying on, or applying, their prior experience and expertise. Relevant to the current research, older adults, as well as younger adults to some extent, may have some prior experience with medically-related information, particularly medication, and it is likely that age-related differences in memory for such information would not be found.

A growing body of literature is investigating the problem of medication misuse, abuse, and drug-drug interactions (Blanco et al., 2007). Currently, about one in three adults in the United States is taking some form of complementary medicine such as a vitamin or herbal supplement, usually with their doctors being unaware as patients fail to mention their use or doctors fail to ask about their use (Magee, 2007). Issues with taking medication properly or potential side effect risks are further compounded by poor and/or confusing labels and warnings on medication (Webb et al., 2008). Two populations seriously at risk for medication problems are students and older adults. Younger adults currently in school have been incentivized to seek a diagnosis of ADHD, even going to efforts of feigning symptoms, for both special accommodations within courses as well as stimulant medication (used as a recreational study aid; DeSantis, Webb, Noar, 2008; Sollman, Ranseen, & Berry, 2010). Older adults are at risk for medication misuse and abuse for several reasons including age-related physiological changes, increased risk for disease, increases in medication use, and increased risk for harmful interactions between alcohol and medication (Hines & Murphy, 2011; Moore, Whiteman, & Ward, 2007). Additionally, prescription and nonprescription medications are commonly used together in later adulthood, which leads to an even greater risk of harmful drug interactions (Qato et al., 2008). With declines in memory, older adults may have greater difficulty in remembering information associated with various drugs they are taking, and thus be especially at risk for drug misuse. Overall, the body of work cited above explains the growing need for researchers to investigate younger and older adults' interactions with medication, and in particular their memory and knowledge for safe medication use.

To a certain degree, the use of memory in the context of remembering medication-related information has been investigated, particularly in the context of prospective memory (i.e., remembering to complete an action sometime in the future; for a review, see Zogg, Woods, Sauceda, Wiebe, & Simoni, 2012). Prospective memory is an important component in proper medication adherence, and interventions designed to improve prospective memory are effective at improving adherence (Haynes, Ackloo, Sahota, McDonald, & Yao, 2007; Liu & Park, 2004). The need for optimal prospective memory is even more critical for older adults, yet older adults do not always show age-related declines in prospective memory performance (illustrating some differences between episodic and prospective memory; see Einstein & McDaniel, 1990; Henry, MacLeod, Phillips, & Crawford, 2004; but see McDaniel & Einstein, 2007). However, the way in which medication-related information is presented to individuals may influence what information they remember. (Morrow, Leirer, & Sheikh, 1988; Webb et al., 2008). Most medication information shown in modern magazine advertisements, television or radio commercials, and other forms of media have to mention the "risk information" associated with that medication (as required by the laws of the US Food and Drug Administration, FDA). However, FDA laws fail to mention specifically how that risk information needs to be presented (see “Fair Balance”: http://www.fda.gov/Drugs/ResourcesForYou/Consumers/PrescriptionDrugAdvertising/ucm072025.htm). As a result, many ads explaining the risk factors associated with a given medication are presented in relatively small font, stated (or presented) rapidly, and numerous risk factors are typically listed. Given these current presentation methods, it seems probable that patients (both younger and older adults alike) are being forced to be selective when learning about these medications, and encoding only the information they consider to be valuable or important. Therefore, it seems worthwhile to investigate whether the information younger and older adults deem important in the context of medication side effects (i.e., risk information) will actually impact their memory for that information.

The specific aim of this research is to explore whether or not value or importance can have an impact in a situation in which people must attend to information associated with medication (i.e., the associated side effects). Since older adults have impairments remembering vast amounts of information, such as when learning about or using medications, could older adults be more sensitive to the information they consider to be valuable and remember it better, relative to younger adults? Furthermore, if value is manipulated either objectively (e.g., different point values) or subjectively (e.g., personal importance to the learner), does the type of value used impact memory differently, and if so, does the impact of value-type on memory change across the adult lifespan? For example, would learners be more likely to remember side effects that they perceive to be severe (e.g., increased risk of stroke) compared to side effects that are relatively milder (e.g., nausea). Also, would learners be more likely to remember specific side effects if those effects, while otherwise mild in severity, may be indicative of a life-threatening complication? In the present study, using a basic design (i.e., learning medication side effects), younger and older adults studied and were tested on a list of side effects, that varied in importance or value to differing degrees. Value was manipulated on an objective (i.e., severity categories or warnings of potential complications) and subjective basis (i.e., the individuals' perceived discomfort from experiencing a specific side effect) across the different variations of the study.

It was predicted that older and younger adults may have equivalent overall recall performance on this task (similar to Castel, 2005; Miller, 2003), but the type of value that is used in each experiment (i.e., objectively labeled values or values the participant subjectively assesses) may influence younger and older adults' sensitivity to value in memory differently. Specifically, younger adults may show larger VDR effects when objective values are used, while older adults show larger VDR effects when subjective values used. The rationale for this claim is that younger adults may find the task more similar to educational contexts when objective values are used, allowing them to use their own prior experience to selectively encode high value information better. Conversely, older adults likely have more experience assessing the qualitative value of information (similar to remembering gist information; see Aizpurua & Koutstaal, 2010; Koutstaal, 2006; Koutstaal & Cavendish, 2006) and will show larger VDR effects when they have to place a subjective value on information. Work by McGillivray and Castel (2011) supports this claim – older adults tend to be more conservative on estimating what information they think they will remember (with practice and when given feedback), and as a result can predict their memory performance more accurately than younger adults.

Experiment 4

In order to explore differences between objective and subjective values in selectivity in a practical, everyday test of memory, younger and older adult participants studied and recalled a list of medication side effects. After recalling as many side effects as possible, participants' subjective ratings of the side effects were collected, in order to investigate how the value participants place on information (which likely varies from person to person) may impact their memory. Due to their prior experience and/or expertise in this domain, it is likely that older adults will remember as many side effects as younger adults because some older adults may be able to integrate their prior knowledge of side effects to support learning (e.g. Castel, 2005; Miller, 2003). However, since younger adults may have an interest in such medication-related information as well (DeSantis et al., 2008; Sollman et al., 2010), it is possible that both age groups would have better memory for important medication-related information relative to other to-be-learned material. Additionally, it is relatively unclear whether or not the subjective value participants will place on information will impact their memory as it would in a more traditional selectivity paradigm in which objective values (point values) were used.

Method

Participants. Twenty undergraduates (2 males, average age = 19.8) from the University of California, Los Angeles and 20 older adults (9 males, average age = 77.3) from the surrounding community were recruited for this task. Older adults were living independently in the Los Angeles area, and recruited through community flyer postings as well as through the UCLA Cognition and Aging Laboratory Participant Pool. Older adults were compensated with $10 an hour for participation, while younger adults received course credit for their participation.

Materials. The materials were a list of 20 common side effects taken from an internet database (http://www.vaughns-1-pagers.com/medicine/prescription-drug-side-effects.htm) for medication side effects (e.g., nausea, sweating, headache, etc.). All of the side effects that were selected for this experiment were temporary in their duration (although the selected side effects could occur chronically as well), and considered relatively mild.

Procedure. All instructions were read out loud to the participant by the experimenter. Participants were told they were going to be shown a list of side effects for a new medication, one at a time. After studying the last side effect, they would be asked to recall as many of the side effects as they could, and in any order. Participants saw each side effect in the center of the screen on a white background in black 44-point Arial font for 4 s before the program automatically advanced to the next side effect. Participants were randomly assigned to view the list of side effects in one of four different fixed-randomized orders. After studying the last side effect, participants had 1 min to verbally recall as many of the studied side effects as they could possibly remember. The experimenter recorded their verbal responses. After the recall task, participants were told they would be asked rate how negative or unpleasant they felt experiencing each side effect would be on a scale from 1 to 10, where 1 would not be an issue, while 10 would be an extremely big issue if they experienced that specific side effect. Participants then, in the order with which they originally studied the side effects, verbally rated how negative they felt each side effect was while the experimenter recorded their responses. Participants had unlimited time to provide their ratings. After completing the rating task, two post-test questions were asked of participants after completing the task: how did they try to remember the side effects they studied, and whether or not they pay attention to the side effects of medication they are currently taking or have taken in the past.

Results and Discussion

Recall and average rating assigned to the side effects based on age group are shown in Figure 17. While there was no difference between age groups for recall, older adults felt, overall, that the side effects would be more unpleasant than younger adults. This was confirmed using independent samples *t*-tests comparing younger and older adult recall and ratings. Older adults (M = .55, SD = .13) remembered as many side effects as younger adults (M = .60, SD = .12), $t(38) = 1.43$, $p = .16$. However, older adults (M = 6.4, SD = 1.62) gave higher negativity ratings on average for the side effects than younger adults did (M = 5.0, SD = .79), $t(38) = 3.50$, $p < .01$, $d = 1.11$. Based upon the post-test questions asked, 55% of younger adults said that they pay attention to the side effects of medication they take, and 50% of older adults said that they pay attention to side effects as well. One older adult reported that they gave up taking a specific medication because of its side effects. There were neither significant main effects nor interactions when subjects were grouped by whether or not they paid attention to side effects (*t*s and *F*s < 1).

In order to investigate the potential impact of participants' ratings (i.e., value) on their memory of specific side effects, participants' memory performance was plotted based on their standardized ratings, and is shown in Figure 18. The top panel of Figure 18 (A), shows the probability of recall based on participants' standardized ratings (within individuals), in order to control for participants having differing levels of variability within the ratings they gave for the side effects. The standardized ratings were then binned based on their percentile within a normalized distribution (i.e., the lowest 33%, the middle 33%, and the highest 33%). However, the associated 2 (Age Group: younger adults, older adults) x 3(Tritile: first, second, third) analysis of variance (ANOVA) did not reveal significant main effects nor a significant interaction. To further investigate this relationship between value and memory, a more robust trend analysis using multilevel modeling was then used to predict recall rates based on the associated ratings for those side effects, shown in the bottom panel of Figure 18 (B) and reported in Table 2. Specifically, ratings participants gave after the recall test were standardized (via a z-transformation, within individuals, to control for differences of variability and scale-use), matched, and compared with their recall performance for each associated side effect. For example, if a participant gave a particular side effect a relatively high rating, would the probability that they recalled that specific side effect at test be higher than another side effect they gave a relatively lower rating to? Log-odds[2] of recall was predicted from the following variables: the linear value of the participants' standardized ratings, the quadratic (or curvilinear) value of the participants' standardized ratings, age group (coded -1 for younger adults and 1 for older adults), and the associated interaction terms with age group, with participants being the level-2 unit. Inconsistent with predictions of value-directed remembering, both the linear and quadratic terms were non-significant in the model ($\beta = 0.06$, $p = .40$ and $\beta = -0.02$, $p = .73$, respectively). Importantly, there was not a significant effect of age group in the model either ($\beta = -0.12$, $p = .15$) nor did age interact with either rating variable (ps > .40), supporting the claim that older adults can remember the studied side effects as well as younger adults. However, the merits of this analysis should be taken skeptically. Since the ratings of the side effects were assessed post-recall, the assigned ratings may be confounded with recall performance. For example, a participant may have given a specific side effect a higher rating because they remembered it at test (see Castel, Rhodes, McCabe, Soderstrom, & Loaiza, 2012; Kassam et al., 2009). Despite this issue, the findings from this analysis may provide some indication as to whether or not value-directed remembering will be found in subsequent experiments when more salient cues for value are used, and are either subjectively or objectively defined.

Experiment 5

In Experiment 4, younger and older adults recalled equivalent amounts of the to-be-remembered side effects. The lack of age differences in memory performance is noteworthy, and likely driven by schematic support (Castel, 2005; 2007; Castel, McGillivray, & Worden, in press; Miller, 2003) However, neither younger or older adults recalled significantly more side effects that they rated as more negative (i.e., valuable) relative to less negative. Put differently, value-directed remembering was not found in Experiment 4. It may be that both age groups did not prioritize what was important during encoding, and simply tried to remember as many side effects as possible. The above explanation seems plausible, if not likely; since participants were not initially given instructions to remember the side effects based on their importance. The lack of sensitivity to value may also be due to the fact that the side effects were all relatively mild, and thus were not perceived by participants as having differences in level of importance or severity. If the "uniformity" of the materials contributed to the lack of sensitivity to value in memory, then increasing the "variability" (i.e., the severity) of the materials by using side effects that are moderate and severe in addition to mild in severity may increase the learners' sensitivity to value.

The specific procedure of Experiment 4 also may have contributed to the lack of sensitivity to value in either age group, as negativity ratings were assessed *after* recall had occurred. In typical selectivity paradigms, the value of information needs to be defined during (or prior to) learning. If the motivation to remember particular information (i.e., value) is placed on it after learning (e.g., during testing), the likelihood of that motivation influencing memory disappears completely, similar to if no motivation or value cue was given (Kassam et al., 2009). Additionally, the problem of post-test ratings in Experiment 4 is further confounded by ratings being potentially biased by recall performance (Castel et al., 2012). For example, a participant may have given a specific side effect a lower rating than they would otherwise have simply because they failed to remember that side effect during the prior free recall test.

There were two goals in designing Experiment 5. The first was to further investigate the impact of subjective value on recall by having participants rate the severity of side effects during learning, rather than after recall. If value assignment or assessment can only confer a benefit to memory during encoding, participants may be more sensitive to value in this experiment than they were in Experiment 4. The second goal was to create more variability within participants' subjective ratings by using side effects that could be classified into three "severity" categories: mild, moderate, and severe (as only mild side effects were used in Experiment 4). By manipulating the side effects in this way, participants should give quantitatively different ratings to each side effect category-type, as well as explore how those subjective ratings may impact their recall performance.

Method

Participants. Twenty-four undergraduates (5 males, average age = 20.3) from the University of California, Los Angeles and 24 older adults (11 males, average age = 73.5) from the surrounding community were recruited for this task. Older adults were recruited and compensated (i.e., $10 an hour) in the same manner as in Experiment 4, and younger adults were compensated in the same manner as Experiment 4 as well (i.e., course credit).

Materials. The materials were similar to those in Experiment 4, but with the following exception. Twenty-one side effects were selected from the same internet resource as Experiment 4. However, the selected side effects were classified into three separate severity categories: mild (e.g., itching), moderate (e.g., heartburn), and severe (e.g., stroke) side effects. Each side effect was categorized based on the classification of the side effects in drug advertisements, as well as by independent raters and the experimenters, into one of the three groups based on how severe each side effect was, such that there were seven side effects in each severity category.

Procedure. The procedure of Experiment 5 was similar to Experiment 4 with the following exceptions. During the learning phase of the experiment, participants were asked to rate the severity of each side effect on the same scale as Experiment 4, as each side effect was presented. Participants studied and provided a rating for each side effect during the 5 s it was on the screen, before the next side effect was shown. The order of the side effects was blocked randomized such that within each block of three items, one side effect from each of the severity categories was presented. The order of each block was randomized across four versions of the experiment.

Results and Discussion

Recall and average ratings assigned to the side effects based on age group are shown in Figure 19. Overall, older adults rated the side effects as more negative compared to younger adults, and participants' ratings closely matched the categories side effects were assigned to (i.e., "severe" side effects were rated more negative than moderate effects, and moderate effects were rated more negative than mild effects). A 2 (Age Group: younger adults, older adults) x 3 (Severity Category: mild, moderate, severe) mixed ANOVA on ratings found a marginal effect of age group such that older adults (M = 5.88, SD = 1.36) rated the side effects as more negative overall than younger adults (M = 5.30, SD = 0.73), $F(1, 46) = 1.43$, $MSE = 3.59$, $p = .07$, $\eta_p^2 = .068$. There was also a main effect of severity category, $F(2, 92) = 561.45$, $MSE = .68$, $p < .001$, $\eta_p^2 = .924$, such that severe side effects (M = 8.65, SD = 1.07) were rated more negative overall than moderate effects (M = 5.02, SD = 1.39), which were rated more negative than mild effects (M = 3.11, SD = 1.44). There was also a marginal interaction between age groups and severity categories, *F*(2, 92) = 2.09, *MSE* = .68, *p* = .13, such that older adults (*M* = 3.59, *SD* = 1.80) rated mild side effects as more negative than younger adults (*M* = 2.62, *SD* = 0.69), *t*(46) = 2.47, *p* = .02, *d* = .71. The comparisons between younger and older adult ratings for moderate and severe side effects were not different from one another (*t*s < 2).

