Effects of Error Generation on Episodic Memory among Healthy Younger and Older Adults

by

Andrée-Ann Cyr

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Psychology
University of Toronto

© Copyright 2015 by Andrée-Ann Cyr
Abstract

Effects of Error Generation on Episodic Memory among Healthy Younger and Older Adults

Andrée-Ann Cyr
Doctor of Philosophy
Graduate Department of Psychology
University of Toronto
2015

Making mistakes during everyday learning is the norm rather than the exception, but its consequences for episodic memory are still unclear. A pedagogically motivated literature among younger adults shows that generating errors at study leads to improvements in memory for correct information (i.e. the error generation effect). By contrast, a clinically motivated literature among older adults suggests that error generation hinders episodic memory and should be avoided. The goal of this dissertation is to bridge these two literatures and uncover the conditions under which errors help or harm memory among healthy younger and older adults. Study 1 showed that relative to avoiding errors, generating errors at study led to increases in target memory when learning from conceptual cues, but decreases in target memory when learning from nonconceptual lexical cues. Moreover, this dissociation was found among both younger and older adults. Study 2 replicated this conceptual error generation benefit and showed that it was not modulated by cue constraint in both age groups (Study 2A), but that it was enhanced among younger adults when errors and targets were strongly relative to weakly related (Study 2B). Finally, Study 3 found that for both age groups, general knowledge errors were more likely to be corrected when they were produced with high relative to low confidence: This ‘hypercorrection effect’ was stronger in younger than older adults when a free recall learning paradigm was used (Study 3A), but this age difference was eliminated when a multiple-choice paradigm was employed (Study 3B). Overall, these results suggest that the conditions yielding positive or negative effects of error generation on episodic memory are similar among younger and older adults: Lexical (i.e., nonconceptual) errors hinder memory, but conceptual errors enhance memory as a function of the amount of semantic processing they afford. These findings are interpreted within a framework of error processing, which posits the role
of cognitive control in reducing error interference and enhancing activation of target information.
Acknowledgements

It takes a village.

First and foremost, sincerest thanks to Nicole Anderson for taking on the (thankless) role of supervisor with such unwavering dedication (*just keep swimming!*). Thank you for playing to my forest strengths by giving me considerable creative license in all my research projects, and my tree weaknesses by teaching me the importance of well-made graphs and meticulous research design. Importantly, thank you for being a great human being, especially at times in my journey when compassion rather than reason strengthened my resolve. Thank you also to my esteemed committee members Gus Craik and Cheryl Grady for their valued input and encouragement over the years. I feel very fortunate to have received such expert guidance.

To the army of research assistants and volunteers that have made my studies possible: Thank you! In no particular order, I am grateful to Mariam Iskander, Andrea Maoine, Eileen Jung, Vinay Kansal, Laura Stefanik, Nisha Kansal, and Janice Babins. Thanks also to the best labmates a student could ask for: Emma Guild, Katherine Krpan, and Brandon Vasquez.

Many thanks to my family for being supportive - albeit at times incredulous- spectators to my feat of endurance. Thanks to Horia Vulpe, Yashar Salek and Matthew Price for walking with me at different points in time, as well as Alexa Tullet and Dan Lee for being kindred spirits in life and research.
Contents

1 Introduction .................................................. 1
   1.1 Defining healthy aging and episodic errors .................. 1
       1.1.1 Age-related memory changes .......................... 1
       1.1.2 Memory errors ...................................... 3
       1.1.3 Diverging views on episodic learning among younger and older adults .... 4
   1.2 Errors are bad: Overview of the older adult literature .......... 6
       1.2.1 The errorless learning benefit ....................... 6
       1.2.2 Cognitive mechanisms of the errorless learning benefit .......... 6
       1.2.3 Summary of the older adult literature .................. 9
   1.3 Errors are good: Overview of the younger adult literature ...... 10
       1.3.1 Retrieval-based learning ............................ 10
       1.3.2 The testing effect .................................. 11
       1.3.3 The error generation effect ......................... 12
       1.3.4 The hypercorrection effect ......................... 14
   1.4 Bridging the older and younger adult literature ............... 15
       1.4.1 Retrieval-based learning and aging ................... 15
       1.4.2 Evidence of an error generation effect among older adults ........ 16
   1.5 Thesis overview ......................................... 18

2 Dissociation between type of processing and effects of error generation on memory 21
   2.1 Study 1 .................................................. 24
       2.1.1 Method ............................................. 25
       2.1.2 Results ........................................... 28
       2.1.3 Discussion ......................................... 32
List of Tables

2.1 Mean (SD) scores on neuropsychological measures as a function of age group and condition. .................................................. 26
List of Figures

1.1 Estimates of familiarity (left panel) and recollection (right panel) as a function of age and encoding context in Cyr and Anderson (2012). ........................................ 17

2.1 Mean cued recall performance for targets as a function of age and condition (bars represent ± 1 standard error). ................................................................. 29

2.2 Mean number of prior guesses recalled as a function of age and condition, condition-alized on successful recall of target word (bars represent ± 1 standard error). ....... 30

2.3 Proportion of prior guesses recalled as a function of age and condition, given successful recall of target word (bars represent ± 1 standard error). ................. 31

3.1 Memory for target word endings as a function of age and condition in Study 2A (bars represent ± 1 standard error). ................................................................. 43

3.2 Memory for target words as a function of age and trial-and-error condition in Study 2B (bars represent ± 1 standard error). ......................................................... 48

3.3 Memory for prior errors as a function of age and trial-and-error condition. .................. 49

4.1 Mean accuracy on initial and final free recall as a function of age in Study 3A (bars represent ± 1 standard error). ................................................................. 56

4.2 Mean final free recall accuracy as a function of confidence in initial errors and age in Study 3A (bars represent ± 1 standard error). ........................................ 57

4.3 Mean Somers d coefficient between confidence in the initial error and retest accuracy as a function of age in Study 3A (bars represent ± 1 standard error). ........ 58

4.4 Mean accuracy on initial and final free recall as a function of age in Study 3B (bars represent ± 1 standard error). ................................................................. 61

4.5 Mean final free recall accuracy as a function of confidence in initial errors and age in Study 3B (bars represent ± 1 standard error). ........................................ 62

4.6 Mean Somers d coefficient between confidence in the initial error and retest accuracy as a function of age in Study 3B (bars represent ± 1 standard error). ........ 63
5.1 Schemata representing the action of dual modes of cognitive control for error resolution. .............................................................. 67
5.2 Schemata representing strength of associations as a function of type of processing. ........................................ 69
5.3 Schemata representing strength of associations as a function of semantic proximity. ................................................. 70
5.4 Schemata representing strength of associations as a function of confidence. ......................................................... 71
Chapter 1

Introduction

The ability to acquire and remember new information is at the core of our sense of self and continuity. In practical terms, most years of human life in North America are spent gaining knowledge in educational or workplace settings, and the failure to do so efficiently is acutely felt on both personal and social levels. The aging trend in our demographics aptly captures this felt importance of cognitive integrity: For the aging individual, age-related memory failures represent a primary concern (Hertzog & Dixon, 1994). For society, declining learning ability with age is tied to a dwindling workforce and increased social and healthcare burdens. As such, uncovering the nature of age-related learning deficits should be a prime concern for cognitive psychologists. The objective of my dissertation is to elucidate the effect of making errors on episodic memory among healthy younger and older adults. How we recover from mistakes is central to improving educational practice across the lifespan and affords insight into age-related changes in episodic learning.

1.1 Defining healthy aging and episodic errors

1.1.1 Age-related memory changes

The concept of age calls to mind both pathological processes leading to morbidity (e.g., dementia, cerebral vascular disease), and more ‘wear-and-tear’ processes leading to normal declines in almost all areas of biological functioning. In memory terms, this latter form of normal or healthy aging is typically accompanied by declines in the ability to consciously remember personally experienced events, e.g., episodic memory (Zelinski & Burnight, 1997). Episodic memory is measured using direct tests of memory such as recognition and recall tasks, which are differentially impacted by aging: Recognition performance is relatively spared whereas recall performance is reduced (for a review, see Craik & Jennings, 1992). Recall is more difficult and unsupportive relative to recog-
nition, requiring older adults to self-initiate strategies to organize retrieval (Craik & Byrd, 1982). Older adults are also impaired in the ability to remember details about the context of studied information, whether external (Park & Puglisi, 1995; Naveh-Benjamin & Craik, 1995; McIntyre & Craik, 1987) or internal (Cohen & Faulkner, 1989; Degl’Innocenti & Backman, 1996). Losing contextual details makes older adults more prone to over relying on feelings of familiarity when making decisions about items (e.g., Jacoby, Woloshyn, & Kelley, 1989), leading to false memories (Schacter, Koutsas, & Norman, 1997a). In the dual process memory model (Jacoby, 1991), the process of recollection counteracts these misleading influences of familiarity during recognition as it contains source information. While familiarity is typically spared with aging, recollection is reliably reduced (for a review, see Prull, Dawes, Martin, Rosenberg, & Light, 2006). In sum, episodic memory declines with healthy aging, particularly its aspects that require self-initiation and effortful reinstatement of encoding details.

Like episodic memory, semantic memory also requires conscious and intentional remembering, but for general knowledge about the world. Semantic memory is relatively spared by aging: Verbal knowledge actually increases with age (Park et al., 2002), and there are no age differences in knowledge of concepts (Eustache, Desgranges, Jacques, & Platel, 1998) or production of category exemplars (Mayr & Kliegl, 2000). When age differences do emerge on semantic memory tasks, they are generally due to older adults susceptibility to nonsemantic aspects of responding, like tip-of-the-tongue. This inability to retrieve a word, accompanied by the strong subjective sense that it is known, is experienced more commonly and takes longer to resolve by older relative to younger adults (Burke, Mackay, Worthley, & Wade, 1991).

Remembering that is unconscious and automatic, such as implicit memory, also holds up with healthy aging (Light & La Voie, 1993; 1994). For instance, older adults show priming effects on indirect tests of previously studied information, such as word stem completion, even if they fail to recall the items consciously on a direct test (Light & Singh, 1997). This age-related dissociation of implicit and explicit memory processes can have similar memorial consequences as that of familiarity and recollection in that if unopposed by explicit memory, implicit memory can lead to memory errors (Baddeley & Wilson, 1994). In sum, the cognitive picture of normal aging depicts a decline in conscious remembering of personally lived experiences, and general sparing of conscious retrieval of knowledge as well as more automatic, unconscious remembering.