Proportion of side effects recalled is shown on the right portion of Figure 19. Older adults remembered more severe side effects relative to mild or moderate effects, while younger adults had better recall for mild and severe effects (i.e., the extreme categories). A 2 (Age Group: younger adults, older adults) x 3 (Severity Category: mild, moderate, severe) mixed ANOVA on recall found no significant effect of age group, but an effect of severity group, *F*(2, 92) = 4.91, *MSE* = .04, *p* < .01, η_p^2 = .097, such that moderate side effects (*M* = .52, *SD* = .21) were recalled less often than mild (*M* = .61, *SD* = .20) or severe side effects (*M* = .64, *SD* = .19). Additionally, age group interacted with severity category *F*(2, 92) = 3.24, *MSE* = .04, *p* = .04, η_p^2 = .066, such that older adults had equivalent levels of recall for mild and moderate side effects (*t* = .02), but higher recall for severe effects (*t*s > 2.0). Younger adults showed a different pattern of results such that mild effects were recalled more often than moderate effects, *t*(23) = 3.36, *p* < .01, *d* = .96, while severe side effects were remembered marginally more often than moderate effects, *t*(23) = 1.97, *p* = .06, *d* = .62. Additionally, although a larger proportion of older adult participants reported they pay attention to the listed side effects of medications they are currently taking (58%) relative to younger adults (38%), however this factor did not influence performance on either dependent measure of ratings or recall (similar to the results of Experiment 4).

A trend analysis using multilevel modeling was conducted to further explore the relationship between participants' negativity ratings and likelihood of recalling a specific side effect. The pattern of results and associated beta coefficients are shown in Figure 20 and Table 2, respectively. In addition to the variables included from the Experiment 4 analysis, the model for Experiment 5 also includes the severity category of each side effect (coded -1, 0, and 1 for mild, moderate, and severe categories, respectively) and the associated interaction with age group variable. Consistent with the observation from the ANOVA, the quadratic term was significant ($\beta = 0.34$, $p < .001$), but had a marginally stronger influence on younger adults based on the associated interaction term ($\beta = -0.14$, $p = .09$). Importantly, there was a strong, but marginal influence of the linear term on recall ($\beta = 0.22$, $p = .07$), supporting the claim that valuable information (i.e., side effects rated more negatively) were more likely to be recalled than less valuable information. Replicating the finding from the analysis in Experiment 4, no influence of age group was found on recall ($\beta = -0.08$, $p = .28$).

Additionally, while the trend lines in Figure 20 are peculiar for younger adults, in that mild side effects are notably higher (i.e., more likely to be remembered) than moderate or severe effects, that result is due to the range of the scale younger adults used for mild, moderate, and severe effects. Specifically, the average standardized rating provided for mild side effects by younger adults was a z-score of -3.67, and based on the associated beta coefficients, the average log-odds of recall for a mild side effect with that standardized rating would be 6.58 (a high probability of recall). However, the average standardized rating provided by younger adults for severe side effects was 4.32, making the average log-odds of recall, based on the beta coefficients, 9.75 (a relatively equivalent probability of recall). Thus, even though the trend line is “higher” for mild side effects compared to moderate or severe effects when standardized ratings are equal to zero, the range of standardized ratings that younger adults used for each category of side effects explains why the probability of recall is equivalent for mild and severe effects.

The results of this experiment illuminate an interesting relationship between subjective ratings (or value), age group, and memory. Specifically, older adults remembered more severe side effects than mild or moderate side effects. This pattern of results was found in both the ANOVA, as well as the trend analysis. However, younger adults showed a different pattern than older adults, in that they recalled as many mild side effects as they did severe side effects, as reflected in both the ANOVA and the quadratic-age group interaction coefficient from the trend analysis. This curvilinear relationship between value and memory has been found in other research (Castel, Friedman, McGillivray, Flores, & Murayama, in preparation), but could be explained by two alternatives. The first explanation is that younger adults have better memory for extreme values, rather than only the high-value information like older adults (reward salience and reward value, respectively; see Madan & Spetch, 2012). This explanation is plausible, as participants were never provided with instructions to remember specific, important, side effects over less important ones. The alternate explanation is one of relative experience. Namely, younger adults likely have had less experience with the moderate or severe side effects relative to the mild ones. As a result, younger adults may have elected to focus their encoding and rehearsal efforts on the side effects they had experienced in the past (i.e., the mild ones). Several participants (both younger and older) reported post-test that they tried to remember side effects they had previously experienced, giving this explanation some validity. Conversely, older adults may have relatively more experience with side effects from all three categories, which could explain why they selectively focused on only the severe side effects. Another alternative may be that older adults elected to focus on the severe side effects, because they could not remember everything that was studied, so they chose to focus on the side effects they felt were most important.

Experiment 6

Experiment 4 showed that older adults are as capable of recalling side effects as younger adults, and rate the side effects as more negative based upon post-test subjective value ratings. Experiment 5 showed that both younger and older adults have significantly higher recall for side effects they considered to be severe than mild (as assessed by their subjective ratings during learning). However, using arbitrary values to indicate importance, as in Experiments 4 and 5, may not truly reflect how people normally assess value in their everyday lives. Additionally, Experiments 4 and 5 did not instruct learners to remember the side effects based on their importance or value. Put differently, no instructions to be selective were given to participants. The previous experiments also only investigated participants' episodic memory for the side effects, so it is relatively unclear whether or not value may impact other types of memory (e.g., source memory).

Experiment 6 was designed to investigate younger and older adults' episodic and source memory for valuable or important information, when seemingly mild side effects suddenly become important to remember. Like in Experiment 4, participants studied a list of mild side effects; however a subset of effects was identified during study as important to remember by a warning of a severe complication with the medication being associated with those side effects. Participants were instructed they would need to contact their doctor immediately if they experienced any of those specific side effects, as they were indicative of a serious complication with the medication (labeled "contact-your-doctor side" effects, or "CYDs"). Many contemporary medication advertisements come with similar warnings that seemingly mild side effects may be indicative of a severe complication, making relatively low-value information suddenly important to remember. Thus, some relatively mild side effects (e.g., sore throat) may be critical to remember, as they are indicative of a potentially life-threatening condition. Memory for the side effects was assessed using a free recall test, but was then followed by a source memory task, in which participants had to identify which of the side effects were indicative of the severe complications (i.e., the CYD effects).

It is likely that older adults will have greater difficulty with the source memory task compared to younger adults. Specifically, a lifetime of experience with the mild to-be-remembered side effects likely will lead to a greater degree of proactive interference at identifying the CYD effects. After all, why should a side effect like a cough, which an older adult has viewed their entire life as common, chronic, and not important, suddenly become more important to remember than back pain? Such a finding may illustrate one potential cost of schematic support in older adulthood, in that older adults may be able to remember as much information as younger adults, yet may have an inability to update the value of information in memory due to extensive proactive interference. Work by Jacoby and colleagues have shown that older adults have relatively higher false alarm rates than younger adults due to a reliance on familiarity and not recollection, which makes the above prediction seem likely (Jacoby, 1991; Jacoby, Bishara, Hessels, & Toth, 2005; Jacoby, Walheim, Rhodes, Daniels, & Rogers, 2010; Jones & Jacoby, 2005; Rhodes, Castel, & Jacoby, 2008).

Method

Participants. Twenty-four undergraduates (10 males, average age = 21.5) from the University of California, Los Angeles and 24 older adults (11 males, average age = 71.2) from the surrounding community were recruited for this task. Older adults were recruited and compensated (i.e., $10 an hour) in the same manner as in Experiments 4 and 5, and younger adults were compensated in the same manner as in Experiment 4 and 5 as well (i.e., course credit).

Materials. The materials were similar to those in Experiment 4 with the following exception. Of the 20 side effects used in Experiment 4, seven of the side effects were removed and replaced with alternate mild side effects. These effects were replaced because some participants in Experiment 4 thought the side effects were more serious than they actually were (e.g., heart palpitations), or were too vague (e.g., muscle pain). The replacement side effects were equivalently mild to moderate and temporary in duration (e.g., back pain, itching, etc.).

Procedure. The procedure was identical to Experiment 5, in that side effects were studied and rated simultaneously before a free recall test. However, participants were instructed that a subset of the side effects that they studied, if experienced, were indicative of a serious complication, and that they would need to contact their doctor immediately if they experienced any of those specific symptoms (contact-your-doctor side effects, CYDs). That is, some of the to-be-remembered side effects were objectively valuable and important for the participant to remember relative to others. Those objectively severe or CYD side effects were a subset of five of the 20 side effects, randomly selected across the entire list and counterbalanced across participants. The CYD side effects and were shown in red font rather than black font to simplify the learner’s ability to identify the CYD effects during study. The order of the side effects studied was blocked randomized such that within each block of 4 side effects, one side effect was identified as a CYD effect. Across participants, each of the 20 side effects was equally labeled as a CYD effect, such that there were four different versions of the study phase of the experiment (with five different side effects labeled as CYD effects within version).

Following the 1 min free recall test, participants then engaged in a source memory test to identify the five CYD side effects out of the 20 side effects initially studied. The experimenter read aloud the name of each side effect to the participant in the order in which they were presented during the learning portion of the experiment. As each side effect was read aloud, the participant had to respond by saying “yes” or “no” as to whether or not that specific side effect was one of the CYD side effects. After the participant made their judgment on the last side effect, he or she proceeded to answer the post-test questions before being debriefed.

Results and Discussion

Recall and average ratings assigned to the side effects based on age group and severity category are shown in Figure 21. As in Experiments 4 and 5, on average older adults gave higher ratings for the l side effects relative to younger adults, and both age groups gave higher ratings to CYD side effects than non-CYD effects. A 2 (Age Group: younger adults, older adults) x 2 (Warning Category: non-CYD, CYD) mixed ANOVA on ratings found an effect of age group $F(1, 46) = 4.85$, $MSE = 3.63$, $p = .03$, $\eta_p^2 = .095$, such that older adults ($M = 6.76$, $SD = 1.38$) gave higher ratings to the side effects than younger adults ($M = 5.91$, $SD = 1.32$). There was a main effect of warning category $F(1, 46) = 13.12$, $MSE = 1.52$, $p = .001$, $\eta_p^2 = .222$, such that CYD side effects were rated more negatively ($M = 6.79$, $SD = 1.21$) than non-CYD side effects ($M = 5.88$, $SD = 1.06$). There was also a marginal interaction between age group and warning category $F(1, 46) = 2.20$, $MSE = 1.52$, $p = .15$, $\eta_p^2 = .046$, such that older adults gave higher ratings than younger adults for non-CYD side effects, $t(46) = 2.84$, $p = .01$, $d = .82$, but there was no such difference between age groups for CYD side effects ($t < 1$).

The proportions of side effects recalled based on age group and warning category is shown on the right portion of Figure 21. Older adults remembered as many side effects as younger adults, while warning category only had a minor influence on recall. A 2 (Age Group: younger adults, older adults) x 2 (Warning Category: non-CYD, CYD) mixed ANOVA on recall found only a marginal effect of warning category $F(1, 46) = 2.51$, $MSE = .04$, $p = .12$, $\eta_p^2 = .052$, such that side effects that were categorized as CYD ($M = .58$, $SD = .16$) were recalled slightly more often than side effects not categorized as CYD ($M = .52$, $SD = .11$). Both the main effect of age group and the associated interaction between age group and warning category were non-significant.

A trend analysis using multilevel modeling was also conducted, using the same parameters as that in Experiment 5, to explore the relationship between participants' ratings and the likelihood of recalling a specific side effect. The pattern of results and associated beta coefficients are shown in Figure 22 and Table 2, respectively. Warning category was recoded for this experiment to reflect only two levels of the variable (-1 and 1 for non-CYD and CYD side effects, respectively). Consistent with the ANOVA, there was a marginal influence of warning category on recall ($\beta = 0.12$, $p = .16$), supporting the claim that CYD side effects (i.e., valuable information) would be more memorable than non-CYD side effects. However, all other variables, including age group and the associated interactions with the age group variable, were non-significant (*p*s > .20).

Mean proportions of side effects identified as "CYD" on the post-recall recognition test are shown in Figure 23. Despite similar performance on free recall, older adults correctly identified CYD side effects (i.e., hits) less often than younger adults, yet misidentified more non-CYD side effects as CYD (i.e., false alarms) than younger adults as well. A 2 (Age Group: younger adults, older adults) x 2 (Measure: hits, false alarms) mixed ANOVA revealed a main effect of measure $F(1, 46) = 152.07$, $MSE = .04$, $p < .001$, $\eta_p^2 = .77$, such that participants overall had more hits (*M*

= .60, *SD* = .26) than false alarms (*M* = .10, *SD* = .12). Critically, age group interacted with measure $F(1, 46) = 8.58$, $MSE = .04$, $p < .01$, $\eta_p^2 = .16$, such that older adults (*M* = .52, *SD* = .22) had significantly fewer hits than younger adults (*M* = .68, *SD* = .27), $t(46) = 2.35$, $p = .02$, $d = .65$. However, the opposite pattern was found for false alarms $t(46) = 2.22$, $p = .03$, $d = .63$, such that older adults (*M* = .13, *SD* = .13) misidentified more side effects as CYD than younger adults (*M* = .06, *SD* = .09). The above age group by measure interaction supports the claim that older adults are more reliant on using familiarity and not recollection when recognizing important information (e.g., Jacoby, 1991) and illustrates that, while older adults are sensitive to remembering information based on its value (i.e., in free recall), they also have difficulty binding that importance to specific information (the associative deficit hypothesis; see Naveh-Benjamin, 2000). Lastly, although twice as many older adult participants reported they pay attention to the listed side effects of medications they are currently taking (67%) compared to younger adults (33%), this factor did not influence performance on either dependent measure of ratings or recall, nor performance on the post-recall recognition test.

The results of Experiment 6 have some similarity to the findings from Experiments 4 and 5. Importantly, there was no effect of age group on recall of the side effects. However, there were differences between objective (i.e., CYD warnings) and subjective value (i.e., participants' negativity ratings) and their impact on recall. Namely, CYD side effects were better recalled than non-CYD side effects (i.e., objective value), as assessed by the ANOVA. However, there was no apparent influence of participants' severity ratings (i.e., subjective value) on recall, based on the results from the trend analysis. This discrepancy between objective and subjective value, and possible explanations for the result, is further discussed in the General Discussion below. Lastly, hit and false alarm performance on the severity identification task provides further evidence for gist-based memory in older adults (Koutstaal & Schacter, 1997) and also use of familiarity contributing to false recognition in older adults (Jacoby, 1991; Rhodes et al., 2008). Younger adults were able to correctly identify more severe side effects than older adults (i.e., hits), while older adults incorrectly identified more non-severe side effects as severe compared to younger adults (i.e., false alarms). This result is surprising given that older adults recalled as many severe side effects as younger adults during the free recall task that occurred immediately prior to the task. This result can be explained through the combination of value-directed remembering and gist-based memory (see Castel, Farb, & Craik, 2007, Friedman & Castel, in press), as well as the reliance on familiarity leading to false recognition for older adults (Jacoby, 1991; Jacoby et al., 2005; 2010; Jones & Jacoby, 2005; Rhodes et al., 2008). In summary, while older adults are capable of remembering high-value or important information better than less important information, monitoring the specific values associated with that information may quickly deteriorate in older adults (similar to an associative deficit, Naveh-Benjamin, 2000), making source monitoring judgments difficult (Chalfonte & Johnson, 1996; Johnson, Hashtroudi, & Lindsay, 1993).

General Discussion

The overall goal of this line of research was to explore the impact of different value types (i.e., objective and subjective) and the occurrence of VDR in different age groups in the context of memory for medication side effects. Experiment 4 was a simplistic design in which younger and older participants studied, recalled, and rated the severity of mild and common side effects. Experiment 5 manipulated the objective severity of the side effects, by introducing more variability within the studied side effects, while Experiment 6 manipulated the objective value of specific side effects by indicating whether or those side effects were associated with a severe complication with the medication. Across all three experiments there was no effect of age on recall – older adults remembered as many side effects as younger adults did. Importantly, participants were sensitive to value in their memory for the side effects in both Experiments 5 and 6. Participants were not sensitive to value in Experiment 4, but since all of the side effects used in that experiment were uniform in their severity, and no instructions were given to remember side effects based on their value or importance, the lack of a VDR effect may not be surprising. In Experiment 6, post-recall identification of the "severe" effects found the only age-related effect on memory – older adults had a greater frequency of false alarms and fewer hits than younger adults did.

The lack of age-related differences on recall is likely due to older adults relying upon schematic support, or their relatively greater experience with medication and associated side effects, to make up for age-related declines in memory in this domain (Castel, 2005; 2007; Castel et al., in press; Miller, 2003; Morrow, Leirer, Altieri, & Fitzsimmons, 1994). Across all three experiments reported here, older adults rated the side effects as more severe overall compared to younger adults. Based on these relatively higher ratings, it stands to reason that older adults could have found the to-be-learned material to be more valuable than younger adults did, potentially resulting in no age differences on performance. Regardless, short-term memory for side effects appears to be one domain in which older adults do not show age-related declines. Since the need for medication use increases in later adulthood, the need for additional awareness in medication-related issues such as potential side effects is critical (Hines & Murphy, 2011; Moore et al., 2007; Qato et al., 2008). Practically, this work may help doctors and medical providers better inform their patients about their prescribed medications, and help to prioritize treatment plans based on what symptoms or side effects the patient views are more severe or adverse (as side effects people perceive as more negative appear to be better remembered). For example, a patient that fervently hates the experience of nausea should not be prescribed a medication by their doctor in which nausea may result from its use, thus increasing the patient's likelihood of taking the medication properly and advancing their recovery.

The source memory test in Experiment 6 provides support for the idea that older adults rely on familiarity when retrieving information (Jacoby 1991; Jacoby et al., 2005; 2010), in that they incorrectly identified significantly more side effects as the critical CYD effects compared to younger adults, and may illustrate one potential cost of schematic support. Specifically, older adults may have greater difficulty updating the value or importance of to-be-remembered information (Friedman & Castel, in press), and creating new associations with that information (Naveh-Benjamin, 2000), especially if competing yet irrelevant information has to be suppressed (inhibitory deficits; Hasher & Zacks, 1988; Lustig, Hasher, & Zacks, 2007). It is possible that the source monitoring test used in Experiment 6 may have been too difficult, in that older adults may have struggled to successfully update the relative low value of specific mild side effects to be temporarily important. The "incongruency" between those two types of value should be further investigated, by incorporating the source monitoring test and the CYD warning manipulation into the materials used in Experiment 5 (i.e., side effects of differing levels of severity). Specifically, would older adults be able to correctly identify the important CYD effects with greater accuracy if they were severe side effects also or would they false alarm with greater frequency?

In relation to value-directed remembering, both younger and older adult showed sensitivity to value in their recall of the medication side effects. In Experiment 5, VDR was found when the severity of the side effects was manipulated (i.e., mild, moderate, and severe). In Experiment 6, VDR was found when value was objectively manipulated when warnings about severe complications were associated with specific side effects. Conversely, learners did not show VDR when all of the to-be-remembered side effects were uniformly mild and no warnings were associated with any of the side effects (Experiment 4). The above pattern of results across these three experiments may provide theoretical insight as to what conditions are critical or necessary for value to influence memory. Three conditions appear to encourage people to be sensitive to value in memory. The first is that a sufficiently broad range of values that encourages learners to be selective in what they encode seems necessary to some degree, based on either the explicit association of value to the information or by the implicit value naturally associated with specific information. The second component is that learners may need instructions that specific to-be-learned material is more important to remember than other material for an upcoming test of memory (agenda-based regulation of learning; see Ariel, Dunlosky, & Bailey, 2009). The last condition is that learners may choose to ignore value, if they think they will be able to remember all of the to-be-learned material (i.e., perform at ceiling).

Future work should examine value in other ways, in order to fully explore how the impacts of objective and subjective values may differ on memory. For example, in a modification of the paradigm used in this work, value could be manipulated quantitatively (e.g., 75% of users reported itching when using this medication) or qualitatively (e.g., severe itching). Such manipulations would further tease apart the different influences of objective and subjective value on memory. Also, the current study examined short-term recall, and greater retention intervals will likely lead to decrements in memory for both age groups, but may be more ecologically valid as medication is usually taken on the order of weeks or months. Additional work could also explore whether subjective value has an impact on memory in other domains (e.g., learning trivia facts or information on subjectively interesting or non-interesting topics), as well as other medically-related domains to see if age-related declines in memory are reduced, if not non-existent, more broadly in the area of medicine. For instance, value could be creatively manipulated by teaching learners how to identify negative interactions between various medications. Ultimately, the need for additional work in this domain (memory for medically-related information) is important, as it has the potential to benefit both younger and older adults alike. The research in this chapter provides further insight into how value can play a role in memory in our everyday life, and permeates our thoughts and actions in yet to be revealed ways. Importantly, these experiments show that prior experience or schematic support may reduce age-related declines in memory more than perhaps would be expected in a laboratory study of memory, and older adults may need specific conditions and processes to selectively remember important side effect information, especially if that information is valuable to remember.

Chapter 4: Summary and Future Directions

Across six experiments, the impact of value or importance on learning and memory was examined. The goal of Chapter 2 (Experiments 1-3) was to investigate whether or not value-directed remembering (VDR) could occur in the context of category induction (Kornell & Bjork, 2008), similar to how value can influence episodic recall. In each of the three experiments reported in Chapter 2 VDR was found, but only when categories were studied in interleaved schedules, and not in blocked schedules. Based on post-test responses, a large majority of participants reported that they thought that blocked schedules of learning were more effective than interleaved schedules, despite participants actually performing better when the categories were studied in an interleaved schedule. Subsequent research (Experiments 2 and 3) investigated whether participants would give similar responses during the actual learning of the categories (Judgments of Learning, JOLs) and during testing (confidence ratings). In both experiments, the impact of learning schedules on metacognitive judgments was small, if non-existent. Interestingly, in both Experiments 2 and 3, participants' accurately predicted the influence of value on learning, illustrating how value or the importance attached to information may serve as a extrinsic cue that can accurately guide metacognitive judgments (Koriat, 1997).

The goal of Chapter 3 (Experiments 4-6) was to investigate whether there would be any age-related differences in memory for medication side effects, as well as whether or not different types of value (objectively defined or subjectively rated by participants) would impact younger and older adults differently. In all three experiments, older adults recalled as many side effects as younger adults, reflecting the use of schematic support in later adulthood to assist with learning (Castel, 2005; 2007; Castel et al., in press; Miller, 2003). Additionally, participants were sensitive to value based on both their subjective ratings (Experiment 5) and when value was defined by critical complications associated with specific side effects (Experiment 6). Interestingly, when participants were given a source memory test in Experiment 6, older adults had significantly fewer hits and significantly more false alarms than younger adults, reflecting a potential cost of schematic support in later adulthood and the use familiarity during retrieval (Jacoby 1991; Jacoby et al., 2005; 2010).

The research reported in this thesis illuminates several important findings associated with encoding information based on its value or importance. Firstly, Experiments 1, 2 and 3 illustrate that the influence of value on memory is robust, and not only restricted to episodic memory. VDR can also occur in category learning. The fact that learners were sensitive to value, a process thought to be cognitively active and effortful since it requires selective encoding and rehearsal strategies, is relatively striking in these experiments since VDR occurred in the context of category learning, a particular type of learning thought to be less cognitively active (i.e., require little working memory; Ashby et al., 2002). Importantly, the finding that VDR occurs outside of episodic memory highlights that goal-directed learning may not be restricted to episodic memory tasks only. Future research should investigate whether or not VDR would occur in other types of category learning (i.e., rule-based categories), as well as other forms of learning (e.g., motor learning). If VDR was found in these other types of learning as well, it may reflect that selective prioritization processes may help to improve learning or performance, by having learners focus on critical components of what they are trying to master, one component at a time.

From Experiments 1, 2, and 3, it also appears that learners are most sensitive to value when items associated with different levels of value (i.e., low, medium, and high value) are interleaved together rather than when different levels of value are blocked. The fact that VDR occurred only in interleaved schedules of learning for all three experiments supports the idea that selective encoding and rehearsal are most effective when high value information is learned with low value information, rather than have all of the high value information be presented together at once. The claim that VDR is best under conditions of interleaved study needs to be examined with word-value pairs and episodic memory (as opposed to category learning and induction), but this claim does make intuitive sense. If all of the to-be-remembered valuable information was studied in one large block, it would become progressively difficult to rehearse specific items as other high value items would compete for attention immediately. Similar to the influence of presentation rates on serial position effects, information in long-term memory (i.e., primacy items) is remembered in lower quantities if the presentation rate is relatively fast (Rundus, 1980). In the context of a selectivity task, if all of the high value information is studied together it is analogous to rapid presentation, making the encoding and rehearsal of valuable information difficult.

In relation to metacognition, Experiments 2 and 3 illustrate that learners can use value as a salient extrinsic cue to monitor memory accurately (Koriat, 1997), in the sense that learners gave relatively higher JOLs and confidence ratings for valuable items and both experiments showed an impact of value on performance. Importantly, the sensitivity learners have to value in their monitoring is not restricted only to learning (Experiment 2), but retrieval as well (Experiment 3). However, the finding that learners are influenced by value at retrieval may not be that surprising, given previous work on motivation during retrieval (Kassam et al., 2009; Soderstrom & McCabe, 2011), yet the results from Experiment 3 show that learners are sensitive to value even when the values are not present during retrieval. Given the findings from Experiments 2 and 3, it appears as though learners are aware of the influence value can have on learning to some degree, and can use value as an effective guide to monitor (and by extension, improve) learning. Future work should further investigate the role of value as a salient cue for the basis of metacognitive judgments, as goal-directed learning appears to influence and guide study behaviors (Ariel et al., 2009; Castel et al., 2013).

From Chapter 3 (i.e., Experiments 4-6), it appears that older adults can remember medication side effects as well as younger adults can, at least in the context of short-term memory. Since the task used in these experiments was relatively more practical than a typical laboratory test of memory, it may not be surprising that no age differences in performance were found, since this task may better reflect how older adults retain information in everyday life (Zacks & Hasher, 2006). Specifically, older adults may rely on prior experience, or schematic support, to better remember information (Castel, 2005; 2007; Castel et al., 2013; Miller, 2003). However, the use of schematic support in memory may come at a cost, and lead to poor source monitoring (Experiment 6). Interestingly, all three experiments in Chapter 3 found that older adults rated the side effects to be more negative than younger adults, which may reflect older adults' general interest in this domain (and potentially why no age-related differences in memory were found). Future work should aim to find other domains in which older adults have relatively high interest, and examine whether age-related declines in memory for that information are absent as well (e.g., memory for weather-related information).

Based on the results from Chapter 3, it appears that the range or frequency of values used in a selectivity experiment is relatively important for VDR. Specifically, if the range of values used is relatively small, learners will be unable to identify the "important" information (regardless of value being either objectively or subjectively defined), and have difficulty rehearsing information. Since all information would be equally valuable to remember, learners would likely try to encode everything and not be selective. Critically, valuable information needs to be qualitatively different from the rest of the to-be-learned information in some aspect (either explicitly defined or intrinsically different). Subsequent research should try to investigate how critical the frequency of values is for VDR. For example, do learners show more robust VDR effects when values are on a continuous range or categorically binned (values 1-10 points and 1 vs. 5. vs. 10 points, respectively)? Additionally, is it the relative difference between values that impact VDR or the absolute differences between values that impact VDR (1 vs. 2 vs. 3 points and 1 vs. 5 vs. 10 points, respectively)?

Finally, it should be emphasized that VDR occurs only when it is needed. In Experiments 1, 2, 3, and 6, participants were given a goal to study the to-be-learned information based on its associated value and VDR was found in each experiment. In Experiments 1, 2, and 3, participants were told to pay attention to categories with higher point values over those with lower point values, while in Experiment 6 participants were told to remember the medication side effects associated with life-threatening complications (i.e., CYD effects). The results of these experiments support that learners do take goals into account during encoding and can remember information based on its associated value (Ariel et al., 2009). By comparison, Experiment 4 did not show any VDR effects on memory performance, but this finding may not be surprising since no instructions to attend to valuable or important information were given. However, no instructions to attend to valuable information were given in Experiment 5, yet participants were sensitive to value in their memory performance. While conjectural, the VDR effect found in Experiment 5 may be due to participants rating specific side effects as relatively more negative and unpleasant (e.g., stroke, seizure, etc.), and then the participants internally elected to try and remember the relatively severe side effects over the more mild ones. If such an explanation were true, then learners may spontaneously choose to be selective under certain contexts, even if no selectivity instructions are given to them.

To conclude, learners’ sensitivity to valuable or important information is a critical component in learning and memory. In the context of education, value-based encoding may help students better attend to critical material for upcoming examinations, especially if the student can acquire what information their instructor finds to be important. Additionally, value or importance may help students monitor their learning accurately, and assist them in allocating their study time and resources more effectively. Value may also help older adults remember important information precisely, make up for age-related declines in memory, and assist people in aging successfully. Although there are many research questions that remain to be answered, the work detailed in this thesis illuminates the potential of value-directed remembering in learning, and the critical components that may help facilitate the selective encoding and rehearsal of valuable information.

Footnotes

[1]The results of Experiments 1, 2, and 3 could be interpreted relatively differently if the number of items correctly identified was compared to the number of items incorrectly identified (i.e., difference scores). For example, the effect of value on interleaved categories would be relatively small, if not non-existent, in Experiment 1, while performance would be progressively lower as value increased for blocked categories.

[2]Log-odds are a different metric for expressing probability values. The trend analyses reported in Experiments 4, 5, and 6 use log-odds – a probability value converted into odds, before being converted into log-odds. Unlike probability, log-odds values can range from $-\infty$ to $+\infty$. As a shorthand rule, a probability of .25 is the equivalent log-odds value of -1.10, a probability of .50 is the log-odds value of 0, and a probability of .75 is the log-odds value of 1.10.