Theories have been proposed to account for age-related decline in episodic memory, emphasizing the roles of speed of processing, inhibition, processing resources, and associative memory. The speed of processing account relies on evidence of strong correlations between memory deficits and reductions in processing speed with age (Salthouse, 1996): Performance on effortful aspects of memory require a sequence of cognitive operations; if this process takes time, the products from earlier operations dissolve before they can be used for later operations. The inhibitory deficit account (Hasher & Zacks, 1988), on the other hand, proposes that memory is hindered by products
Chapter 1. Introduction

that did not dissolve: Irrelevant information is not successfully inhibited by older adults, creating interference which reduces target memory. The theory of reduced processing resources (Craik & Byrd, 1982) underscores the fact that age differences are especially pronounced in more difficult tasks that require self-initiation. Such effortful tasks tax a pool of attentional resources among older individuals that is already smaller than that of younger adults, leading to less efficient processing. When we shrink the pool of resources among younger adults through divided attention at encoding, their performance approximates that of older adults (Anderson, Craik, & Naveh-Benjamin, 1998). The associative deficit account provides a caveat to the latter by pointing to evidence that even when we equate processing resources among both age groups, older adults still perform worse on memory tasks that require binding of information (Naveh-Benjamin, 2000). All of these theories have their own explanatory strengths and weaknesses, but they are united in postulating the importance of cognitive control for initiating and maintaining operations necessary for memory. Control over cognitive operations is under the jurisdiction of executive functioning and the frontal lobes (Stuss, Craik, & Sayer, 1996), which are known to be less efficient with aging (West, 1996).

This profile of age-related memory change is critical to understanding the effects of error generation on episodic learning. At first glance, it would seem that errors would exacerbate age-related vulnerabilities to interference and source misattributions resulting from reduced cognitive control. However, learning errors are not all created equal, and differ on a number of dimensions that can potentially modulate their effects on aging episodic memory. Before reviewing the evidence, it is important to define what is meant by episodic learning errors, and how they differ from other memory errors.

1.1.2 Memory errors

Enhancing learning and memory can be a question of how to increase the quantity of target mnemonic content, e.g., memorizing more programming code, more vocabulary, and more names of individuals. However, successful learning is not merely a question of getting correct information to stick, but also of making sure that what is incorrect does not. Some errors are introduced because of memory distortions resulting from processes inherent to encoding and retrieval. For example, in the Deese-Roediger-McDermott paradigm (Roediger & McDermott, 1995), younger – but especially older – adults are likely to falsely remember that a word (e.g., sleep) was studied among a list of semantic associates (e.g., dream, bed, slumber). These false memory errors are the result of processing distracters that are external to the studied information. The word sleep was not encountered during learning: It is a product of spreading activation from processing semantically related words. Conversely, some memory errors are the result of interference caused by concurrent distracters (for a review, see Lustig, Hasher, & Zacks, 2007). Older adults are more likely to over encode distracting information relative to younger adults, like processing the meaning of words.
when only their ink colour is relevant to the task (Gopie, Craik, Hasher, 2011). This is evidenced when memory for the distracter words is measured with both indirect and direct tests of memory (Gopie et al., 2011). Memory errors can also arise when information that was previously relevant is not deleted from the current focus of attention (Lustig et al., 2007). Older adults are vulnerable to both proactive and retroactive interference relative to younger adults, and produce more memory errors in the form of intrusions (May, Hasher, & Kane, 1999; Kane & Hasher, 1995). In sum, errors can take the form of false or veridical but inappropriate memories, and aging leads to an increase in the frequency of both.

A different source of errors, however, is the focus of my dissertation: Wrong guesses generated by learners during acquisition of new information, i.e., trial-and-error learning. For example, what is the effect of initially incorrectly guessing that the capital of Australia is Sydney on ones later memory for the correct answer (i.e., Canberra)? What is the memorial consequence of incorrectly guessing at a persons profession before learning it? In these situations, the individual generates guesses prior to receiving feedback, and these guesses are coined learning errors once they are known by the learner to be incorrect. However, the two examples above differ in the extent to which learners can bring prior knowledge to bear to make guesses more or less plausible. In the case of learning capitals, guessing is constrained by a persons knowledge of geography; guessing a persons profession, however, is much more arbitrary and relies on provided cues (e.g., She wearing a suit, so probably works in an office). Understanding the effects of trial-and-error learning on memory is therefore complicated by individual and age-related differences in knowledge and experience. However, learning of entirely novel and somewhat arbitrary associations like word pairs (e.g., desk – computer) can circumvent these issues to get at the core mechanisms of error resolution. Studies 1 and 2 of my dissertation take this latter approach to examining learning errors, whereas Study 3 looks at errors that are constrained by general knowledge.

1.1.3 Diverging views on episodic learning among younger and older adults

The critical question then becomes: Do these learning errors help or harm episodic memory? To learn from ones mistakes is an adage often recounted in educational settings and has currency according to much experimental work among younger adults. As we will see, much research has found that generating errors during learning affords memorial benefits for correct information relative to when errors are avoided. For example, an early study by Kane and Anderson (1978) found that participants better remembered the endings of determined sentences (e.g., The dove is a symbol of ___) when they had been asked to guess beforehand relative to if it was read in full. As will be discussed, many have argued that trial-and-error learning is more active and engaging than passive reading, leading to deeper and more elaborative encoding. For younger adults, learning errors
may foster a learning environment that is amenable to episodic memory formation. By contrast, evidence from the aging literature has advocated the opposite for older adults, namely that errors should be avoided in service of memory. For older adults, errors are thought to create interference and compete with correct answers at retrieval, taxing already compromised processing resources, inhibitory abilities and cognitive control. A more fitting idiom for aging individuals according to this literature may then be that you cant teach an old dog new tricks. In sum, research conducted separately among both age groups has led to opposing conclusions regarding the effects of episodic errors on memory.

My main thesis is that this disparity is not scientifically warranted. As I will show, a comparison of the two literatures yields a puzzling paradox: the purported mechanisms underpinning error benefits among younger adults are the very same that are portended to minimize age differences in episodic memory. Specifically, encoding that emphasizes semantic processing is known to enhance memory among both age groups (Craik & Lockhart, 1972), and may lead to greater improvements among older adults (Backman, 1986; Craik & Simon, 1980; Park, Smith, Morrell, Puglisi, & Dudley, 1990). It follows that if generating errors affords elaborative processing as the experimental literature suggests, then older adults too should benefit from them. The fact that the conclusion from the aging literature stands in opposition to this suggests two possibilities: 1) The noise created by errors overpowers any elaborative benefits they may supply, leading to a net harm for older adults; 2) Methodological differences between the two literatures have led to divergent conclusions. To foreshadow, evidence presented here will primarily support the latter.

The objective of my dissertation is to bridge these two solitudes and systematically examine the effects of learning errors on episodic memory in healthy younger and older adults. I propose that errors can be either beneficial or harmful for both age groups, and that the deciding factor is the type of processing engendered by error generation, whether conceptual or nonconceptual. A review of the existing research will show that the younger and older adult literatures have come to divergent conclusions using learning paradigms that differ precisely along this critical dimension. Studies among younger adults typically had participants learn from conceptual, meaning-based cues (e.g., *A word related to ‘desk’?*), and find a benefit of making errors. By contrast, studies among older adults are also those that typically use nonconceptual, form-based cues (e.g., *A word that begins with the letters ‘br’?*) and consequently find that making errors is detrimental. I will present studies that untangle this confound to elucidate how age and processing requirements independently influence error processing. But first, let us review these literatures in turn to clarify the source of discrepancy.
1.2 Errors are bad: Overview of the older adult literature

1.2.1 The errorless learning benefit

The consensus of error avoidance for healthy older adults has its origins in a clinical literature that sought to rehabilitate memory among populations with more pronounced memory impairments. Older adult groups were included as a baseline in studies where learning paradigms were tailored to retrain abilities vulnerable to pathological rather than healthy aging (e.g., learning of single words, face-name associations, spatial routes). While errors in these studies are true learning errors, they are generated in the context of acquiring inflexible and nonconceptual relations.

In a seminal study, Baddeley and Wilson (1994) had healthy older and amnesic individuals learn a list of words under either trial-and-error learning or so-called errorless learning. In the trial-and-error learning condition, participants were told: “I’m thinking of a five-letter word that begins with ‘qu’. Can you guess what it might be?” Following the generation of one to four guesses on behalf of the participant (e.g., quill? queen?), the experimenter would provide the correct answer: “No, good guesses, but the word is ‘quote’, please write that down.” In the errorless condition, errors were avoided by simply providing the target from the onset: “I’m thinking of a five-letter word that begins with ‘qu’ and the word is ‘quote’. Please write that down.” For both groups, errorless learning led to better subsequent target memory on a cued recall task relative to trial-and-error learning; younger adults were also included in this study but their memory performance was at ceiling in both learning conditions, making it impossible to draw conclusions. These findings reverberated in clinical circles and spawned enthusiasm for the potential of errorless learning to retrain a broad set of skills including word processing (Hunkin, Squires, Aldrich, & Parkins, 1998b; Jokel & Anderson, 2012), and learning of proper names (Parkin, Hunkin, & Squire, 1998) and face-name associations (Clare, Wilson, Carter, Roth, & Hodges, 2002; Kalla, Downes, & Van den Broeck, 2001) among various memory-impaired populations and healthy older adults (for a review, see Clare & Jones, 2008; Middleton & Schwartz, 2012).

1.2.2 Cognitive mechanisms of the errorless learning benefit

Implicit and explicit memory processes

To account for the benefit of error avoidance, Baddeley and Wilson (1994) argued that errorless learning reinforces solely the correct target at encoding: Either implicit or explicit memory processes can lead to correct retrieval of the target. Thus, errorless learning capitalizes on the relatively intact implicit processes of older adults and individuals with amnesia. By contrast, trial-and-error learning introduces the need to discriminate between errors and targets at retrieval, something that can only be accomplished via explicit memory processes, which are compromised in these populations. Findings from other studies have upheld this interpretation of errorless mechanisms
Chapter 1. Introduction

(Anderson & Craik, 2006; Anderson, Guild, Cyr, Roberts, & Clare, 2012; Evans et al., 2000), although conflicting reports exist (Hunkin et al., 1998b; Tailby & Haslam, 1003). For instance, Hunkin et al. (1998b) examined correlations between the magnitude of the errorless benefit, and direct and indirect measures of explicit and implicit memory respectively, among patients with amnesia. They found no support for the hypothesis that a greater errorless benefit was related to better performance on indirect tests of implicit memory (for a similar conclusion, see also Tailby & Haslam, 2003). However, assessing memory by means of separate direct and indirect measures fails to capture the reality that both explicit and implicit processes likely contribute to memory performance in any given task. Moreover, it is well known that measures of both implicit and explicit memory are not process pure (Jacoby, Toth, & Yonelinas, 1993). Therefore, this approach limits our ability to tease apart the relationship between mnemonic mechanisms and the memorial benefits of error avoidance.

Anderson and Craik (2006) bypassed these drawbacks when they examined errorless learning among healthy younger and older adults. They employed a modified version of Hay and Jacobys (1996, 1999) habit forming paradigm which pits implicit and explicit processes against each other and allows for estimations of their respective contributions to performance. In an initial habit forming training phase, a word (e.g., knee) was presented in one of two pairings (e.g., knee – bone or knee – bend). Pairings were rendered typical or atypical by manipulating presentation frequency: For example, if knee – bone accounted for 75% of pairings shown with the word knee, and knee – bend the remaining 25%, they would respectively be referred to as typical and atypical word pairs. A study phase then followed, which presented word pairs that were either made typical or atypical in the training phase. The key here is that some to-be-remembered word pairs were also entrenched by habit in the training phase (e.g., knee – bone) whereas others were not. Of interest on a later cued recall task (e.g., knee – ?) is whether learners can counteract the familiarity of the typical pairing (i.e., lure) when a studied atypical word pairing is to be recalled. In this case, good performance relies on recollection, i.e., consciously remembering the pair along with contextual details. The learner must explicitly remember the correct word from the study phase to successfully suppress the more familiar typical response. However, if a typical word pairing is to be recalled, good performance can be guided by recollection or familiarity, i.e., automatic remembering of the item but devoid of contextual details. In the latter case, the correct word may come to mind simply because it was made very familiar by the training phase, or because the learner explicitly remembers it from the study phase.

In their study, Anderson and Craik (2006) estimated the contributions of familiarity and recollection following this learning paradigm, with and without error generation. The original paradigm served as the errorless learning condition, whereas a trial-and-error variant required participants to guess which of the words would be the target during the habit forming and study phases. Estimates of recollection and familiarity were derived using the process-dissociation procedure (Jacoby, 1991)
which subtracts typical from atypical word pair memory performance to yield a familiarity score, and back solving to yield a recollection score. Echoing prior research, older adults displayed overall lower recollection than their younger peers but no difference in familiarity (Hay & Jacoby, 1999; Jennings & Jacoby, 1997). In terms of learning conditions, the results showed that errorless learning led to a reduction in familiarity for the misleading typical lures for both age groups; however, it also led to a reduction in recollection for younger adults. In sum, errorless learning dampened the familiarity of the lures relative to trial-and-error learning, which is beneficial in the context of this learning paradigm. However, this benefit was offset among younger adults by a decline in recollection, which the authors interpreted as evidence that error generation affords deeper processing among younger adults.

The findings from this study resonate with the general argument that errorless learning acts through implicit, familiarity-based channels to benefit memory among older adults. However, the paradigm used in this study is not an ecological reflection of everyday trial-and-error learning in that the correct answer changed arbitrarily during the training phase. Also, the results spoke to the activation of lures following errorless and trial-and-error learning, not targets. Nevertheless, this study foreshadowed the point that error generation can enrich a memory trace as indexed by recollection.

Frontal lobes and executive functioning

Age-related declines in frontally-mediated executive functioning (Moscovitch & Winocur, 1992) have also been proposed as central to understanding the errorless learning benefit. Inspired by Hebbian learning models, some have argued that during trial-and-error learning both errors and targets are neurally reinforced (Fillingham, Sage, & Ralph, 2005; McClelland, Thomas, McCandliss, & Fiez, 1999). When the memory is triggered at a later time, the same pattern of neural response is likely to be activated and responses must then be monitored for accuracy on the basis of prior feedback for targets and errors to be distinguished. Hence, older adults may display difficulties in differentiating targets from learning errors due to error-monitoring deficits which are under the domain of executive functioning (Mitchell & Johnson, 2000; Baddeley, 2000; Wagner, 2002). The idea that older adults may have difficulty discriminating targets from self-generated errors is congruent with findings that older adults are prone to confusing actual and imagined experiences (Johnson, Hashtroudi, & Lindsay, 1993), a failure of frontally mediated source memory (Glisky, 2001). Perhaps a parsimonious way of relating errors to frontal/executive functioning is with the concept of cognitive control (Stuss, Craik, & Sayer, 1996). Error resolution is likely a function of simultaneously activating the target and suppressing of errors, and cognitive control fuels both the initiation and inhibition required to accomplish these functions. In sum, compromised frontal functioning with age may lead to greater difficulties in error-monitoring, source discrimination, and cognitive control processes required for error resolution more broadly.
With respect to learning errors more specifically, however, our prior work examined the independent contributions of executive and episodic memory functions to the errorless learning benefit among older adults (Anderson, Guild, Cyr, Roberts, & Clare, 2012). We had participants learn lists of words in a paradigm identical to Baddeley and Wilsons (1994), followed by tests of free recall, cued recall, and recognition. Critically, older adults had previously completed the Glisky neuropsychological battery which yields composite scores to reflect cognitive abilities mediated by frontal (i.e., executive) or medial temporal (i.e., explicit memory) functioning. Therefore, older adults in this study were categorized as high or low in frontal lobe (FL) and medial temporal lobe (MTL) domains based on their performance, giving four neuropsychological profiles. Results showed an errorless advantage in free and cued recall, and that those with low-MTL function benefited marginally more. The low-MTL group was also more likely to erroneously classify their errors as targets in the recognition task. This suggests that MTL rather than FL function is at the core of the errorless advantage. Interestingly, it was not the case in either the free or cued recall tasks that prior errors were produced as targets, arguing against a proactive interference explanation which would predict a higher number of prior error intrusions (see also Lubinksy, Rich, & Anderson, 2009). Rather than being mistaken as targets, errors appear to add noise at retrieval which dampens the strength of the target signal.

Dopamine and error processing

The activity of dopaminergic neurons in the anterior cingulate cortex plays a role in a number of functions critical to learning, including the detection and processing of errors (Botvinick, Cohen, & Carter, 2004; van Heen & Carter, 2002a; Scheffers & Coles, 2000). Given that dopaminergic function is known to undergo pronounced change with age (for a review, see Braver & Barch, 2002), it is a potential cause of age-related declines in the ability to resolve error interference. To this point, an electrophysiological study by Eppinger, Kray, Mock, and Mecklinger (2008) had healthy younger and older adults engage in a probabilistic learning task where the validity of feedback was manipulated. They found that older adults were less affected by negative feedback as indexed by the feedback-related negativity, a potential emanating from the anterior cingulate cortex (but see Ferdinand & Kray, 2013). In sum, normal aging may be associated with a change in how our brains respond to errors, which is at the core not only of reinforcement learning but also of trial-and-error episodic learning.

1.2.3 Summary of the older adult literature

Taken together, studies from the neuropsychological literature have converged on the finding that errors are generally harmful for those experiencing age-related memory decline. For older adults, errors in episodic learning tax already compromised explicit and dopaminergic processes required
to discriminate between right and wrong information. However, it is important to remember that the studies cited above have looked at learning in contexts that are known to disadvantage older adults. Most have had participants learn from form-based cues (e.g., a word that begins with qu) that draw attention to the superficial features of memoranda. This kind of item-specific processing is known to be more amenable to implicit memory, and may explain in part why performance hinges so heavily on it (Clare & Jones, 2008). It remains to be seen whether the errorless benefit holds up in learning paradigms that are known to minimize age differences in memory, e.g., learning that is conceptual and meaning-based. These conceptual learning paradigms have been widely favoured in the younger adult literature, to which we now turn.

1.3 Errors are good: Overview of the younger adult literature

1.3.1 Retrieval-based learning

The memorial effects of error generation among younger adults are best understood within the framework of retrieval-based learning more generally. Broadly, retrieval-based techniques aim to enhance learning by harnessing the fact that retrieving information modifies memories (Bjork, 1975). One reason retrieval is thought to modify memory is because it often encourages deep processing, a notion which has wide currency among cognitive psychologists and educators. Fostering deep processing of verbal material typically involves drawing the learners attention to the meaning of information under study as opposed to its superficial features (Craik & Lockhart, 1972). Indeed, learning is not a question of copying and storing information, but rather mapping it on and linking it to information already in memory. As such, deep encoding conditions that lead to successful memory are often those which activate or retrieve stored concepts. One example is the generation effect (Slamecka & Graf, 1978), the finding that generating a word (e.g., hot − c_) leads to better memory than simply reading it (e.g., hot − cold). The act of retrieving the answer forces the learner to attend to the meaning of the association, and invokes desirable difficulty (Bjork, 1994). Relative to reading, retrieval is an effortful and constructive process which strengthens and modifies the memory trace (Bjork, 1975). Importantly, retrieval need not be so targeted to the specific answer to be advantageous. As we will see, retrieving studied information is beneficial for remembering it later on (e.g: answering a question that we studied previously), but so is retrieving unstudied information that is related to it (e.g: answering a question about studied materials).

However, increasing the difficulty of encoding can come at a cost because the retrieval process is fallible and wrong information can be entrenched. Research on testing effects and error generation capture this paradox: both introduce beneficial retrieval and potentially harmful errors, the balance of which is tipped by external factors (e.g., study-test formats, study materials, environmental
context) and internal factors (e.g., age, prior knowledge, mood, confidence).