Table 1
Factors and manipulations could modify the value-directed remembering effect in a selectivity task

Increase Task Difficulty	Better Future Selectivity Performance	Magnitude of Value-Directed Remembering Effect
Excessive number of to-be-remembered items on each list	Multiple study-test lists	Value scale ▪ Categorical or continuous values ▪ Use of negative values
Rapid presentation rate	Providing feedback on performance	
Distractions during learning		Test format ▪ Recall or recognition ▪ Delay between study and test

Table 2
Trend Analysis Beta Coefficients

Measure	Experiment 1	Experiment 2	Experiment 3
Intercept	0.30**	0.38**	0.12
Age Group	-0.12	-0.08	0.09
Standardized Rating	0.06	0.22*	0.06
Standardized Rating * Age Group	-0.06	0.03	-0.02
Squared Standardized Rating	-0.02	0.34**	0.01
Squared Standardized Rating * Age Group	-0.03	-0.14*	-0.02
Severity Category (Exp. 2) or Warning Category (Exp. 3)	--	-0.22*	0.12
Severity or Warning Category * Age Group	--	0.18	0.07

* $p < .10$
** $p < .05$

Note. Age group was coded as -1 and 1 for younger and older adults, respectively. Standardized and squared (or quadratic) standardized ratings were coded as z-scores, standardized within individual participants. Severity category was coded as -1, 0, and 1 for mild, moderate, and severe side effects, respectively, for Experiment 2 while in Experiment 3 severity category was coded -1 and 1 for non-severe and severe side effects, respectively. Experiment 1 did not include any severity category manipulation. Below each main effect is the associated effect by age group interaction.

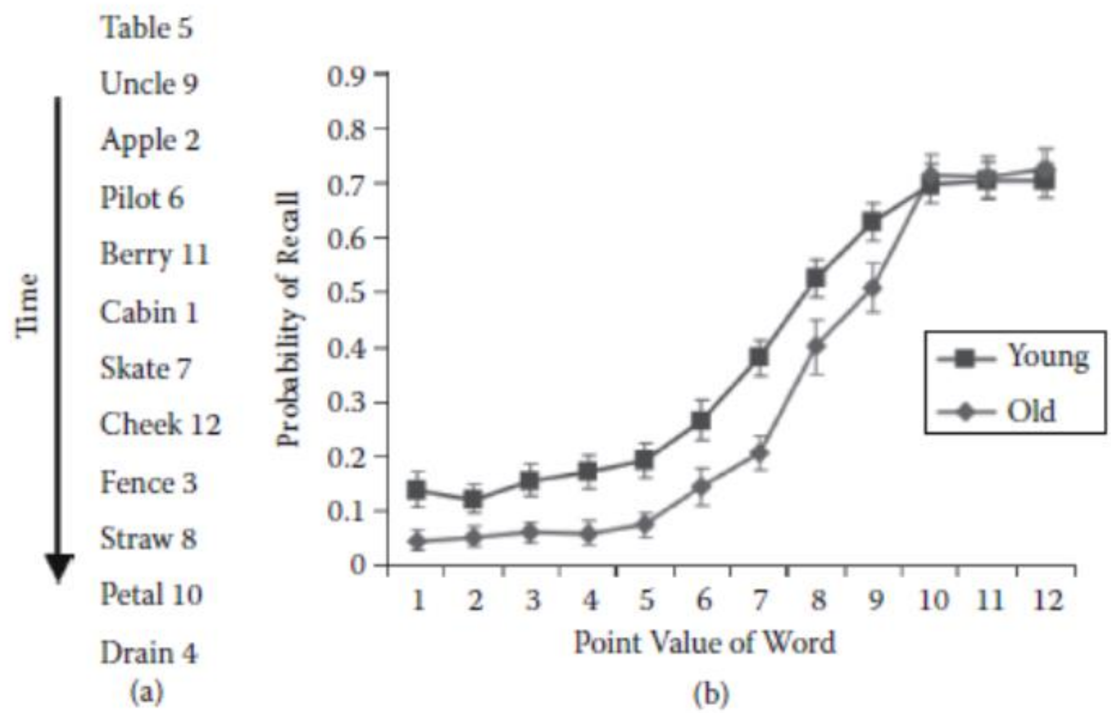


Figure 1. The procedure (a) and results (b) of a basic selectivity paradigm. Panel A: The participants are sequentially presented with words, each paired with a value. Participants recall the words with the goal of maximizing their score. Panel B: The average probability of recall plotted as a function of point value for younger and older adults (Adapted from Castel et al., 2002).

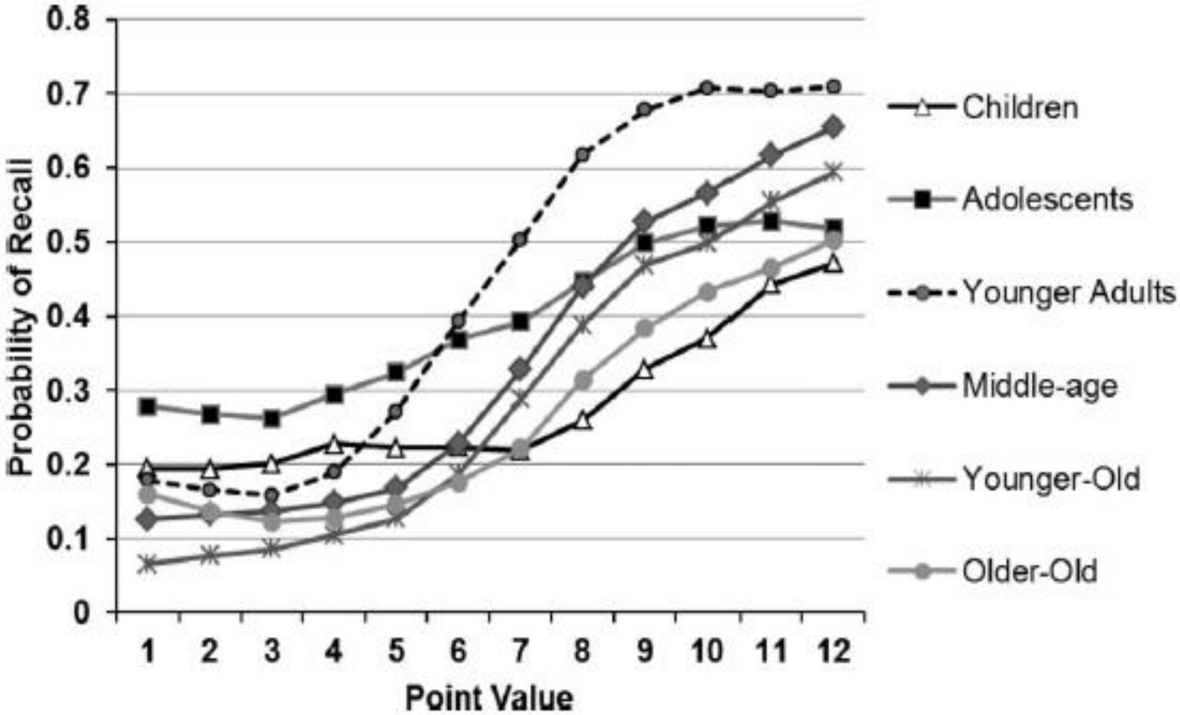


Figure 2. The probability of recall plotted as a function of value for different age groups across the human lifespan (adapted from Castel et al., 2011).

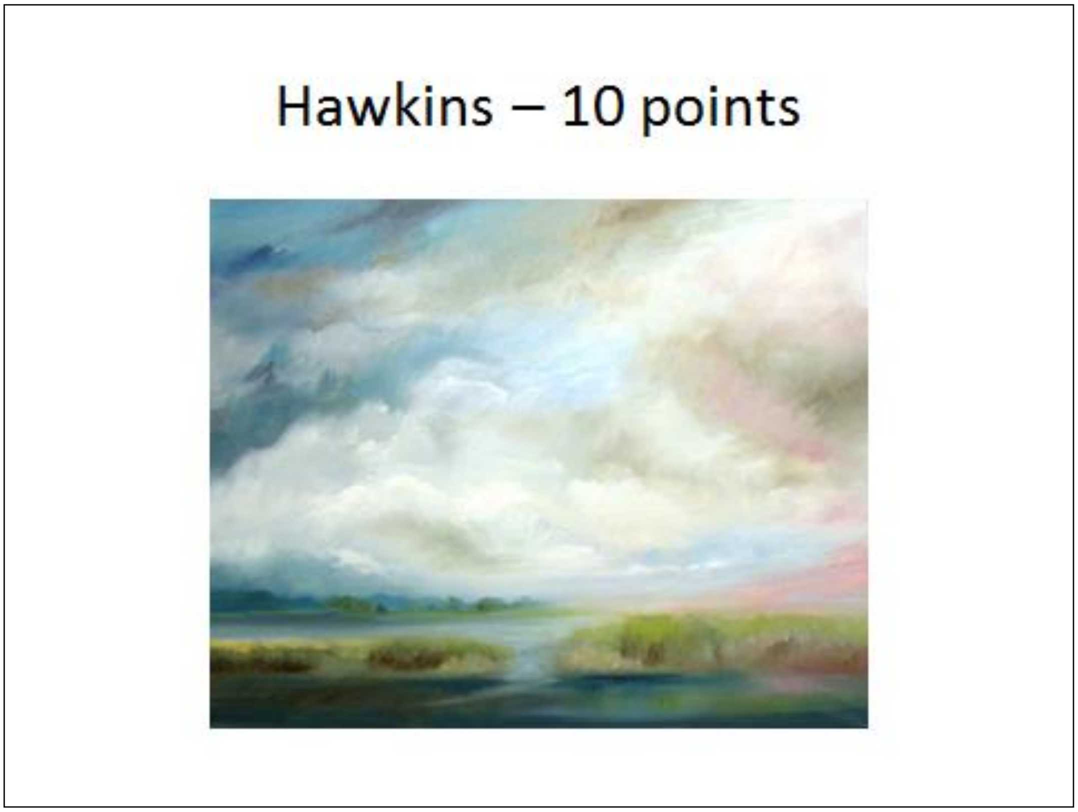


Figure 3. An example of the stimuli shown during the learning phase of Experiment 1.

Figure 4. An example of the stimuli shown during the test phase of Experiment 1.

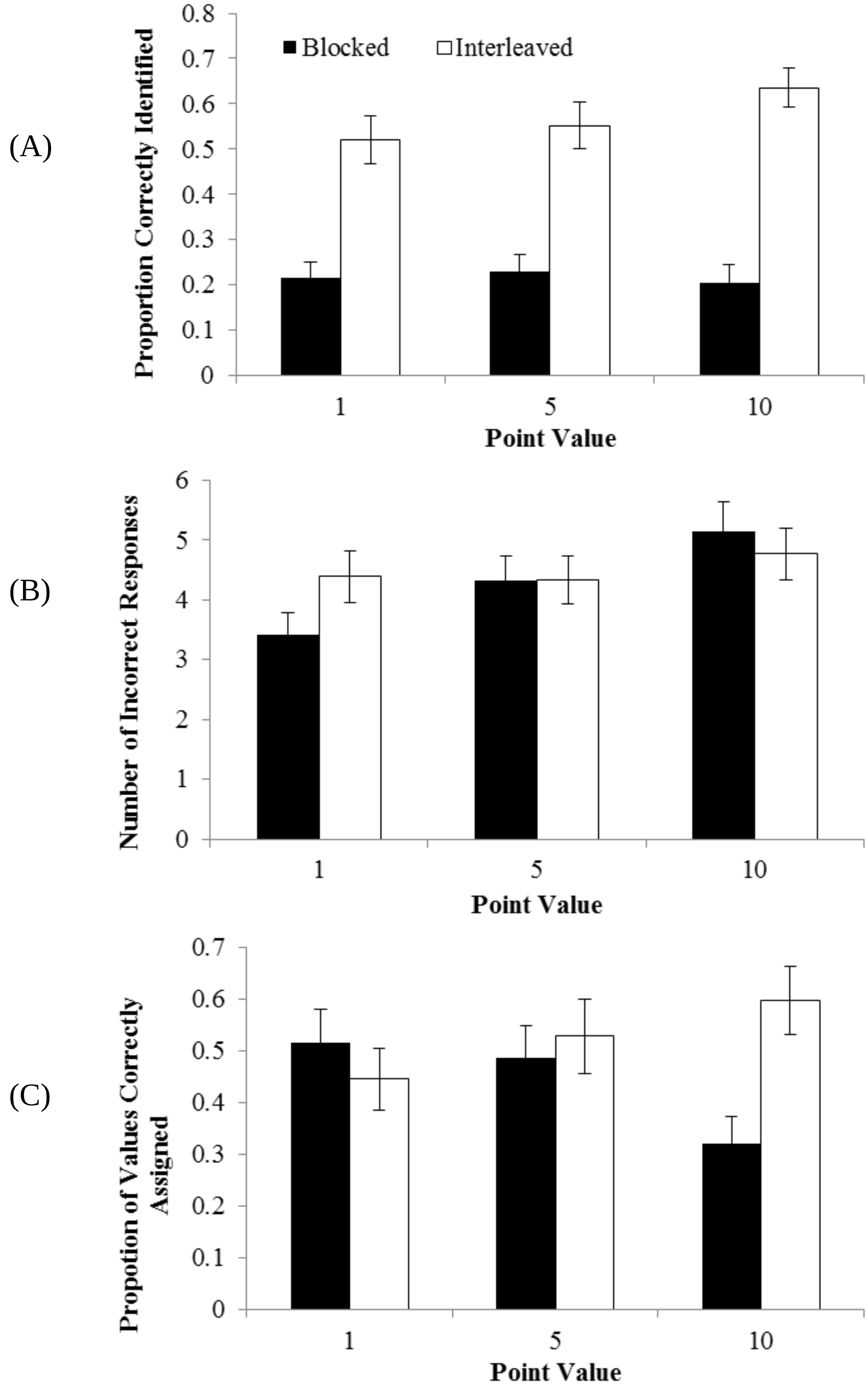
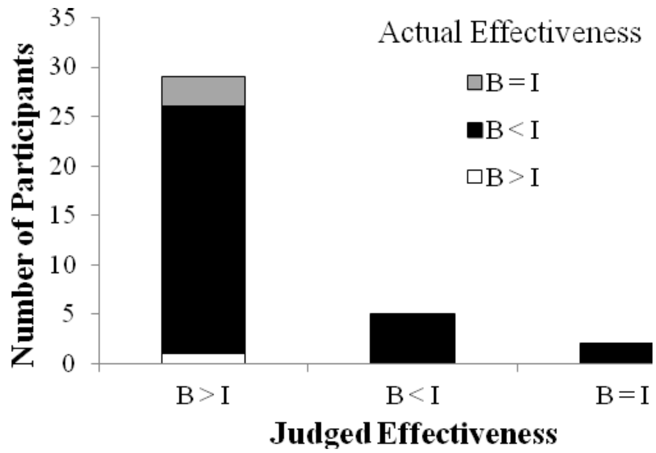


Figure 5. The average performance levels of participants that were assigned to take the induction test in Experiment 1 based on study schedule and point value conditions. (A) shows participants' performance on the induction test, (B) shows the number of incorrect responses participants made, and (C) shows the proportion of values correctly assigned by participants based on the artists, post-test. All error bars represent standard error values.

(A)



(B)

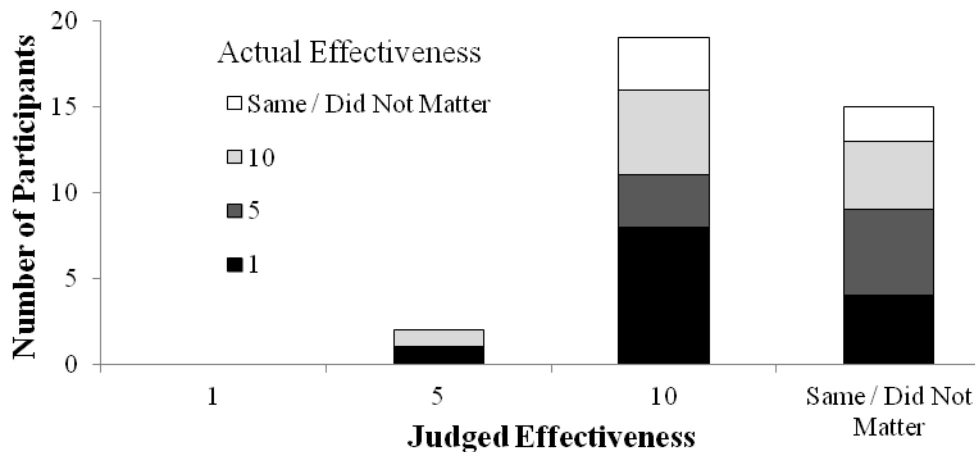


Figure 6. The reported judged effectiveness of participants and the actual effectiveness of study schedules and point values for the participants assigned to the induction test in Experiment 1. Based on their post-test responses, (A) shows the judged and actual effectiveness of study schedules (with blocked schedules symbolized as B and interleaved schedules symbolized as I), while (B) shows the judged and actual effectiveness of point values.

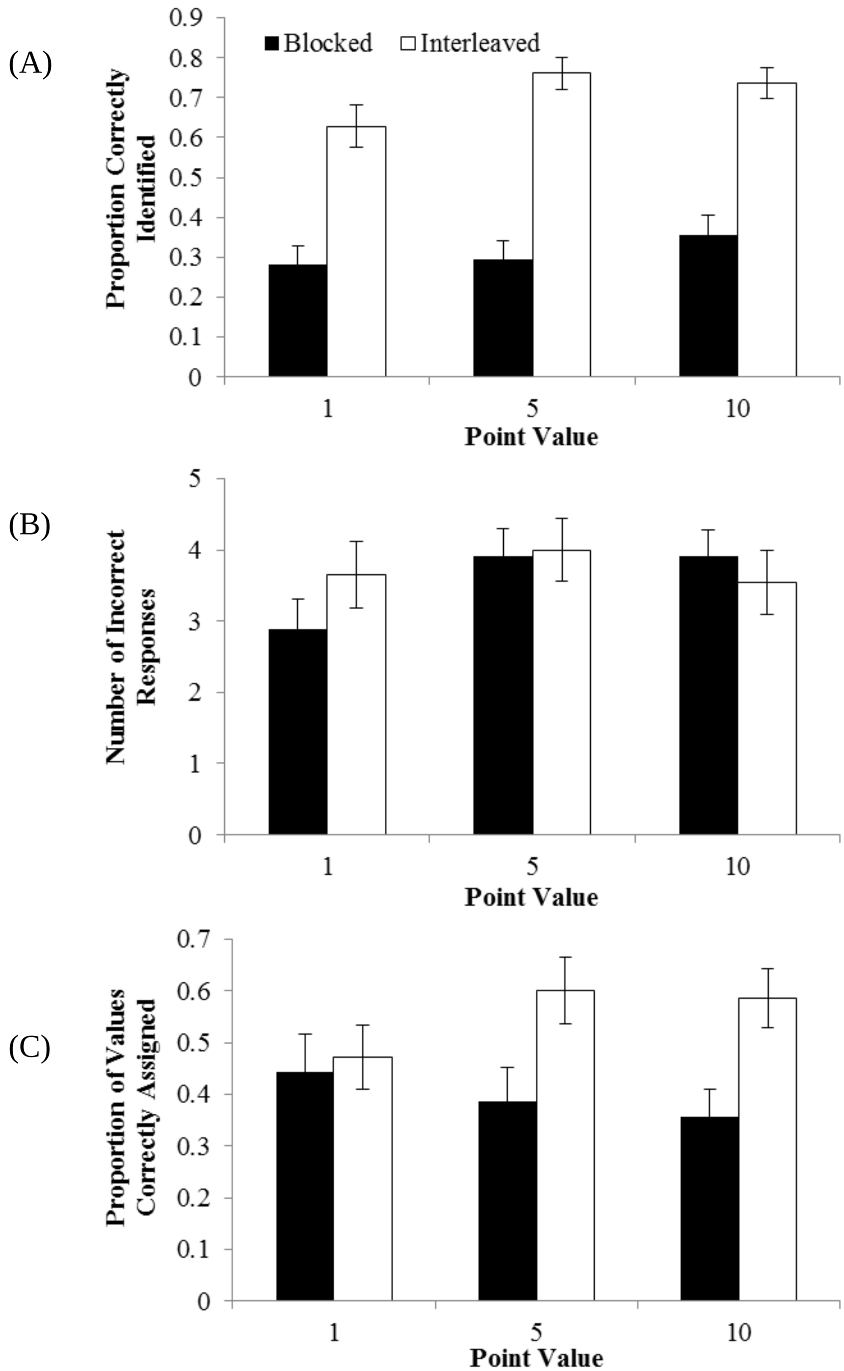


Figure 7. The average performance levels of participants that were assigned to take the memory test in Experiment 1 based on study schedule and point value conditions. (A) shows participants' performance on the memory test, (B) shows the number of incorrect responses participants made, and (C) shows the proportion of values correctly assigned by participants based on the artists, post-test. All error bars represent standard error values.

(A)

(B)

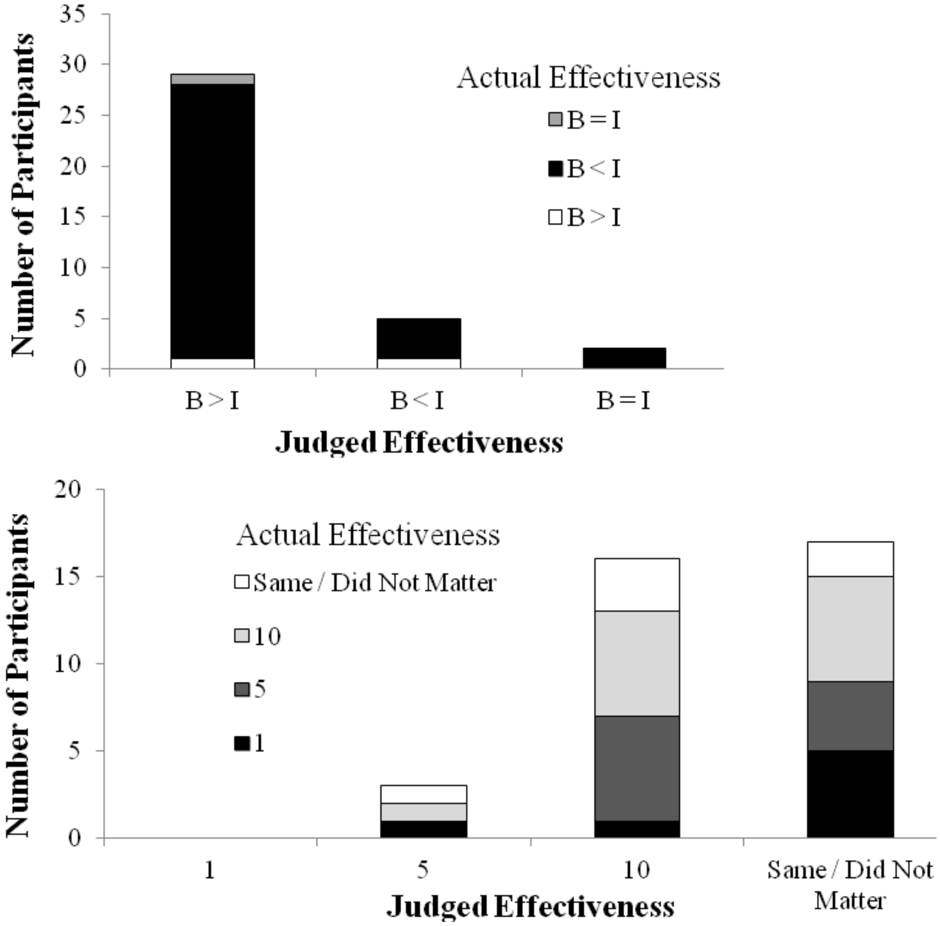


Figure 8. The reported judged effectiveness of participants and the actual effectiveness of study schedules and point values for the participants assigned to the memory test in Experiment 1. Based on their post-test responses, (A) shows the judged and actual effectiveness of study schedules, while (B) shows the judged and actual effectiveness of point values.

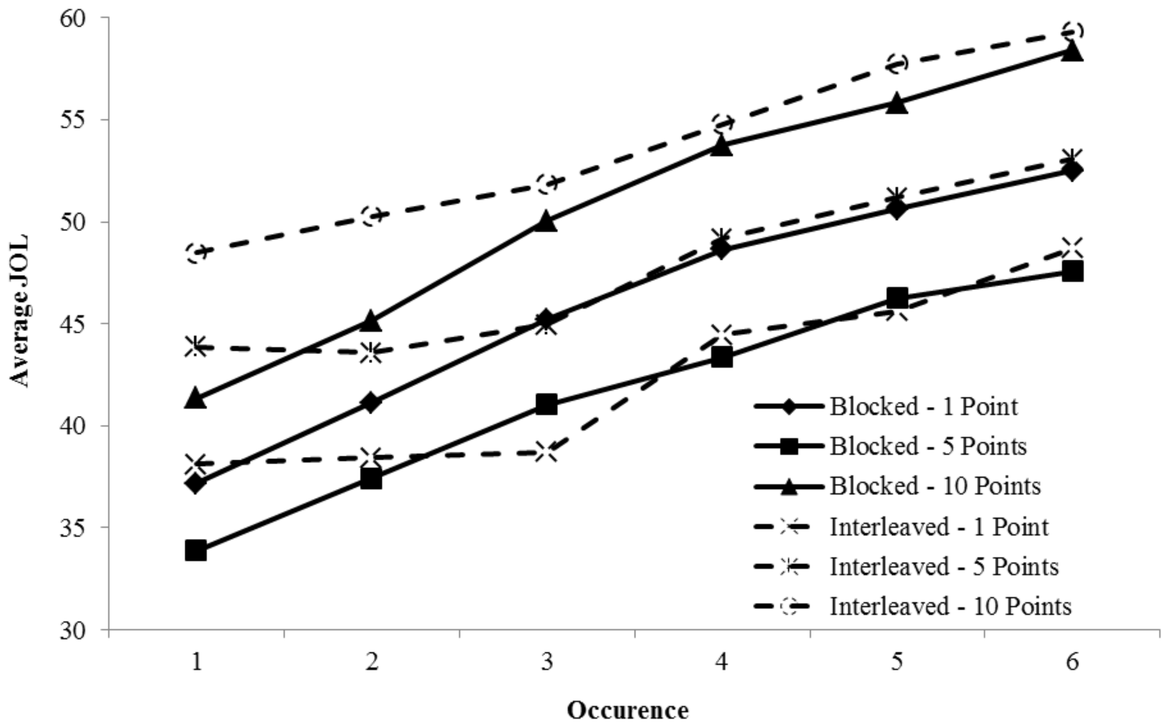


Figure 9. The average judgment of learning value given by participants in Experiment 2 during the learning phase, based on study schedule and point value conditions. The results are plotted based on the number of paintings (or occurrence) shown for that given schedule-point value condition.

(A)

(B)

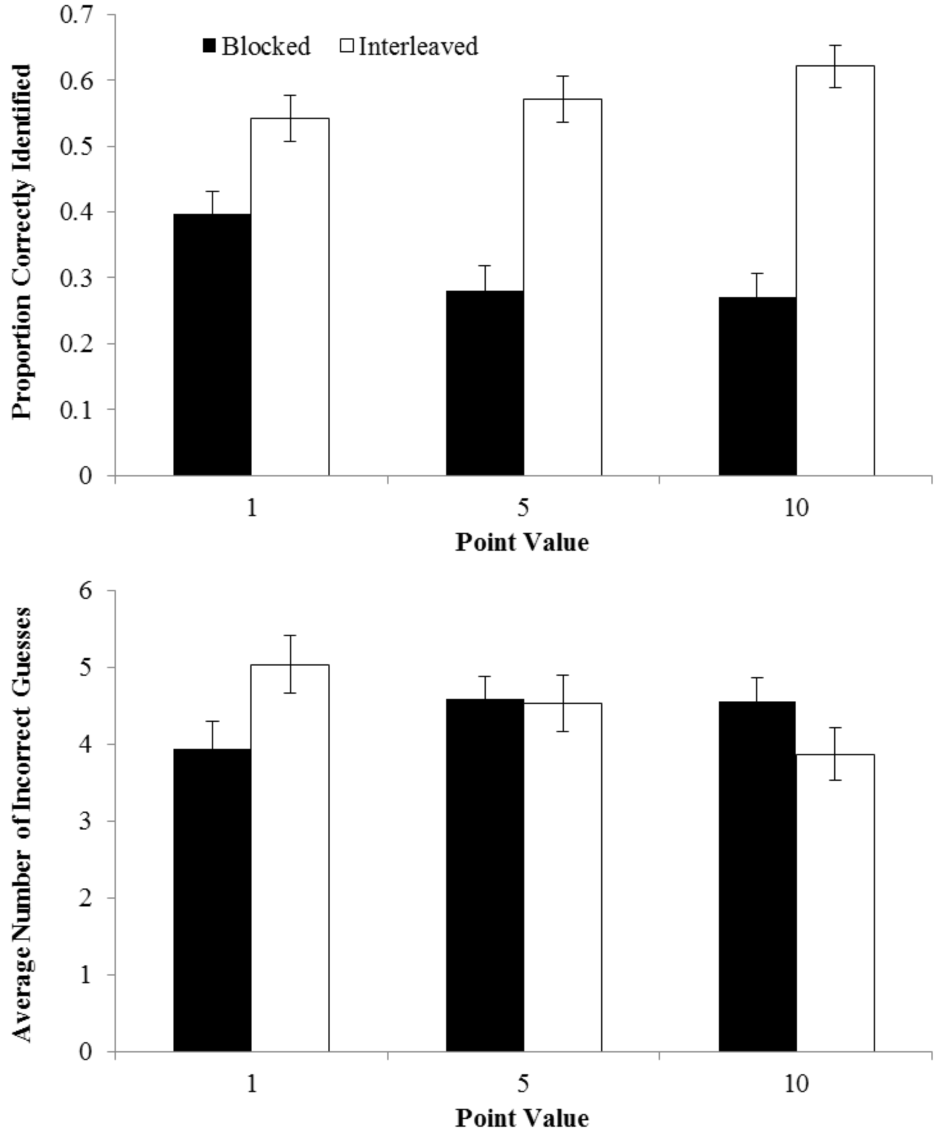


Figure 10. The average performance levels of participants during the induction test in Experiment 2 based on study schedule and point value conditions. (A) shows participants' performance on the induction test, and (B) shows the number of incorrect responses participants made. All error bars represent standard error values.