1.3.2 The testing effect

The testing effect refers to the finding that delayed memory for studied information is better if it was tested rather than restudied. Standard testing paradigms follow a Study-Test-Retest or Study-Restudy-Test procedure, where the critical contrast is retention of items on the final test as a function of whether they were tested or restudied. In a representative study, Karpicke and Roediger (2008) had younger participants study Swahili-English word pairs, followed by a succession of restudy (mashua – boat) or test (mashua – ?) opportunities. Students in one condition underwent the standard testing paradigm of repeated study-test trials (ST). In the other three conditions, once the student had successfully recalled the word, it was kept for further study but dropped from testing (STn), kept for further testing but dropped from study (SnT), or dropped from both study and testing (SnTn). Performance on a cued recall test one week later showed that students who had engaged in repeated testing of items (ST and SnT) did markedly better than those who repeatedly restudied (STn) or neither restudied nor tested (SnTn).

This effect has been widely replicated among younger adults, and it has also been demonstrated that testing effects are stronger if retrieval is more effortful and difficult (Karpicke & Roediger, 2007; Pyc & Rawson, 2010; Carpenter, 2009). For instance, using a free recall format for the test leads to greater testing gains relative to recognition (Kang, McDermott, & Roediger, 2007); indeed, recognition tests can invoke more false alarms (Butler, Karpicke, & Roediger, 2008). Critical to my thesis, however, are findings that the testing effect shines through even if learner generated errors are the cost (Metcalfe & Kornell, 2007; Kang, Pashler, Cepeda, Rohrer, Carpenter, & Mozer, 2011). Indeed, while feedback maximizes testing benefits (Pashler, Cepeda, Wexler, & Rohrer, 2005; Butler et al., 2008), it does not even appear to be always necessary (Hays, Kornell, & Bjork, 2010; Carpenter & DeLosh, 2006). In sum, learning errors from the standpoint of the testing literature are viewed as assets, and, at their worst, benign for episodic memory. In these testing paradigms, errors index effortful retrieval, and effort is a powerful ally of retention.

To explain the testing effect, researchers have referenced retrieval-based learning mechanisms, namely deep processing, elaboration and desirable difficulty. However, some have provided more direct evidence that elaboration is the locus of the testing benefit by looking at what learners self-generate during testing trials. For instance, Pyc and Rawson (2010) had participants study Swahili-English word-pairs and generate mediators, i.e., keywords that look or sound similar to the Swahili cue and are semantically related to the English target. On a later memory task, they found that Swahili cues that were accompanied by the participants memory for the corresponding mediator were more likely to elicit retrieval of the target. In other words, recalling self-generated information that linked the cue and the target was beneficial for memory. Congruently, Carpenter
has shown that cue-target pairs that share a weak (basket – bread) relative to strong (toast – bread) relationship benefit more from testing opportunities because they force learners to covertly generate mediators to link them (e.g., basket – restaurant – table – bread). Overall, these findings show more directly that the semantic elaboration learners do in between seeing the cue and the target supports later episodic memory. However, there is also evidence that a semantic elaboration does not wholly account for testing benefits (Karpicke & Blunt, 2011; Blunt & Karpicke, in press; Lehman, Smith, & Karpicke, in press). In one study, Lehman et al. (in press) had students learn words and either undergo standard retrieval practice, or generate semantic associates for each word. They found that retrieval practice led to greater delayed memory relative to generating associations, suggesting that semantic elaboration is not the locus of testing benefits. They argued in favour of an episodic context account of testing benefits: Each successful retrieval event binds the retrieved item to new context features, thus multiplying the number of contexts that can be reinstated to cue retrieval on a final test in service of memory (Karpicke, Lehman, & Aue, 2014).

In sum, these studies on the testing effect are important because they reveal that retrieval-based learning is possible even in light of learning errors. Moreover, they provide evidence that the generation of semantic information plays a key role in this process, a point which has relevance for aging. However, learning errors in testing contexts remain qualitatively different from those resulting from trial-and-error learning, to which we now turn.

### 1.3.3 The error generation effect

Mistakes made in trial-and-error or error generation paradigms are incorrect retrievals generated prior to studying novel information, whereas mistakes in testing paradigms are essentially unsuccessful retrieval attempts. In error generation studies, there is no preexisting relationship known to the learner between the cue and the answer, and she must generate a guess that is more or less plausible depending on what is given by the context. In a representative study examining error generation effects, Kornell, Hays, and Bjork (2009) had younger adults study related cue-target word pairs by either reading them (e.g., factory – plant) or by first trying to guess the target (e.g., factory – ?) before seeing it. They found that making errors during learning led to better memory for the correct targets in a subsequent cued recall task relative to passively reading them. Other studies with younger adults have since replicated this error advantage for learning related word pairs (Huesler & Metcalfe, 2012; Grimaldi & Karpicke, 2012; Bridger & Mecklinger, 2013; Knight, Ball, Brewer, DeWitt, & Marsh, 2012; Vaughn & Rawson, 2012; Kornell, 2014) and more ecological materials such as facts and concepts (Kornell et al., 2009; Kornell, 2014; Richland, Kornell, & Kao, 2009; Potts & Shanks, 2013; Pressley, Tanenbaum, McDaniel, & Wood, 1991; Hays, Kornell, & Bjork, 2013). Unlike testing effects, feedback appears to be more critical for error generation benefits to appear, and better if immediate rather than delayed (Hays et al., 2013; Vaughn & Rawson,
While the usual suspects of retrieval-based learning mechanisms are summoned to explain the error generation benefit, there is evidence that the nature of the cue-target relationship is especially important. First, errors do not yield any memory benefit when the target turns out to be semantically unrelated to the cue (e.g., factory-hair) (Huesler & Metcalfe, 2012; Grimaldi & Karpicke, 2012; Knight et al., 2012). In such situations, the learner generates guesses that are linked to the cue (e.g., factory: smoke, worker) but not the target (e.g., hair): The semantic elaboration afforded by guessing does not proactively set the stage for enhanced encoding of the target. One notable exception, however, is a series of studies by Potts and Shanks (2013) where they found an error generation benefit with word pairs that were seemingly unrelated to the learner at the onset. They had younger adults learn the meaning of obscure English words using one-word definitions (e.g., *hispid* – bristly) under trial-and-error and errorless conditions. They found that learner generated errors at encoding led to greater memory gains on a delayed recognition test relative to forced-choice errors (i.e., picking the target from provided options) and passive studying. However, their use of recognition as opposed to cued recall may account for this errorful advantage: it is known to be less sensitive to retrieval-based learning manipulations (Kang et al., 2008; Chan & McDermott, 2007) and encoding conditions that enhance recollection (see Yonelinas, 2002). Indeed, a previous study failed to find an effect of errors on recognition memory among younger adults (Rodriguez-Fornells, Kofidis, & Muente, 2004). Moreover, learning vocabulary is likely inherently more engaging than learning arbitrary word pairs, and conceptually bridging the cue-target pairs in a retroactive fashion is not out of the question (e.g., The word *hispid* reminds me of *hirsute* which I know means ‘hairy’). Indeed, Potts and Shanks pointed to internal factors that may have accounted for their results, namely that the act of self-generating errors triggers curiosity in the learner for the correct target, resulting in greater attentional deployment towards corrective feedback. This idea that curiosity can contribute to error correction is in line with research showing that negative feedback attracts more attention from learners if it is surprising to them (Fazio & Marsh, 2009).

In sum, the benefits of error generation appear to hinge on the relationship between cues, errors, and targets, as well as the extent to which learners can meaningfully bind them in the memory trace. However, as Potts and Shanks (2013) alluded to, internal factors likely influence this meaning-making process, such as a learners interest in the right answer. Another internal factor likely to affect learning is the individuals confidence in their response: An answer that violates a high relative to low confidence response signals a larger discrepancy which enhances learning (see Rescorla & Wagner, 1972). As we will now see, the role of confidence in moderating error correction has emerged as an important consideration.
1.3.4 The hypercorrection effect

The role of confidence in error correction can only be examined in the context of learning information where a learners prior knowledge can be brought to bear. For example, it is unlikely that guesses produced during learning of arbitrary associations (e.g., factory – ?) are held with any amount of confidence by learners (e.g., smoke is just as probable as chimney). By contrast, errors produced during the learning of facts (e.g., What is the capital of Australia?) are more likely to range in confidence given that a persons existing knowledge can constrain possible answers (e.g., Perth is more likely to be the answer than Paris). As such, examining the role of confidence in error correction requires us to look at the relationship between self-reported confidence and memory performance within individuals, and in learning of contexts that tap general knowledge.

The hypercorrection effect describes the finding that deeply entrenched errors are more easily corrected than are loosely held errors (Butterfield & Metcalfe, 2001). Butterfield and Metcalfe (2001) asked participants to generate answers to general knowledge questions and to rate their confidence in the correctness of their answers on a scale, after which they were shown the correct answer. When participants were later asked the same questions, they were more likely to respond correctly to the questions that had produced high rather than low confidence errors. This finding contrasts with a memory strength account which would predict that errors endorsed with high confidence would have been strengthened over time, and therefore be more difficult to update. To account for the hypercorrection effect, many have suggested that feedback following high confidence errors captures attention more readily (Butterfield & Mangels, 2003; Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009): Participants are surprised by the answer and rally their resources to correct their misconception. Others have added that the subject matter surrounding high relative to low confidence errors is generally more familiar (e.g., the capital of Australia is Canberra versus the capital of Congo is Kinshasa, respectively), and that this fluency makes it easier to encode the correct answer (Butterfield & Metcalfe, 2006; Finn & Metcalfe, 2011). Along these lines, a recent study showed that prior knowledge was more predictive of error correction than subjective confidence (Sitzman, Rhodes, & Tauber, 2014).

In sum, internal factors such as confidence can have an effect on our ability to resolve errors. The standard hypercorrection paradigm necessarily differs from those used in error generation studies, but is similar in that errors are generated prior to receiving feedback. Another thing these error generation and hypercorrection studies have in common is that they have not included groups of healthy older adults. In the next chapter, I make the case that this is an important endeavour in light of what we already know about the factors that minimize age-related decrements in episodic memory. I also present my previous work which laid the foundation for my thesis that contrary to clinical opinion, healthy older adults do not learn much differently from their errors relative to their younger counterparts.
1.4 Bridging the older and younger adult literature

1.4.1 Retrieval-based learning and aging

As discussed, retrieval-based learning is at the core of many effective techniques applied to younger adults including testing, error generation, and hypercorrection effects. Proposed explanations for all of these learning techniques invoke well known mechanisms known to boost memory: generation (Slamecka & Graf, 1978), deep processing (Craik & Lockhart, 1975), desirable difficulty (Bjork, 1994), and semantic elaboration (Carpenter, 2009; Pyc & Rawson, 2010). Interestingly, encoding and retrieval manipulations that minimize age-related episodic memory deficits also hinge on these very same mechanisms (Craik & Jennings, 1992).