(A)

(B)

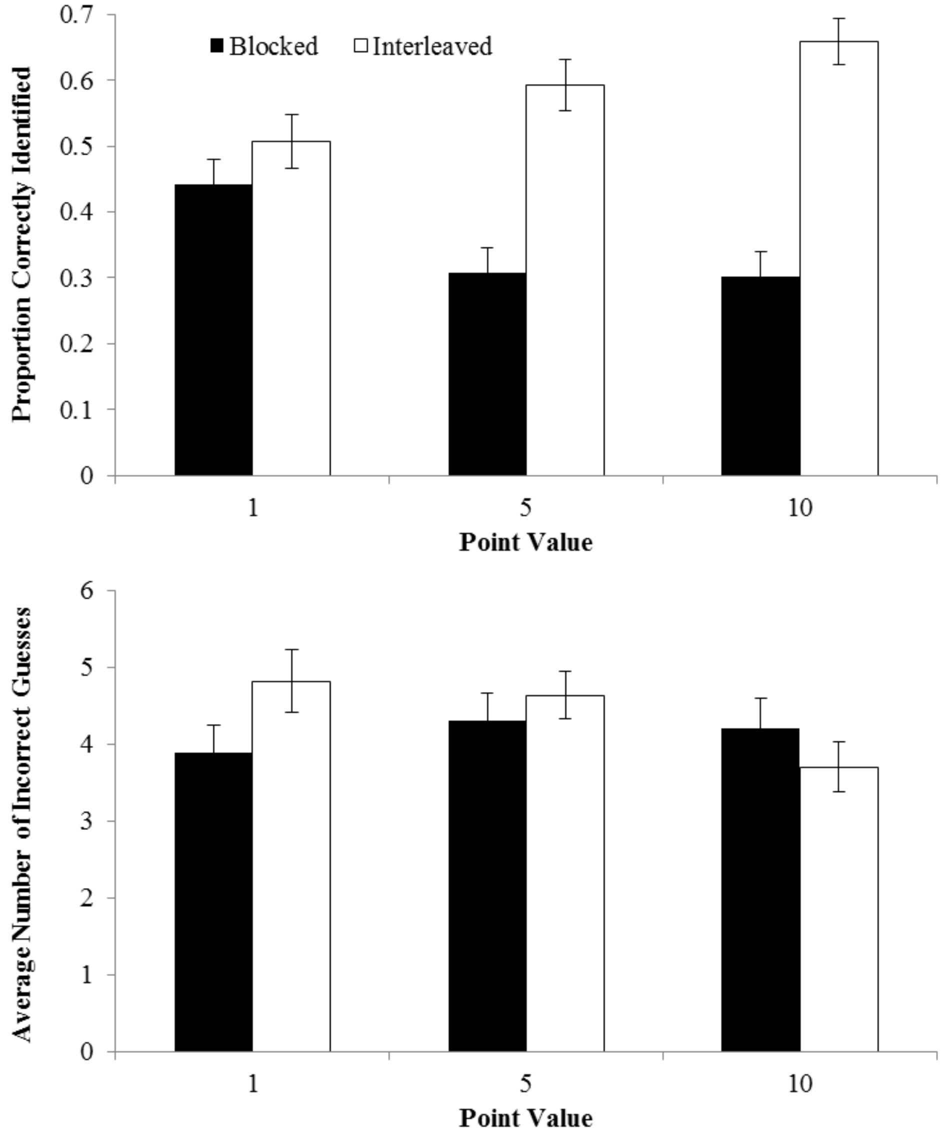


Figure 11. The average performance levels of participants during the memory test in Experiment 2 based on study schedule and point value conditions. (A) shows participants' performance on the memory test, and (B) shows the number of incorrect responses participants made. All error bars represent standard error values.

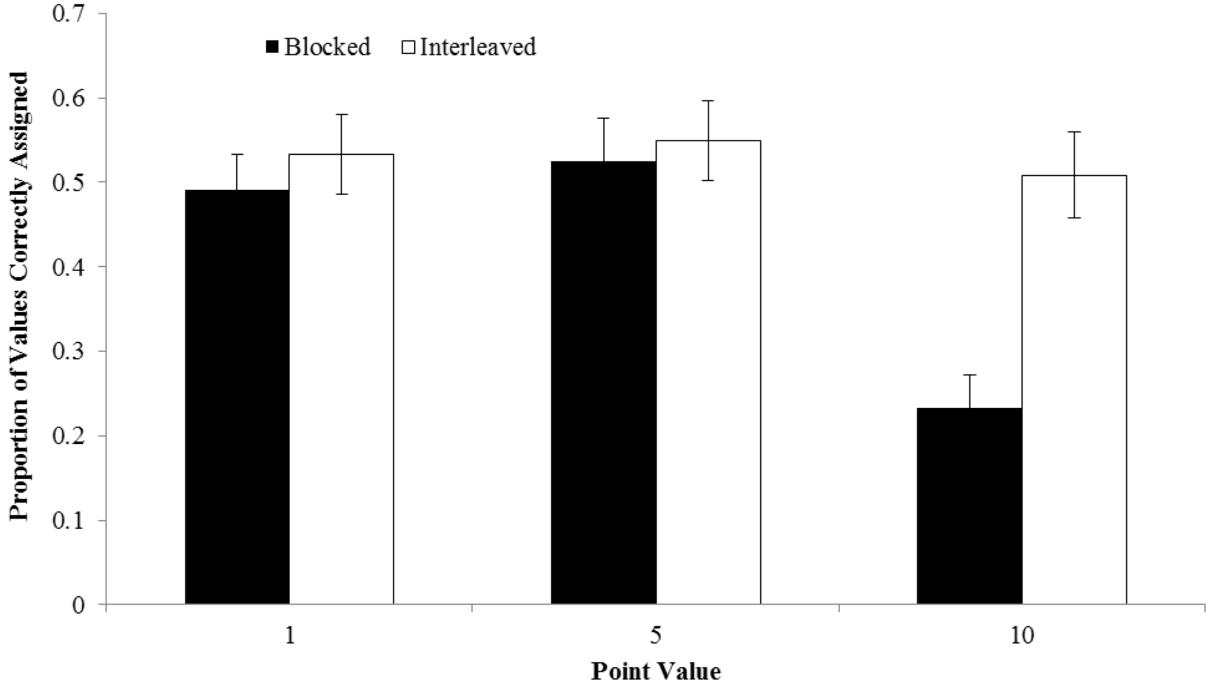


Figure 12. The average proportion of values correctly assigned by participants in Experiment 2, based on the artists' assigned value, post-test. All error bars represent standard error values.

(A)

(B)

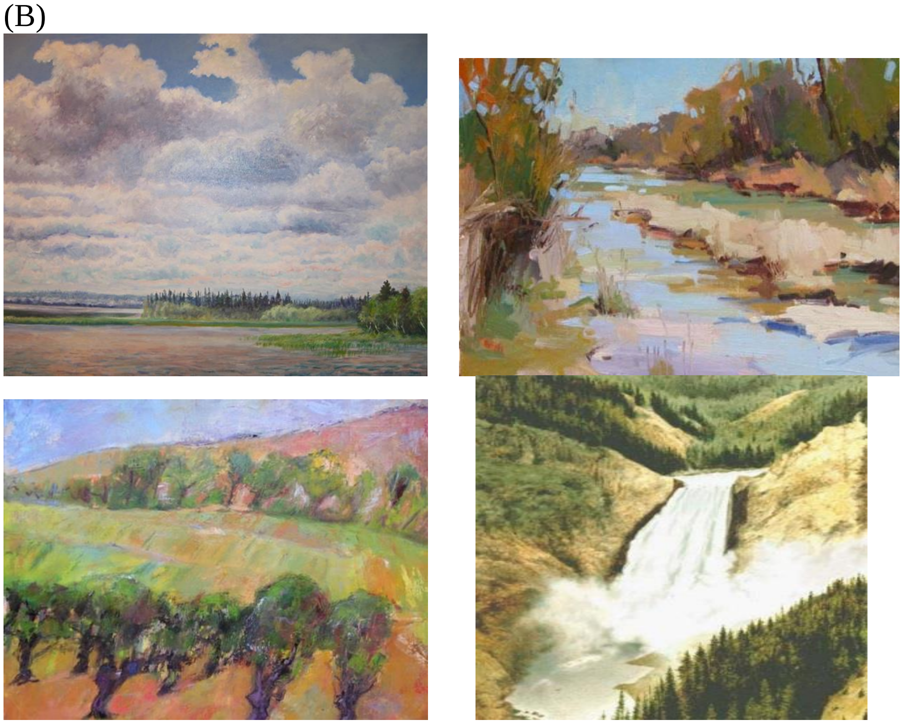


Figure 13. Sample paintings from one of the studied artists (A) and samples of the associated distractor paintings (B) used in the test phase of Experiment 3.

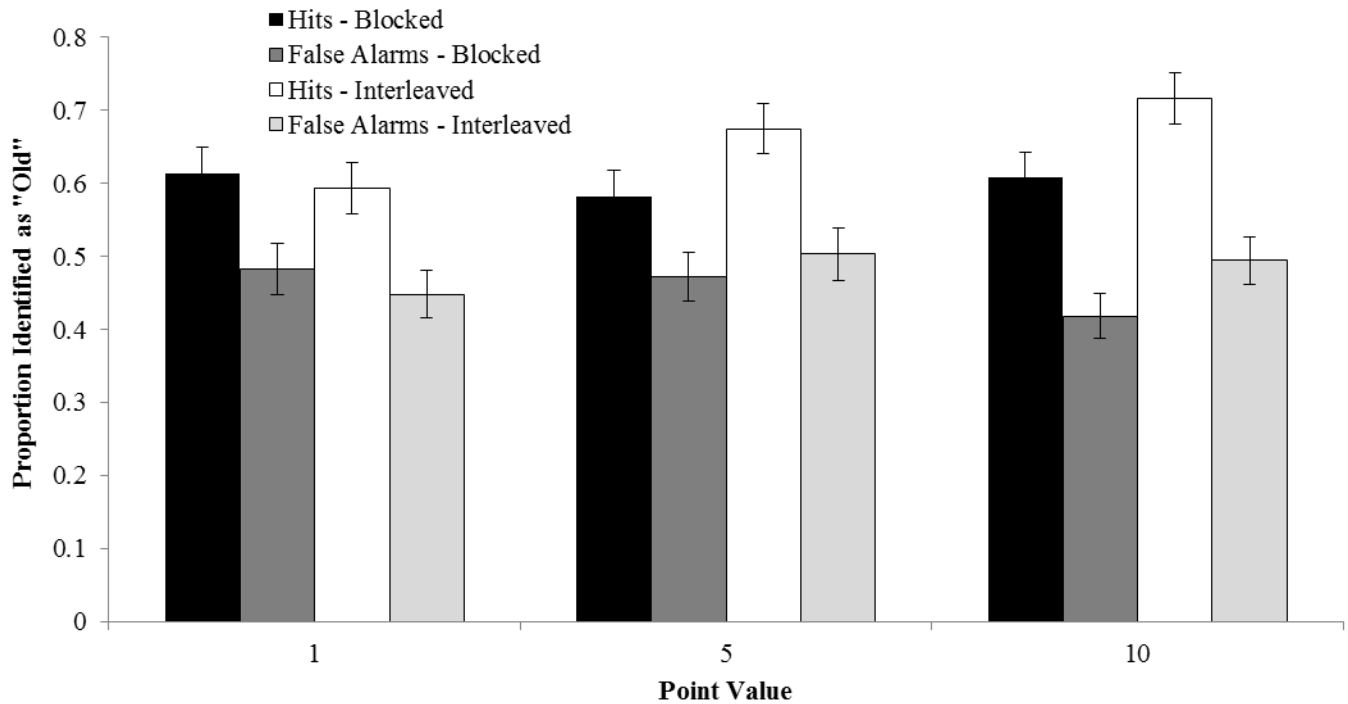


Figure 14. The average proportion of paintings identified as "old" by participants on the recognition test in Experiment 3, based on study schedule and point value conditions. All error bars represent standard error values.

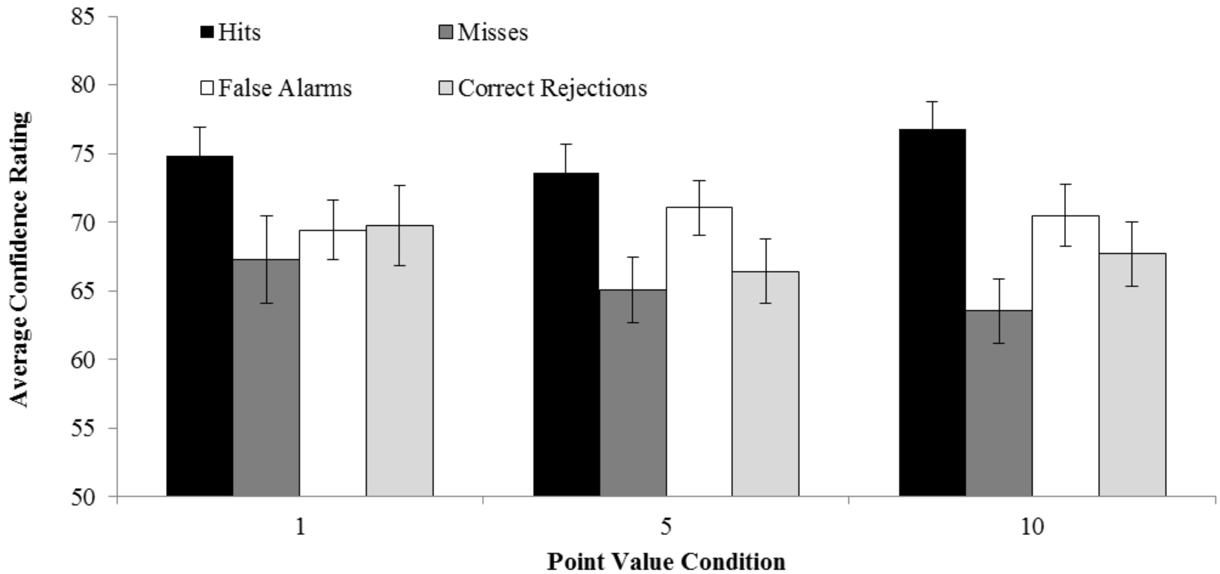


Figure 15. The average confidence ratings given by participants during the recognition test of Experiment 3, based on point value conditions (collapsed across study schedule conditions). All error bars represent standard error values.

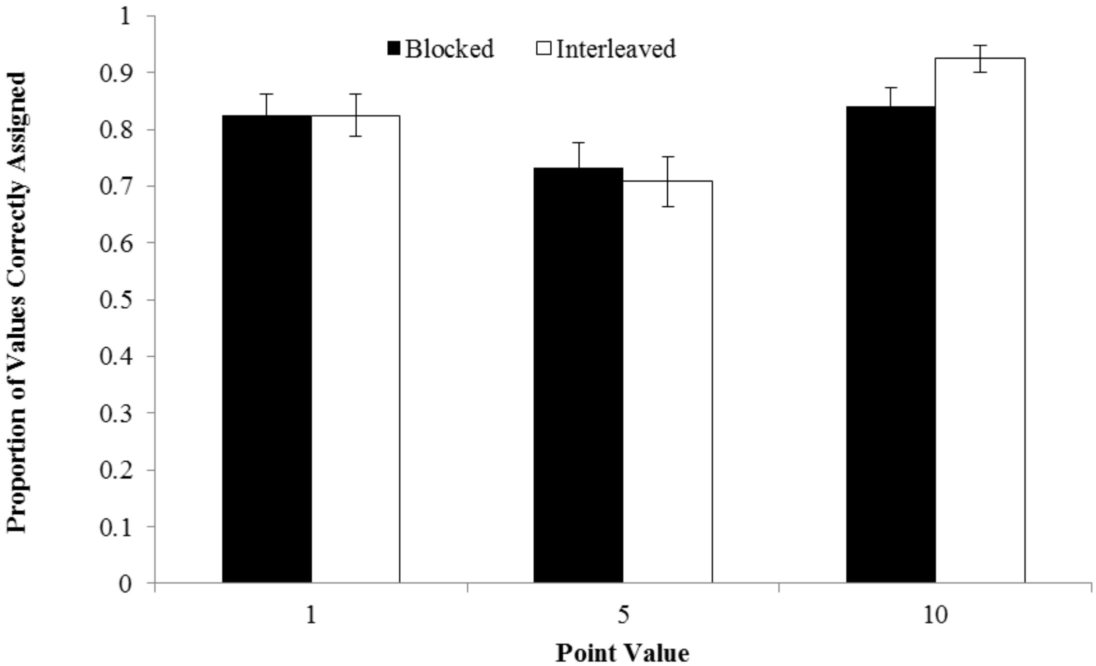


Figure 16. The average proportion of values correctly assigned by participants in Experiment 3, based on the artists' assigned value, post-test. All error bars represent standard error values.

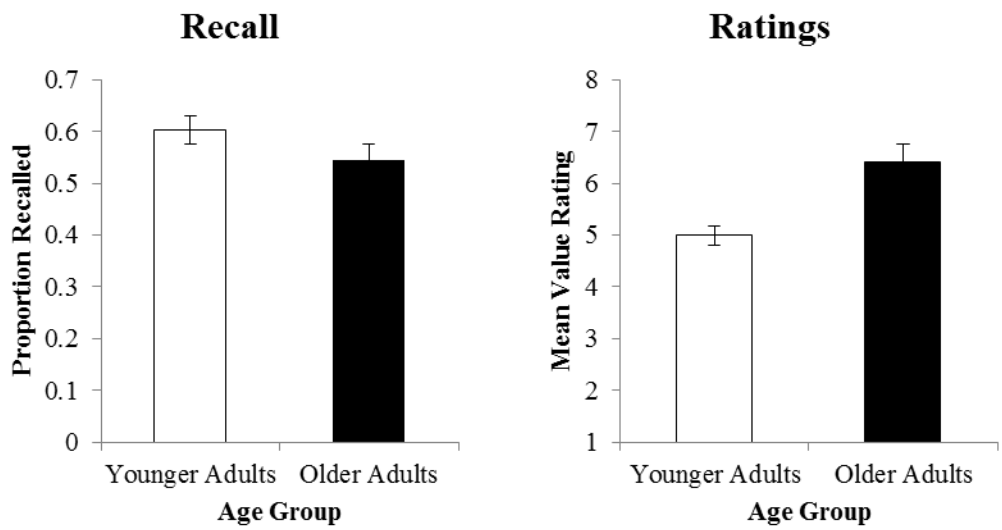
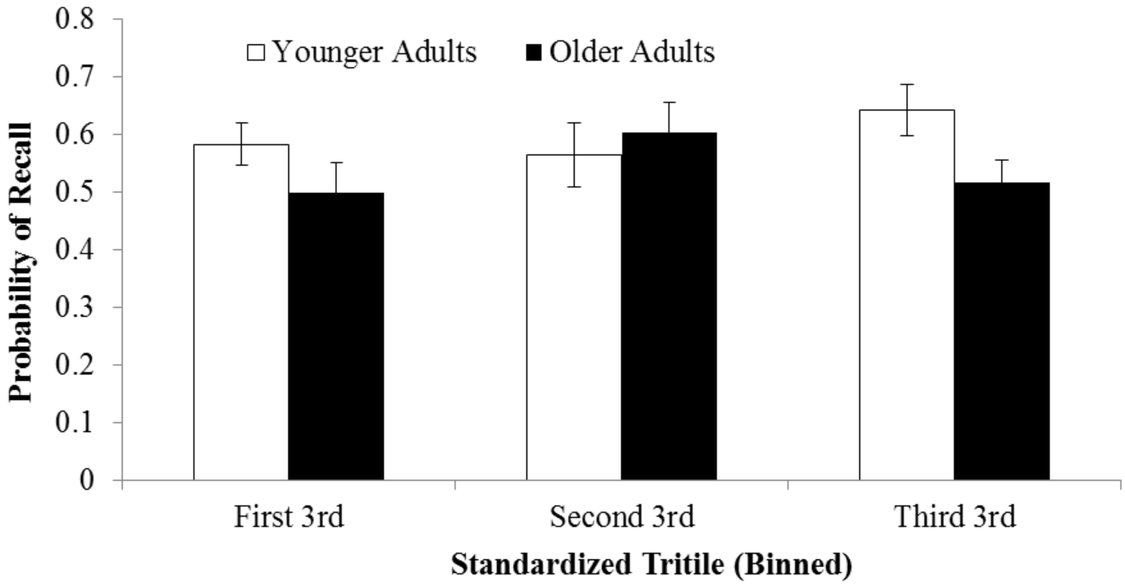


Figure 17. The average recall and average ratings for side effects for older and younger adults in Experiment 4. All error bars represent standard error values.

(A)



(B)

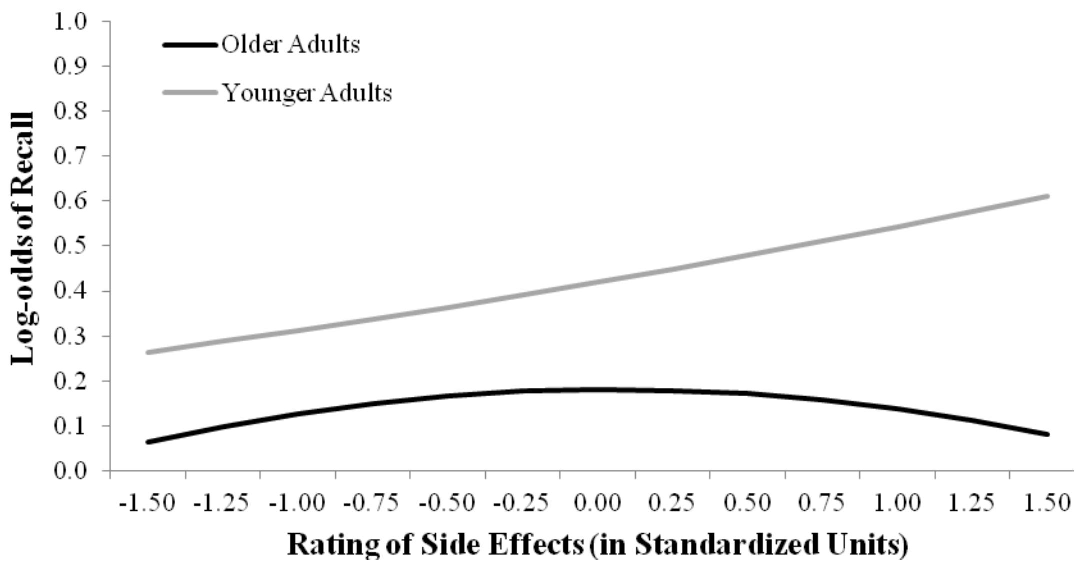


Figure 18. The memory performance of Experiment 4, plotted as a function of younger and older adults' ratings of the medication side effects. The top panel (A) shows the probability of recall based on participants' standardized ratings (within individuals) binned into one of three categories. The first tritile represents any side effect a participant gave a rating below a z-score of -0.43 (i.e., the lowest 33% in a standardized distribution), while the third tritile represents any side effect given a z-score above 0.43 (the highest 33%). The second tritile represents any side effect rated between those two z-scores. The lower panel (B) displays the trend analysis of Experiment 4. The log-odds of recall are plotted as a function of standardized ratings for younger and older adults.

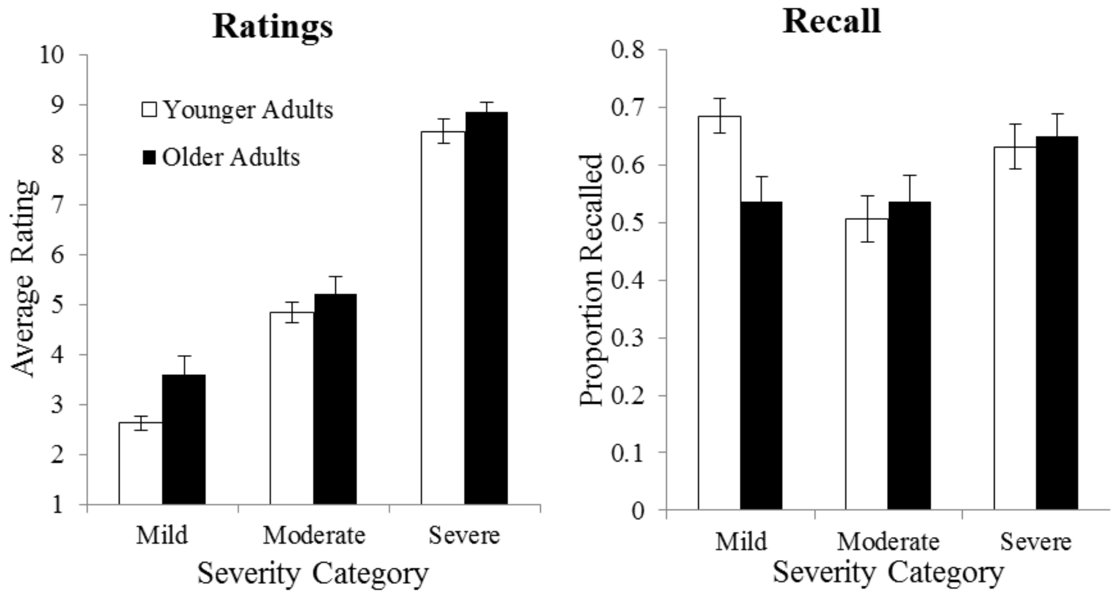


Figure 19. The mean ratings and recall of Experiment 5 based on their assigned severity categories and age group of the participants. All error bars represent standard error values.

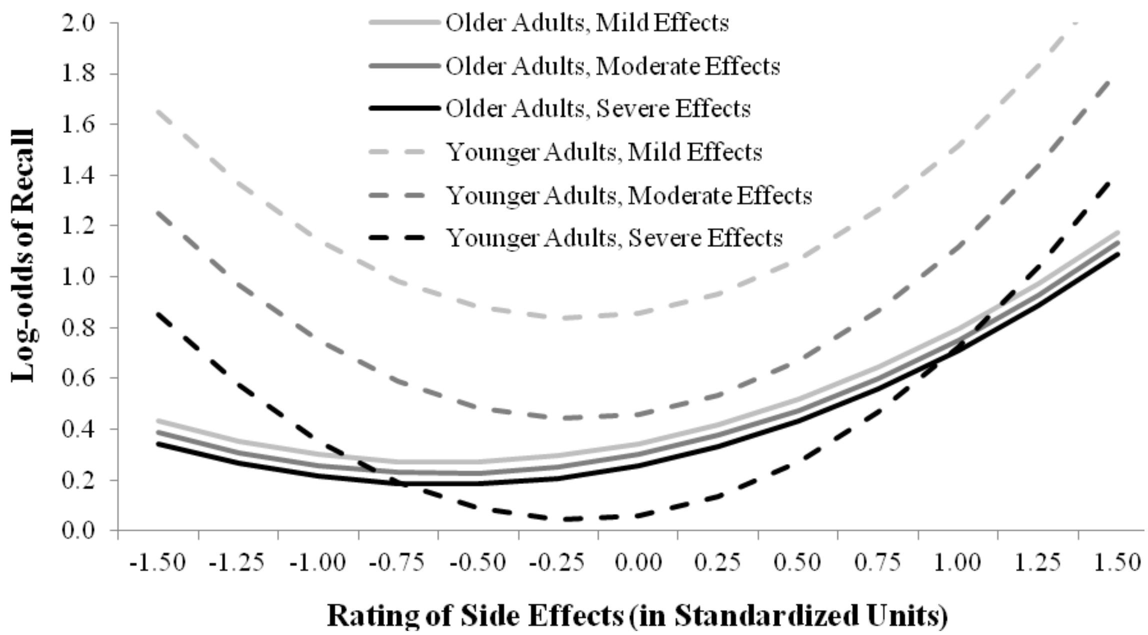


Figure 20. The trend analysis of Experiment 5. The log-odds of recall are plotted as a function of standardized ratings for younger and older adults.

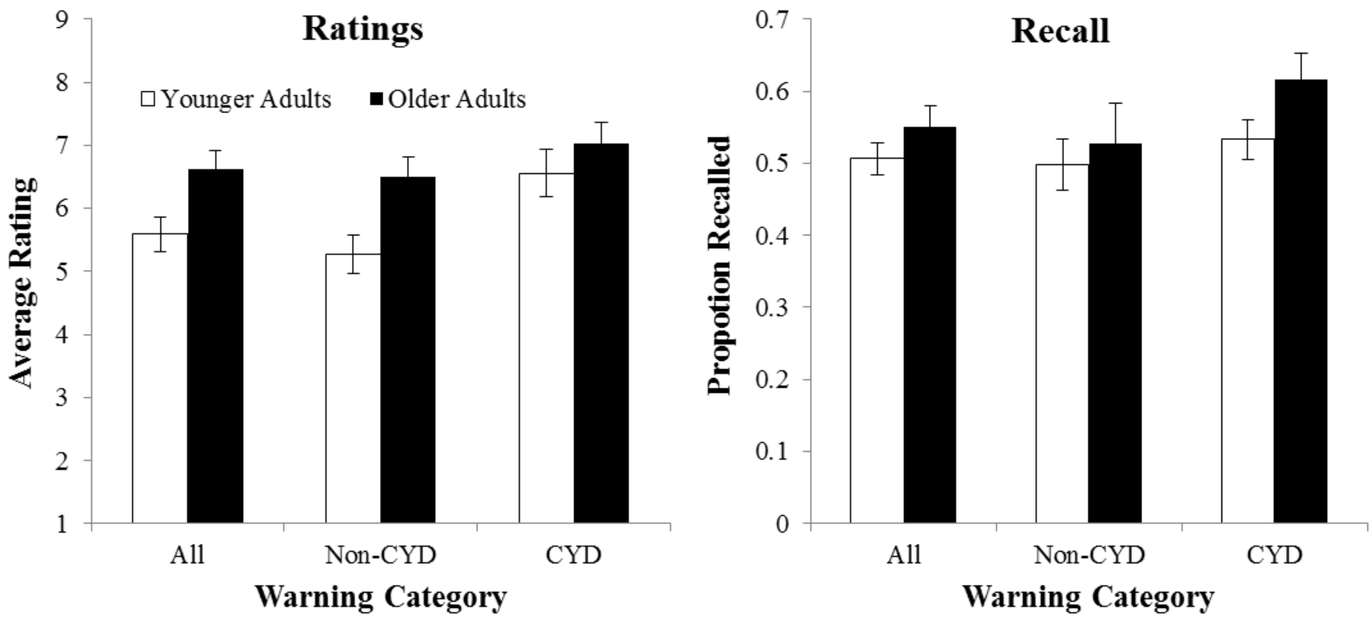


Figure 21. The mean ratings and recall of side effects in Experiment 6 based on their assigned warning categories and the age group of the participants. Since there were three times as many non-CYD side effects as there were CYD side effects, separate bars are shown to depict overall performance (shown on the left side of each graph). All error bars represent standard error values.

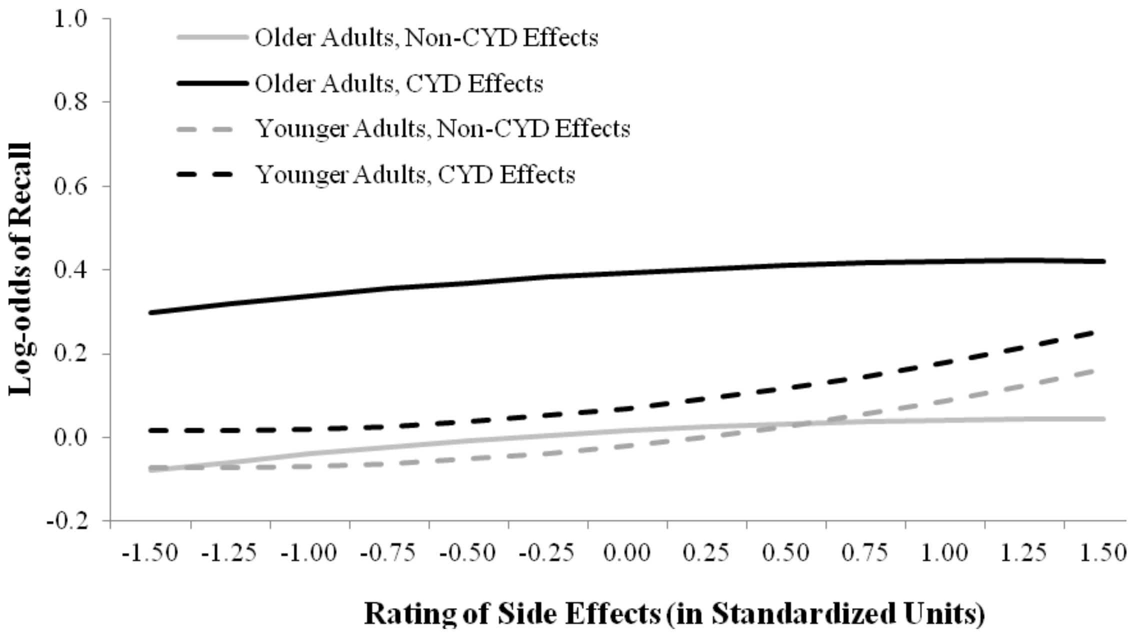


Figure 22. The trend analysis of Experiment 6. The log-odds of recall are plotted as a function of standardized ratings for younger and older adults, and the warning condition of the side effects.