The concepts of environmental support and self-initiated activities were introduced by Craik (1983; 1986) to account for the fact that older adults perform better on recognition than recall tasks. The theory proposes that older adults are less effective at self-initiating encoding and retrieval strategies to maximize memory performance, but are able to execute them if they are provided by the learning environment. This environmental support often comes in the form of manipulations that guide deeper processing or semantic elaboration at encoding, many of which have been suggested to underpin the error generation benefit: Like their younger peers, older adults derive benefits from the generation effect (Mitchell, Hunt, & Schmidt, 1986) and testing effects (Meyer & Logan, 2013; Coane, 2013; Rabinowitz & Craik, 1986). While the error generation effect in conceptual learning has not been examined in older adults, it should be expected given that guessing spurs semantic elaboration, which older adults are less likely to carry out if left to their own devices (Luo, Hendriks, & Craik, 2007).

The suggestion that older adults should show an error generation benefit is in contradiction to the conclusions outlined earlier and the fact that errorless learning is widely viewed as best practice when teaching older adults new information (Clare & Jones, 2008). However, differences in learning paradigms between the error generation and errorless learning literatures, among younger and older adults respectively, represent a major confound. Specifically, paradigms that have reported an advantage of error generation have invoked processing that conceptually binds the cue, errors, and target (e.g., factory → smoke? → plant). By contrast, errorless learning paradigms almost always invoke shallower processing given that the cue, errors, and target are largely bound by form-based similarities (e.g., qu → queen? → quote). Errors in these contexts do nothing to constrain later retrieval as they bear no meaningful relationship to the target or cue, and cannot be strategically culled. Instead, errors may be simply reinforced along with targets at encoding (Baddeley & Wilson, 1994; Anderson & Craik, 2006; Evans et al., 2000; Lubinsky et al., 2009), increasing interference in the absence of a conceptual framework upon which to organize encoding or retrieval.

In sum, errorless learning paradigms have examined errors that afford undesirable difficulty, catering to the rehabilitation needs of individuals undergoing pathological rather than normal aging.
It remains unclear how healthy older adults would perform under conditions that build on their strengths rather than their weaknesses, and the following work was my first attempt to answer this question.

1.4.2 Evidence of an error generation effect among older adults

In my Masters work, we found the first evidence across two studies that older adults can benefit from errors like their younger peers when error generation is conceptual and meaning-based (Cyr & Anderson, 2012). The general methodology used in the study phase is similar to those used in the dissertation studies that follow. We were interested in the contributions of recollection and familiarity to memory for targets acquired via errorless and trial-and-error conceptual learning. If error generation invokes broad conceptual linkages among the cue, errors, and target, it should be associated with improved recollection which is sensitive to depth of processing manipulations (see Yonelinas, 2002). We further predicted that older adults would show a larger boost in recollection given that it would guide them in semantic processing that they are less likely to self-initiate than their younger peers (Luo et al., 2007).

In study 1, we looked at learning of words from conceptual cues acquired through errorless and trial-and-error learning, and derived estimates of familiarity and recollection to memory performance. During a study phase, participants were initially presented with a broad semantic cue (e.g., A type of tooth) and were either simultaneously presented with a word fragment of the target word in the errorless condition (e.g., m_l_r), or were required to generate two guesses in the trial-and-error condition (e.g., canine? incisor?) before the word fragment was presented. Thus, the target word in both conditions was generated by participants before it was shown in full (e.g., molar). In the event that participants generated the assigned target word, the experimenter provided a different one to ensure that trial-and-error learning always yielded two errors. A recognition test followed the study phase, and a process dissociation procedure (Jacoby, 1991) was applied to estimate the contributions of recollection and familiarity to target memory. We applied a variant of the process dissociation procedure known as ‘double exclusion’ (Yonelinas & Jacoby, 1994, 1995) where recollection was defined as the ability to recall the target and the encoding context in which it was studied and reject another encoding context (cf., Luo et al., 2007). Three recognition tests were administered to extract estimates for the errorless and trial-and-error learning conditions by pitting target responses (for which correct ‘yes’ responses could be supported by recollection or familiarity) and errors or nontargets (for which incorrect ‘yes’ responses could only be supported by familiarity in the absence of recollection) against each other. Each recognition list consisted of targets from the errorless and trial-and-error conditions as well as lures that were related or unrelated to the studied semantic categories. Importantly, the experimenter coded the first error generated by participants for categories in the trial-and-error condition (e.g., canine) and included them in the recognition
lists. As such, we could look at the likelihood that participants incorrectly endorsed their own unique prior guesses as targets.

We hypothesized that for both age groups making errors at encoding would enhance recollection of target words relative to errorless learning. Moreover, we predicted that this increase would be more pronounced among older relative to younger adults, reflecting greater gains afforded by semantic elaboration at encoding. Results from study 1 are shown in Figure 1.1 on the next page.

![Figure 1.1: Estimates of familiarity (left panel) and recollection (right panel) as a function of age and encoding context in Cyr and Anderson (2012).](image)

Following errorless learning, we found an age-related deficit in recollection but age-equivalence in familiarity, consistent with previous research (Anderson & Craik, 2006; Jennings & Jacoby, 1997). Relative to errorless learning, recall of words learned via trial-and-error was mediated more by recollection and less by familiarity for both age groups. Congruent with Anderson and Craik (2006), we interpreted the boost in recollection in light of the beneficial elaboration provided by error generation at encoding, and the decrease in familiarity in light of reduced fluency at the expense of contextual details. Consistent with our hypothesis, older adults showed greater gains in recollection from errorless to trial-and-error learning relative to younger adults. However, this benefit was offset by an increase in false alarms to their own prior errors and related lures, which is consistent with evidence that older adults are more susceptible to source-based memory errors (Johnson et al., 1993) and intrusions of irrelevant information at retrieval (Hasher, Zacks, & May, 1999; Hay & Jacoby, 1996).

In study 2, we explored the idea that conceptual error generation increases recollection by making the encoding context more rich and distinctive. Indeed, both recollection and source judgments require the reinstatement of contextual information, and age-related deficits in recollection are
closely related to age-related deficits in source monitoring (Spencer & Raz, 1995). Therefore, we opted for a more direct assessment of source memory in study 2. The learning phase was identical to that used in study 1, but the test phase involved only one test list. During recognition, participants were required to respond whether the word was a target studied without errors, a target studied via trial-and-error, a prior error, or a new word. The results mirrored those of study 1: source attributions of targets were more accurate following trial-and-error learning than errorless learning, and older adults displayed this pattern significantly more than their younger peers. However, older adults were once again more likely to falsely endorse their own prior errors as targets.

These studies were the first to find an error generation benefit among older adults, and suggest that error avoidance should not be embraced in all learning contexts. Findings of an errorless learning advantage stems from a clinically motivated literature that sought to retrain inflexible relations, such as individual words, among memory-impaired populations. When aging is examined in the context of the educationally-motivated learning techniques, however, we find that they can harness error generation in service of episodic memory via the same retrieval-based learning mechanisms operating among their younger peers.

1.5 Thesis overview

The goal of the dissertation studies that follow is to explore the effects of error generation on episodic memory among healthy younger and older adults. Study 1 is a direct test of the ideas proposed by my Masters work: Are the effects of episodic errors on younger and older memory a function of the type of processing engendered by error generation? I had participants learn words based under errorless and trial-and-error learning instructions, and based on cues that were either conceptual (e.g., a flower: rose, orchid, tulip) or nonconceptual (e.g., br___: bridge, bride, brick). I predicted that when error generation is meaning-based, both younger and older adults would benefit relative to errorless learning. By contrast, when error generation is form-based, both age groups should do better when errors are avoided. In line with my earlier work, I hypothesized that this dissociation would be larger among older adults. To foreshadow, all but the latter prediction were confirmed: both age groups showed a double dissociation between the effect of errors on target memory and the type of processing. In addition, I provide compelling evidence that these effects on target memory are linked to memory for prior errors. When a conceptual target is successfully recalled, it is likely to be accompanied by retrieval of more rather than fewer errors; when a nonconceptual target is successfully recalled, it is likely to be accompanied by retrieval of fewer rather than more errors. Put otherwise, errors are enablers in the former and interferers in the latter. I propose that conceptual errors act as stepping stones that scaffold retrieval toward the target. Nonconceptual errors, however, are not activated in concert with the target as they do not belong to the same semantic network and thus cannot provide scaffolding.
In Study 2, I explored this idea that relatedness between the error and the target is at the crux of the error generation benefit seen for memory of meaning-based associations. If errors act as stepping stones to the target, then making them closer together should be beneficial for memory. In Study 2A, I explored this by manipulating the extent to which sentence stem cues constrained possible word endings: Later memory for the target word endings was greater under trial-and-error relative to errorless instructions for both age groups, but this benefit did not interact with cue constraint. Study 2B examined semantic relatedness between errors and targets more directly by having younger and older adults learn related homograph-target word pairs with and without errors, and manipulating whether the target was related or unrelated to the meaning of their generated error. We replicated the error generation benefit among both groups, but only younger adults showed an enhancement for related relative to unrelated targets following trial-and-error learning, suggesting that older adults may not be as affected by semantic relatedness due to more flexible semantic networks.

In Study 3, I moved away from learning of arbitrary relations and examined error correction with more ecologically relevant stimuli: general knowledge facts. A qualitative difference between learning from conceptual and nonconceptual cues in the studies I have mentioned is that the former presumably engages learners more readily. Selection of a guess based on meaning (e.g., *a flower*) more likely taps prior knowledge than if it is based on form (e.g., *br____*). Study 3 goes farther on this continuum by examining misconceptions, i.e., errors that are held with some degree of confidence by learners. As previously explained, error correction for this kind of information is mediated by confidence among younger adults. Specifically, initial errors held with higher confidence are more likely to be corrected on a delayed test, a phenomenon dubbed the hypercorrection effect. Study 3 is the first investigation of error confidence among older adults. In Study 3A, I find that older adults also display a hypercorrection effect using the standard paradigm, but to a lesser extent than their younger counterparts. In Study 3B, I used a different approach to minimize methodological factors that may disadvantage older adults in the standard paradigm, which eliminated age differences in the magnitude of the hypercorrection effect.

Overall, these studies suggest that the parameters and mechanisms that influence the effect of error generation on episodic memory are the very same among healthy younger and older adults. The errorless learning advantage can be reframed as the benefit of avoiding errors in contexts that provide no conceptual context, regardless of age. When the playing field is levelled, both younger and older adults can benefit from errors thanks to processes afforded by retrieval-based learning. This dissertation will end by discussing the results within the broader literature, as well as the introduction of a framework of error generation and its effects on episodic memory from both a cognitive and neural perspective.

Please note that all of the dissertation studies that follow are either published or have been submitted for publication, and the manuscript format has been preserved. As such, there will be
redundancies across chapters.
Chapter 2

Dissociation between type of processing and effects of error generation on memory

Errors have long been the allies of educators as catalysts of successful learning. For instance, when students are asked questions about a topic before they begin to study it (to which they always receive negative feedback), they learn more from the subsequent study opportunity than do students who are shown the same questions but are not required to answer them (Richland, Kornell, & Kao, 2009; Pressley, Tanenbaum, McDaniel, & Wood, 1991). Even forced guessing does not appear to interfere with learning among students, provided that feedback is given (Pashler, Rohrer, Cepeda, & Carpenter, 2007; Kang, Pashler, Cepeda, Rohrer, Carpenter, & Mozer, 2011).

In cognitive research, the effect of errors on memory is typically examined in the context of learning related word pairs (ex: wing – eagle). In a representative study, Kornell, Hays, and Bjork (2009) had participants study semantically related word pairs under conditions that either elicited or avoided self-generated errors. In the generation condition, participants were asked to produce a word related to the first word in the pair (factory – ?) before being shown the correct pairing (factory – plant). This was contrasted with a study condition in which participants read the word pairs presented in full, thus bypassing errors. Across a number of studies, they found better cued recall performance if the initial study involved error generation. They, along with others, have argued that errors foster elaborative encoding by creating associations with the study material (Cyr & Anderson, 2012; Huesler & Metcalfe, 2012; Kane & Anderson, 1978), and set the stage for enhanced processing of the feedback (Fazio & Marsh, 2009; Potts & Shanks, 2013). In other words, making an error – even if it is a random guess – engages the learner in deeper processing which
ultimately rewards episodic memory (Craik & Tulving, 1975; Craik & Lockhart, 1972).

However, errors have earned a conversely negative reputation among researchers interested in memory rehabilitation and cognitive aging (Middleton & Schwartz, 2012; Clare & Jones, 2008). In a seminal study, Baddeley and Wilson (1994) examined memory for a list of words among persons with amnesia as well as healthy younger and older adults following errorless and trial-and-error learning. In the errorless condition, participants were told “I am thinking of a word that begins with QU and it is ‘quote’.” Under trial-and-error instructions, they were told “I am thinking of a word that begins with PE, can you guess what it is?” and were given the correct word after the participant made four incorrect guesses (or 25 seconds had elapsed). Both people with amnesia and healthy older adults showed an errorless advantage in memory, i.e. greater cued recall performance following errorless relative to trial-and-error learning (younger adults data were not interpretable due to ceiling effects). We and others have since replicated this errorless benefit using similar learning paradigms among older adults (Anderson, Guild, Cyr, Roberts, & Clare, 2012; Lubinsky, Rich, & Anderson, 2009) and memory-impaired populations (for a review, see Clare & Jones, 2008). Our data furthermore support Baddeley and Wilson’s argument that errorless learning capitalizes on the intact implicit memory of older adults as it reinforces a single correct response (Anderson & Craik, 2006; Lubinsky et al., 2009; Guild & Anderson, 2012; Cyr & Anderson, 2012; but see Tailby & Haslam, 2003; Hunkin, Squires, Parkin, & Tidy, 1998). By contrast, discriminating errors from targets following trial-and-error learning requires the engagement of episodic memory, which is known to be compromised in older adults. On the basis of these findings, people with poorer episodic memory, including healthy older adults, have been advised categorically to avoid making errors in service of episodic memory. Thus, the young adult literature and aging/clinical literature have converged on opposing views regarding the effects of errors on memory: beneficial among the former and detrimental among the latter.

Differences between the younger and older adult literatures

While it may be tempting to point to age as the pivotal factor in determining harm or benefit of errors, a closer look at these literatures reveals another important methodological difference. Among younger adults, educational interests have motivated the use of learning materials that mirror those employed in the classroom to study errors, e.g., learning of facts and vocabulary (Kornell et al., 2009; Kornell, 2014; Pashler et al., 2007; Pressley, 1991; Richland et al., 2009; Potts & Shanks, 2013; Karpicke & Roediger, 2008; Kang et al., 2011). Even in paradigms with less ecological validity, younger adults typically emphasize learning of conceptual, albeit arbitrary, cue-target relationships (e.g., factory – plant) (Kornell et al., 2009; Huesler & Metcalfe, 2012; Grimaldi & Karpicke, 2012; Vaughn & Rawson, 2012). On the other hand, paradigms in healthy aging and groups with memory impairment have been an offshoot of endeavours to retrain single word learning (Tailby & Haslam, 2003; Hunkin et al., 1998; Jokel & Anderson, 2012; Page, Wilson, Shiel, Carter, & Norris, 2006;
Baddeley & Wilson, 1994), face-name associations (Dunn & Clare, 2007; Ruis & Kessels, 2005; Kalla, Downes, & van den Broek, 2001; Evans et al., 2000) and procedural skills (Glisky, Schacter, & Tulving, 1986; Kessels, van Loon, & Wester, 2007), all of which are less conceptually driven tasks. In sum, studies that find a benefit of errors typically examine conceptual learning among younger adults; those that find a detriment of errors typically examine nonconceptual, and often lexical learning, among older adults and memory-impaired individuals.

Experimental evidence in support of a dissociation between the effects of conceptual and nonconceptual error generation on episodic memory

In this study, we aim to show that younger adults can display an errorless benefit in lexical learning, and conversely that older adults can show an error generation benefit in conceptual learning. While aging studies largely espouse errorless learning, there is support for the notion that errors can boost memory among healthy older adults. In our previous work (Cyr & Anderson, 2012), we had younger and older participants generate guesses from conceptual cues (e.g., A farm animal: “Is it horse? Is it cow?”), and found that trial-and-error learning led to greater recollection for correct targets than if the target was provided immediately by the experimenter (i.e., errorless learning). Moreover, this boost afforded by errors was greater among older adults, eliminating age differences in recollection. In a second experiment, we replicated this finding using a source memory task, revealing that the retrieval of targets from the trial-and-error condition was accompanied by more contextual details of the study episode than was retrieval of targets from the errorless condition, particularly for the older adult group. We suggested that conceptual error generation can minimize age-related episodic memory declines via the very mechanisms championed by the experimental literature; namely, encoding conditions that promote semantic elaboration are known to minimize age differences in episodic memory (Craik & Rabinowitz, 1985; Rabinowitz, Craik, & Ackerman, 1982) or even eliminate them (Luo, Hendriks, & Craik, 2007). In conceptual learning (A farm animal), error generation inherently creates associations (“Is it horse? Is it cow?”), for richer and more distinctive encoding that is amenable to later recollection (Mitchell & Johnson, 2000; Jacoby, Shimizu, Daniels, & Rhodes, 2005; Jacoby, Kelley, & McElree, 1999). Conversely, guessing based on lexical cues (A word that begins with QU), simply generates competitors that share surface-level features (Queen, Quote) as opposed to meaning. Under these ‘noisy’ conditions, age-related susceptibility to error interference may come into play (Anderson & Craik, 2006; Lustig & Hasher, 2001; Hay & Jacoby, 1996) and overwhelm any memorial benefits of effortful retrieval (Gardiner, Craik, & Blesadale, 1973). Therefore, the processing requirements of the learning paradigms espoused in the aging and clinical literature may have engendered the negative influence of error generation on memory. One study among individuals with Alzheimers dementia illustrates this point nicely. Haslam, Gilroy, Black, and Beesley (2006) contrasted the effects of errorless and trial-and-error learning on memory for low-level, specific information (i.e., names), as well as high-level, more
conceptual information (i.e., the person's occupation) associated with faces. While guessing led to worse memory for names, it had no effect on memory for occupations. This suggests that learning of more flexible representations is differentially affected by errors than narrow, non-semantic representations, such as names.

There is also support for the notion that younger adults' memory can be harmed by errors. In line with the idea that they afford semantic elaboration, studies using paradigms similar to Kornell et al.'s (2009) have shown that errors fail to benefit memory when they do not share a preexisting relationship with the cue and target (e.g., pond – radio as opposed to pond – frog) (Grimaldi & Karpicke, 2012; Huelser & Metcalfe, 2012). Attending to superficial relations does not yield deep processing to the same extent as elaborating on semantic relations (Craik & Tulving, 1975), and is more amenable to implicit, recollection-based memory (Mitchell & Johnson, 2000). A recent study by Bridger and Mecklinger (2014) speaks to a potential dissociation based on processing requirements at encoding. They contrasted learning of pairings based on word cues (angel - ?) with word-stems (bro____). While guessing predictably boosted memory for targets in the word cue condition (angel – church), it reduced it in the word-stem condition (bro – brother). However, the effects of aging were not examined in that study, so it is unclear whether older adults would show a similar pattern.