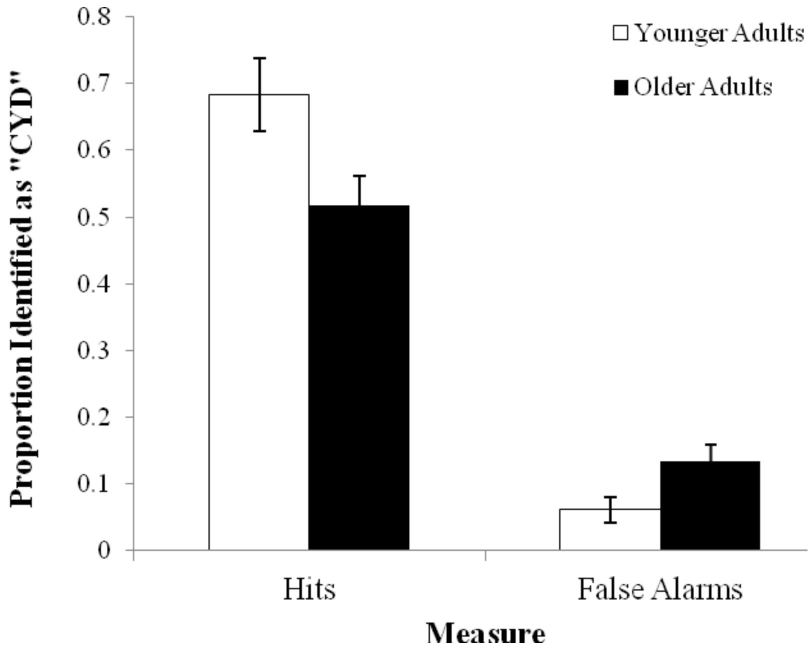


Figure 23. The mean proportion of items identified as “CYD” in the recognition task that followed the recall test in Experiment 6. All error bars represent standard error values.

References

Adams, C. (1991). Qualitative age differences in memory for text: A lifespan developmental perspective. *Psychology and Aging, 6*, 323–336.

Adams, C., Smith, M. C., Nyquist, L., & Perlmutter, M. (1997). Adult age-group differences in recall for the literal and interpretive meanings of narrative test. *Journal of Gerontology: Psychological Science, 57B*, P28–P40.

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. E. (2006). Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron*, *50*, 507–517.

Aizpurua, A., & Koutstaal, W. (2010). Aging and flexible remembering: Contributions of conceptual span, fluid intelligence, and front functioning. *Psychology and Aging, 25*, 193-207.

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*, 126.131.

Ariel, R., Dunlosky, J., and Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138*, 432-447.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30*(5), 666-677.

Baltes, P. B., & Baltes, M. M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. In P. B. Baltes and M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences* (pp. 1–34). New York, NY: Cambridge University Press.

Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28,* 610-632.

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 309-338). Hillsdale, NJ: Lawrence Erlbaum.

Benjamin, A. S., Bjork, R. A, & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55-68.

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. *Memory & Cognition, 41*, 392-402.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe and A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp.185-205). Cambridge, MA: MIT Press.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society* (pp. 56-64). New York: Worth Publishers.

Blanco, C., Alderson, D., Ogburn, E., Grant, B. F., Nunes, E. V., Hatzenbuehler, M. L., & Hasin, D. S. (2007). Changes in the prevalence of non-medical prescription drug use and drug use disorders in the united states: 1991-1992 and 2001-2002. *Drug and Alcohol Dependence*, *90*(2-3), 252-260.

Bui, D. C., Friedman, M. C., McDonough, I. M., & Castel, A. D. (in press). False memory and importance: Can we prioritize encoding without consequence? *Memory & Cognition.*

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. . *Psychology and Aging, 20*, 718–721.

Castel, A. D. (2007). Aging and memory for numerical information: The role of specificity and expertise in associative memory. *Journal of Gerontology: Psychological Sciences, 62*, 194-196.

Castel, A. D. (2008a). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 225-270). London: Academic Press.

Castel, A. D. (2008b). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition, 36*, 429-437.

Castel, A. D., Balota, D. A., & McCabe, D. P. (2009). Memory efficiency and the strategic control of attention at encoding: Impairments of value-directed remembering in Alzheimer's Disease. *Neuropsychology, 23,* 297-306.

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.

Castel, A. D., Farb, N., & 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.

Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S., & Link, I. (2013). Selecting valuable information to remember: Age-related differences and similarities in self-regulated learning. *Psychology and Aging*, *28,* 232-242.

Castel, A. D., Friedman, M. C., Murayama, K., McGillivray, S., & Flores, C. (in preparation). Memory for money associated with faces: Age-related similarities and differences in remembering gains and losses.

Castel, A. D., Humphreys, K. L., Lee, S. S., Galván, A., Balota, D. A., & McCabe, D. P. (2011). The development of memory efficiency and value-directed remembering across the lifespan: A cross-sectional study of memory and selectivity. *Developmental Psychology, 47,* 1553–1564.

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,* 15-24.

Castel, A. D., McCabe, D. P., & Roediger, H. L., III. (2007). Illusions of competency and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*, 107-111.

Castel, A. D., McGillivray, S., & Friedman, M. C. (2012). Metamemory and memory efficiency in older adults: Learning about the benefits of priority processing and value-directed remembering. In M. Naveh-Benjamin & N. Ohta (Eds.), *Memory and aging: Current issues and future directions* (pp. 245-270). New York: Psychology Press.

Castel, A. D., McGillivray, S., & Worden, K. M. (in press). Back to the future: Past and future era-based schematic support and associative memory for prices in younger and older adults. *Psychology and Aging*.

Castel, A. D., Rhodes, M. G., McCabe, D. P., Soderstrom, N. C., & Loaiza, V. M. (2012). The fate of being forgotten: Information that is initially forgotten is judged as less important. *The Quarterly Journal of Experimental Psychology, 65,* 2281-2287.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24*, 403-416.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging, 12*, 50-71.

DeSantis, A. D., Webb, E. M., & Noar, S. M. (2008). Illicit use of prescription ADHD medications on a college campus: A multimethodological approach. *Journal of American College Health, 57*, 315-323.

Dodson, C. S., Johnson, M. K., & Schooler, J. W. (1997). The verbal overshadowing effect: Why descriptions impair face recognition, *Memory & Cognition, 25*, 129-139.

Dunlosky, J., & Connor, L. T. (1997). Age-related differences in the allocation of study time account for age-related differences in memory performance. *Memory & Cognition, 25,* 691-700.

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

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34,* 906-911.

Friedman, M. C., & Castel, A. D. (in press). Memory, priority encoding, and overcoming high-value proactive interference in younger and older adults. *Aging, Neuropsychology, and Cognition.*

Friedman, M. C., & Castel, A. D. (2011). Are we aware of our ability to forget? Metacognitive predictions of directed forgetting. *Memory & Cognition, 39,* 1448-1456.

Gerlach, K. D. (2008). *Value-directed remembering*. Unpublished manuscript.

Goldstone, R. L., Landy, D., & Brunel, L. C. (2011). Improving perception to make distant connections closer. *Frontiers in Psychology, 2,* 1-10.

Hacker, D. J., Bol, L., Horgan, D. D., & Rakow, E. A. (2000). Test prediction and performance in a classroom context. *Journal of Educational Psychology, 92,* 160-170.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 801-812.

Halamish, V., McGillivray, S., & Castel, A. D. (2011). Monitoring one's own forgetting in younger and older adults. *Psychology and Aging*, *26*, 631-635.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology, 56*¸ 208-216.

Hart, J. T. (1966). Methodological note on feeling-of-knowing experience. *Journal of Educational Psychology*, *57*¸ 347-349.

Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior, 6*¸ 685-691.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193-225). New York, NY: Academic Press.

Haynes, R. B., Ackloo, E., Sahota, N., McDonald, H. P., & Yao, X. (2007). Interventions for enhancing medication adherence. *Cochane Database of Systematic Reviews*, *4*, 1-130.

Henry, J. D., MacLeod, M. S., Phillips, L. H., & Crawford, J. R. (2004). A meta-analytic review of prospective memory and aging. *Psychology and Aging*, *19*, 27-39.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood: Spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science*, *20*, 167-173.

Hertzog, C., & Hultsch, D. F. (2000). Metacognition in adulthood and old age. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 417–466). Mahwah, NJ: Lawrence Erlbaum Associates.

Hertzog, C., Kidder, D. P., Powell-Moman, A., & Dunlosky, J. (2002). Aging and monitoring associative learning: Is monitoring accuracy spared or impaired? *Psychology and Aging, 17*, 209–225.

Hines, L. E., & Murphy, J. E. (2011). Potentially harmful drug–drug interactions in the elderly: A review. *American Journal of Geriatric Pharmacotherapy, 9,* 364-377.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

Jacoby, L. L., Bishara, A. J., Hessels, S., & Toth, J. P. (2005). Aging, subjective experience, and cognitive control: Dramatic false remembering by older adults. *Journal of Experimental Psychology: General, 134*, 131-148.

Jacoby, L. L., Wahlheim, C. N., Rhodes, M. G., Daniels, K. A., & Rogers, C. S. (2010). Learning to diminish the effects of proactive interference: Reducing false memory for young and older adults. *Memory & Cognition, 38*, 820-829.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.

Jones, T. C., & Jacoby, L. L. (2005). Conjunction errors in recognition memory: Modality-free errors for older adults but not for young adults. *Acta Psychologica, 120*, 55-73.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26*, 97-103.

Kassam, K. S., Gilbert, D. T., Swencionis, J. K., & Wilson, T. D. (2009). Misconceptions of memory: The Scooter Libby Effect. *Psychological Science, 20*, 551-552.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349-370.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *31*, 187-194.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643-656.

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*, 147-162.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585-592.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65*, 85-97.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in younger and older adults. *Psychology and Aging*, *25*, 498-503.

Koutstaal, W. (2006). Flexible remembering. *Psychonomic Bulletin & Review*, *13*, 84–91.

Koutstaal, W., & Cavendish, M. (2006). Using what we know: Consequences of intentionally retrieving gist versus item-specific information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 778-791.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*(4), 555-583.

Liu, L. L., & Park, D. C. (2004). Aging and medical adherence: The use of automatic processes to achieve effortful things. *Psychology and Aging, 19*, 318-325.

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent developments in a "new view". (pp. 145-162) American Psychological Association, Washington, DC.

Madan, C. R., & Spetch, M. L. (2012). Is the enhancement of memory due to reward driven by value or salience? *Acta Psychologica, 139*, 343-349.

Magee, C. L. (2007). The use of herbal and other dietary supplements and the potential for drug interactions in palliative care. *Palliative Medicine, 21*, 547-548.

McDaniel, M. A., & Einstein, G. O. (2007). Prospective memory components most at risk for older adults and implications for medication adherence. (pp. 49-75) American Psychological Association, Washington, DC.

McGillivray, S., & Castel, A. D. (2011). Betting on memory leads to metacognitive improvement in younger and older adults. *Psychology and Aging*, *26*, 137-142.

McGillivray, S., Friedman, M. C., & Castel, A. D. (2012). Impact of aging on thinking. In K. Holyoak & R. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 650-672). New York: Oxford University Press.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131*, 349-363.

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*, 463-477.

Miele, D. B., Finn, B. & Molden, D. C. (2011). Does easily learned mean easily remembered? It depends on your beliefs about intelligence. *Psychological Science, 22*, 320-324.

Miller, L. M. S. (2003). The effects of age and domain knowledge on text processing. *Journals of Gerontology: Psychological Sciences, 58*, 217-222.

Moore, A. A., Whiteman, E. J., & Ward, K. T. (2007). Risks of combined alcohol/medication use in older adults. *American Journal of Geriatric Pharmacotherapy, 5*, 64-74.

Morrow, D. G., Leirer V. O., Altieri, P. A., & Fitzsimmons, C. (1994). When expertise reduces age differences in performance. *Psychology and Aging, 9*, 134-148.

Morrow, D. G., Leirer, V. O., & Sheikh, J. (1988). Adherence and medication instructions: Review and recommendations.*Journal of the American Geriatrics Society, 36*, 1147-1160.

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*, 1170-1187.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95*, 109-133.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-`73). New York: Academic Press.

Nilsson, L. G. (1987). Motivated memory: Dissociation between performance data and subjective reports. *Psychological Research, 49*, 183-188.

Old, S., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23*, 104–118.

Qato, D. M., Alexander, G. C., Conti, R. M., Johnson, M., Schumm, P., & Lindau, S. T. (2008). Use of prescription and over-the-counter medications and dietary supplements among older adults in the united states. *JAMA: Journal of the American Medical Association, 300*, 2867-2878.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*, 615–625.

Rhodes, M. G., & Castel, A. D. (2009). Metacognitive illusions for auditory information: Effects on monitoring and control. *Psychonomic Bulletin & Review, 16*, 550-554.

Rhodes, M. G., Castel, A. D., & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and Aging, 23*, 239-249.

Rundus, D. (1980). Maintenance rehearsal and long-term recency. *Memory & Cognition, 8*, 226-230.

Salthouse, T. A. (1996). The processing-speed theory of adult age difference in cognition. *Psychological Review, 103*, 403–428.

Shohamy, D., & Adcock, R. A. (2010). Dopamine and adaptive memory. *Trends in Cognitive Science, 14*, 464.472.

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based regulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1236-1242.

Sollman, M. J., Ranseen, J. D., & Berry, D. T. R. (2010). Detection of feigned ADHD in college students. *Psychological Assessment, 22*, 325-335.

Storm, B. C., Friedman, M. C., Murayama, K., & Bjork, R. A. (in press). On the transfer of prior tests or study events to subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Sungkhasettee, V. W., Friedman, M. C., & Castel, A. D. (2011). Memory and metamemory for inverted words: Illusions of competency and desirable difficulties. *Psychonomic Bulletin & Review, 18*, 973-978.

United States v. Libby, 461 F. Supp. 2d. 12 (D.D.C. 2006)

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction.*Cognition, 109*, 163-167.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & Cognition, 39*, 750-763.

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

Webb, J., Davis, T. C., Bernadella, P., Clayman, M. L., Parker, R. M., Adler, D., & Wolf, M. S. (2008). Patient-centered approach for improving prescription drug warning labels. *Patient Education and Counseling, 72*, 443-449.

Zacks, R. T., & Hasher, L. (2006). Aging and long-term memory: Deficits are not inevitable. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 162-177). New York, NY: Oxford University Press.

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15*, 41-44.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition, 41*, 16-27.

Zogg, J. B., Woods, S. P., Sauceda, J. A., Wiebe, J. S., & Simoni, J. M. (2012). The role of prospective memory in medication adherence: A review of an emerging literature. *Journal of Behavioral Medicine*, *35,* 47-62.