2.1 Study 1

In this study, we compared errorless and trial-and-error learning in paradigms that engaged either conceptual or a lexical processing among healthy younger and older adults. Participants learned words associated with either semantic categories (a fruit) in the conceptual condition, or word stems (ﬂ____) in the lexical condition, for a later cued recall task. In the trial-and-error condition, individuals viewed the cue and generated two guesses before viewing the target word, and in the errorless condition participants viewed the cue and subsequently the target word. Our first set of hypotheses follows from our earlier discussion. First, we predicted that younger and older adults would show greater cued recall performance following trial-and-error relative to errorless learning in the conceptual condition, and the inverse in the lexical condition, suggesting that errors act as elaborators in the former and as retrieval competitors in the latter. Second, we expected this dissociation to be greater among older adults, consistent with our previous findings of greater gains in recollection and source memory from conceptual error generation among older relative to younger adults (Cyr & Anderson, 2011): We argued that making conceptual errors guided semantic elaboration, which is known to help older adults catch up to their younger peers in terms of memory performance (Luo et al., 2007). As such, we expected that conceptual errors would enhance semantic processing to minimize age differences in memory, whereas lexical errors would exacerbate age-related vulnerability to memory interference. Our study goes an additional
step to address the nature of the error benefit in conceptual learning. We have previously suggested that trial-and-error learning may have boosted recollection for targets because the errors acted as retrieval cues (Cyr & Anderson, 2012). For example, when presented with the category ‘a flower’ at cued recall, participants may use their recollection that they guessed ‘rose’ to cue their memory for the target word. This idea is similar to Pyc and Rawson’s (2010) mediator effectiveness hypothesis of testing effects. In their study, they had participants study Swahili-English word-pairs and generate mediators, i.e., words or concepts that bridge cues and targets. On a later memory task, they found that Swahili cues that were accompanied by the participants memory for the corresponding mediator were more likely to elicit retrieval of the target. In other words, recalling self-generated information that linked the cue and the target was beneficial for memory. Conceptual errors may be akin to mediators, such that memory for the correct answer varies as a function of whether or not the incorrect guess is recalled as well. Indeed, a study by Knight, Ball, Brewer, DeWitt, and Marsh (2012) found an error generation benefit in memory for related (e.g., door-exit) but not unrelated (e.g., door-shoe) word pairs, and that this pattern was mediated by the ability to recall incorrect guess on the memory test. Given that conceptual and lexical errors vary precisely on the dimension of semantic relatedness, we may expect errors to scaffold retrieval in the former—but not the latter—condition. To address this question, we had participants retrieve not only the correct targets, but also their incorrect prior guesses in the order that they were generated, insofar as possible. We hypothesized that conceptual errors would be better remembered than lexical errors, and that memory for errors would mediate differences in target memory performance. This would suggest that conceptual guesses act as stepping stones towards the target, whereas lexical guesses are not similarly embedded in the memory trace.

2.1.1 Method

Participants

Sixty five younger and 64 older adults volunteered to participate in the study. To be included in the study, individuals had to be free of major health conditions affecting cognition (e.g., stroke, dementia, uncontrolled metabolic dysfunction). We also administered the Modified Telephone Interview Cognitive Screening (Welsh, Breitner, & Magruder-Habib, 1993), and participants had to score above the cutoff of 30/40. One younger participant was excluded due to poor English fluency as identified by the Shipley vocabulary test (<1st percentile). Participants were recruited from volunteer research participant pools or responded to advertisements posted in the community and were paid for their participation. Thirty two participants from each age group were assigned to the conceptual condition, and the other 32 to the lexical condition. Table 1 displays demographic and neuropsychological descriptions as a function of age group and condition.
Table 2.1: Mean (SD) scores on neuropsychological measures as a function of age group and condition.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Younger Adults</th>
<th>Older Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Conceptual (n=32)</td>
<td>Lexical (n=32)</td>
</tr>
<tr>
<td>Age</td>
<td>22.65 (2.82)</td>
<td>22.69 (3.37)</td>
</tr>
<tr>
<td>Education</td>
<td>15.28 (1.73)</td>
<td>15.25 (2.21)</td>
</tr>
<tr>
<td>MMSE</td>
<td>29.22 (0.87)</td>
<td>29.00 (0.95)</td>
</tr>
<tr>
<td>Shipley</td>
<td>31.22 (4.13)</td>
<td>32.84 (4.46)</td>
</tr>
<tr>
<td>HADS-A</td>
<td>6.09 (3.74)</td>
<td>7.13 (4.05)</td>
</tr>
<tr>
<td>HADS-D</td>
<td>2.81 (2.47)</td>
<td>4.29 (2.90)</td>
</tr>
</tbody>
</table>

The mean age was 22.67 (SD=3.09) for the younger group and 72.52 (SD=11.04) for the older group. Years of education did not differ between age group (F <1), but scores on the Mini-Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975), were marginally higher for younger adults than those of their older peers, t(126) = 1.82, p = .07. Consistent with prior research comparing younger and older adults, vocabulary scores on the Shipley Institute of Living Scale (Shipley, 1940) were higher in the older than younger group, t(126) = 5.23, p <.001. Anxiety and depression symptomatology as measured by the Hospital Anxiety and Depression Scale1 (HADS; Zigmond & Snaith, 1983) revealed significantly higher scores among younger relative to older adults, t(125) = 2.84, p <.01, and t(125) = 2.97, p <.01, respectively. However, scores from neither groups neared the clinical cut-off score of 13. None of the measures in Table 1 differed between Condition or revealed a significant Age X Condition interaction, with the exception of HADS-D, F(1,123) = 5.99, p = .02. This was a result of higher (but still sub-clinical) scores among the younger adults in the lexical condition.2

---

1Due to a filing error, the HADS could not be scored for one younger participant.
2Rerunning all the analyses that follow while covarying scores on the HADS-A and HADS-D yielded the same pattern of results, with the exception of the Age X Condition interaction for target memory (see Results section).
Materials

A single pool of 96 nouns made up the study stimuli for both the conceptual and lexical conditions: all participants studied targets from the same set. This pool of nouns was selected such that they represented three exemplars from each of 32 semantic categories for the Conceptual study set (e.g., fabric: silk, cotton, linen), and three word completions for each of 32 two-letter word stems for the Lexical study set (e.g., co: coat, cotton, cow). Each set of 32 cues was then divided into two sets of 16 each, one for errorless learning and one for the trial-and-error learning instructions. In both conditions, the order of the nouns associated with each semantic category or word stem was arranged such that word frequency (Brysbaert & New, 2009; database available at subtlexus.lexique.org) and word length did not differ as a function of set or word position within the set, Fs < 1. The assignment of which word would be selected as an intended target (1, 2, or 3) was counterbalanced across participants.

Design and Procedure

This was a mixed study design with age and processing condition (Lexical or Conceptual) as between-subjects variables, and learning instructions as a within-subjects variable. Participants were tested individually. Stimuli were presented and participants responses were recorded by E-prime software (Psychology Software Tools, Inc., Pittsburgh, Pennsylvania).

Each individual underwent 2 study-test cycles of blocked errorless and trial-and-error learning, each consisting of 16 different categories. Order of study set (1 and 2) and learning instruction (errorless and trial-and-error) were counterbalanced across participants. **Conceptual study condition:** In errorless learning, the experimenter introduced the semantic cue as it appeared at the top of the screen by saying “The category is a pastry”, after which she pressed a key to make the target appear under the cue, saying “textit and the target is tart”. In trial-and-error learning, participants were similarly shown a category (e.g., a fruit) and were prompted to guess the target word (“Is it mango?”) after which they were given feedback by the experimenter (“No, good guess, try again!”). Participants were required to generate two guesses, after which the experimenter made the correct target appear, saying, “The target word is strawberry.”. **Lexical study condition:** Participants in the lexical condition underwent an identical procedure but with word stems as opposed to conceptual categories for cues. In errorless learning, they were shown a word stem at the top of the screen (ta____) followed by the target under it (tart). In trial-and-error learning, they were shown a word stem (st____) and had to guess which word completed it (“Is it stair?”). As in the conceptual condition, they were required to provide two guesses, after which the experimenter displayed the correct target (“The target word is strawberry.”). Across all conditions, the cue-target pair remained on the screen for 4 seconds.

For both errorless and trial-and-error learning conditions, participants were instructed to com-
mit to memory the correct target words for a later memory test. In the trial-and-error conditions, the experimenter always selected a target word that was not generated as a guess by the participant. As mentioned earlier, assignment of the intended target word to exemplar (1, 2, or 3) was counterbalanced across participants. If the participant was assigned to have the first exemplar as the target, but guessed this item during trial-and-error learning, the experimenter pressed key number two to display a word that had not been generated. As a result, all participants generated exactly two errors at study. All targets were shown for three seconds, followed by a one second inter-stimulus interval. Participants were asked to write down the targets as they were shown, making sure to cover their responses with another sheet of paper as they went along to discourage rehearsal.

A 10-minute break followed the study phase during which participants viewed a series of paintings on-screen and were asked to match them to one of two painter names. The HADS and the MMSE were also administered. Once the break had expired, participants began the cued recall task. For the errorless conditions, participants were shown the studied cues and asked to write down the correct target. In the trial-and-error conditions, the recall sheet contained the column headings Guess 1, Guess 2, and Target: Participants were required to write down not only the correct targets, but also the two guesses they had generated under their correct header. For each cue, recall order of guesses and targets was unconstrained: Participants were free to begin by reporting their guesses or the target from trial to trial. There were no new cues, and the cued recall task was self-paced. Following both study-test cycles, participants completed the Shipley vocabulary test and were debriefed and compensated. The entire session lasted approximately 75 minutes. This study was approved by the University of Toronto and Baycrest Research Ethics boards.

2.1.2 Results

An alpha level of .05 was used for all statistical tests.

Cued recall performance: Target words

We first ran a 2(Age) x 2(Learning Instruction: errorless, trial-and-error) X 2(Condition: Conceptual, Lexical) mixed ANOVA. Younger adults recalled more target words than older adults, $F(1,124) = 18.40, p<.001, \eta^2 = .13$, and memory performance was better in the Conceptual than Lexical condition for both age groups, $F(1,124) = 195.46, p<.0001, \eta^2 = .61$. There was no main effect of Learning Instruction, $F(1,124) <1$, or Learning Instruction X Age interaction, $F(1,124)<1$. However, the Learning Instruction X Condition interaction was significant, $F(1,124) = 53.65, p<.0001, \eta^2 = .30$, revealing greater cued recall performance following trial-and-error ($M=0.87$) relative to errorless learning ($M=0.76$), $t(63) = 5.50, p<.001$, in the Conceptual condition, but the opposite pattern (e.g. greater memory following errorless learning) in the Lexical condition ($M=0.48$ and
Chapter 2. Dissociation between type of processing and effects of error generation on memory

There was a significant Age X Condition interaction, $F(1,124) = 4.35$, $p = .04$, $\eta^2 = .30$, which revealed that while both younger and older adults memory was better in the Conceptual relative to Lexical condition, this was more pronounced among older adults (see Figure 2.1). However, when we covaried scores on the HADS anxiety and depression scales, this latter interaction became marginally significant, $F(1,124) = 3.61$, $p = .06$. The Learning Instruction X Condition X Age interaction was not significant, $F(1,124)<1$.

![Figure 2.1: Mean cued recall performance for targets as a function of age and condition (bars represent ± 1 standard error).](image)

Although we counterbalanced the order of learning instructions (errorless and trial-and-error) across participants, we ran a 2(Age) X 2(Order: errorless – trial-and-error, trial-and-error – errorless) X 2(Learning Instruction) X 2(Condition) mixed ANOVA to examine potential order effects on memory for targets. This revealed no significant main effects or interactions with Order (all $p$s > .05), with the exception of a Learning Instruction X Order X Condition interaction, $F(1,120) = 4.74$, $p = .03$, $\eta^2 = .04$, due to better memory for errorless lexical targets when they were studied first relative to second (i.e., following trial-and-error learning). However, a 2(Age) X 2(Order) X 2(Learning Instruction) within the Lexical condition revealed no significant main effects or interactions (all $p$s > .05). In sum, error generation effects were not modulated by order of instructions, but memory for Lexical errorless targets benefited overall from being studied first, likely due to reduced proactive interference.

Cued recall performance: Error intrusions

We then examined whether learners produced their prior guesses as intrusions when target recall was unsuccessful, confusing self-generated errors as the right answers. A 2(Age) X 2(Condition)
ANOVA revealed that older adults ($M=.42; SD=.94$) made more prior error intrusions than younger adults ($M=.11; SD=.36$), $F(1,124) = 6.47$, $p = .01$, $\eta^2 = .05$, and that they were more frequent for both groups in the Lexical ($M=.42; SD=.85$) relative to Conceptual condition ($M=.11; SD=.54$), $F(1,124) = 6.47$, $p = .01$, $\eta^2 = .05$. The Age X Condition interaction was not significant, $F(1,124) = 1.62$, $p = .21$. It should be noted, however, that the mean number of prior error intrusions for each group was less than one.

**Cued recall performance: Prior Guesses**

If errors act like stepping stones to Conceptual but not Lexical targets, they should be more accessible when targets are successfully retrieved. For each participant, we selected the trials where targets were successfully recalled, and computed the mean number of prior guesses (i.e., zero, one, or two guesses) that were correctly retrieved on those trials. We then conducted a 2(Age) X 2(Condition) mixed ANOVA, where the dependent variable was the mean number of prior guesses correctly recalled along with targets. This revealed that a significantly greater proportion of prior guesses were recalled along with targets in the Conceptual relative to the Lexical condition, $F(1,124) = 184.43$, $p < .001$, $\eta^2 = .60$, and that younger adults recalled more of their guesses overall relative to their older counterparts, $F(1,124) = 23.69$, $p < .001$, $\eta^2 = .16$. The Age X Condition interaction was also significant, $F(1,124) = 6.78$, $p = .01$, $\eta^2 = .05$, showing that prior guess recall decreased more from the Conceptual to the Lexical condition among the older relative to the younger group (see Figure 2.2).⁴

![Figure 2.2: Mean number of prior guesses recalled as a function of age and condition, conditionalized on successful recall of target word (bars represent ± 1 standard error).](image)

³We conducted the same analysis for the total number of prior guesses recalled, regardless of whether or not the target was successfully recalled, and this yielded the same pattern of results as above.
Figure 2.3 shows the breakdown of total trials where targets were recalled as a function of whether they were accompanied by retrieval of zero, one, or two guesses. Conceptual targets were more likely to be accompanied by two rather than one or zero guesses for both age groups. In the Lexical condition, younger adults maintained a trend of remembering more rather than fewer guesses whereas older adults displayed the inverse pattern.

One could contend that this relationship between target and prior guess retrieval is simply due to the fact that conceptual guesses are better remembered overall relative to lexical targets. The fact that correct recall of conceptual targets is associated with prior guess retrieval may not mean that they act as stepping stones, but merely that they are more accessible. To adjudicate these possibilities, we investigated the likelihood that remembering ones guesses predicts target recall. Therefore, we examined only the trials where participants successfully retrieved two prior guesses, and computed the proportion of these trials that were followed by successful retrieval of the target. For this analysis, the data of one younger adult in the Conceptual condition, and six older adults in the Lexical condition were excluded as they failed to successfully recall both guesses for any cue (new ns 31 and 26 respectively). We conducted a 2(Age) X 2(Condition) mixed ANOVA where the dependant variable was the proportion of these trials (i.e., those where both guesses were recalled) that were accompanied by target retrieval. This revealed a greater proportion among younger relative to older adults, $F(1,117) = 18.72, p < .001, \eta^2 = .14$, and in the Conceptual relative to Lexical condition, $F(1,117) = 158.25, p < .001, \eta^2 = .58$. The interaction was also significant, $F(1,117) = 6.20, p = .01, \eta^2 = .05$, reflecting that this pattern was stronger among older relative to younger adults. In sum, recalling both guesses was more predictive of successful target recall in the Conceptual relative to Lexical condition, especially among older adults.\footnote{Conducting the same analysis limited to trials where one guess was accurately recalled was problematic as it...}
Finally, we were interested in whether participants also preserved the temporal order of guesses at retrieval. For instance, a participant may accurately recall their prior guesses cotton and linen, but they may not remember which they had generated first and second. We selected the trials where both prior guesses were successfully retrieved, and computed the proportion that showed maintained temporal order. A 2(Age) X 2(Condition) mixed ANOVA showed no main effect of age, $F(1,117) = 1.40$, $p = .24$, and a marginally significant effect of condition, $F(1,117) = 3.23$, $p = .08$, showing that temporal order of guesses was preserved more often in the Conceptual ($M = .91; SD = .14$) relative to Lexical ($M = .84; SD = .27$) condition. The interaction was not significant, $F(1,117) < 1$.

### 2.1.3 Discussion

The consequence of trial-and-error learning on memory has obvious implications for educational testing and memory rehabilitation. From a developmental perspective, it is critical to examine whether healthy aging affects the manner in which we best acquire novel information. In this study, we found that younger and older adults do not learn differently from errors: The parameters that enhance or reduce memory are the same among both age groups. Consistent with previous reports in younger adults, trial-and-error learning led to greater memory for correct information relative to errorless learning, but only if the cue, error, and target were conceptually linked. The fact that older adults also showed this benefit is in line with our previous finding that conceptually related guesses enhance recollection and source memory for older adults (Cyr & Anderson, 2012); however, we did not replicate our finding of a larger enhancement among older relative to younger adults. Indeed, the current study reports the opposite pattern, with the younger group benefiting more from conceptual error generation compared to the older group. This may be attributed to differences in how memory was measured in both studies: The current study used cued recall whereas Cyr and Anderson (2012) used recognition, tasks which are known to exacerbate and minimize age-related differences in memory, respectively (Craik & Jennings, 1992). Recognition may provide the additional environmental support required for older adults to maximally harness conceptual error benefits. By contrast, when a nonconceptual, lexical paradigm was used, we replicated the errorless benefit reported in the clinical literature, and extended this finding to younger adults. As such, differences in the processing requirements of the learning materials, as opposed to age, may have accounted for prior conflicting recommendations in the literature with respect to trial-and-error learning.

required excluding the data from 14 younger and four older adults in the Conceptual condition as well as one older adult in the Lexical condition. However, running the analyses with the remaining participants yielded the same pattern of results.
Proposed theories on the effects of errors on memory

Our results speak to three views on how episodic learning errors can benefit memory. First, some past studies have found that if the target does not match the cue (frog – desk), error generation yields no benefits (Huesler & Metcalfe, 2012; Grimaldi & Karpicke, 2012). These results are consistent with our earlier suggestion that the relationship between errors and targets determines whether or not trial-and-error learning will lead to better target memory (Anderson & Cyr, 2012). When to-be-learned material is conceptually based (a flower), guessing generates semantic elaborators (rose? orchid?) that strengthen the memory trace for the correct cue-target pairing (a flower – iris). Our study contrasts these findings to learning in a lexical condition. In our lexical condition (br____), generated guesses selected from a lexical as opposed to a semantic network (bride? broom?). Thus, targets (brick) are among the activated candidates and are lexically related to both the cue and errors. Our findings suggest that even when the cue-target relationship is logical, trial-and-error learning can hinder memory when it does not activate a semantic network.

One exception to these findings demonstrating that error benefits are a function of semantic relatedness is the results of Potts and Shanks (2013). In a series of studies, they had participants learn one-word definitions of rare English words (e.g., valinch – tube), via either trial-and-error or errorless learning. They found greater memory for correct definitions following trial-and-error learning across all studies, suggesting that a preexisting relationship between targets and errors is not a prerequisite for error benefits. However, while errors in their paradigm are tenuously linked to targets, they still engage the learner in conceptual thinking (e.g., guessing ruler because the word valinch contains the word inch), and we suggest that errors were beneficial in a manner similar to that of learning Swahili-English word pairs, for which it was shown that memory for the mediator facilitates recall of the English target word (Pyc & Rawson, 2010). Also, Potts and Shanks queried memory using recognition as opposed to cued recall, unlike our task and those above. Relative to cued or free recall, recognition is known to be less sensitive to learning conditions that influence recollection-based memory, such as testing effects (Chan & McDermott, 2007). Moreover, the benefits of errorless learning among memory-impaired populations often arise in free/cued-recall but not recognition tasks (Clare & Jones, 2008), suggesting that recognition may not be not sensitive enough to discriminate the memorial effects of these learning styles.

Second, Grimaldi and Karpicke (2012) have put forth an alternative explanation of how errors benefit memory, comparing the benefits of error generation to those of the testing effect (the finding that retrieval practice leads to better delayed memory than restudy opportunities). They suggest that when we see a cue at encoding in trial-and-error learning (e.g., frog) we activate a semantic network of preexisting associations that includes both our errors and the target (e.g., pond). Feedback then selects among the activated words the correct target for retrieval. Therefore, targets acquired via trial-and-error learning profit from covert retrieval practice whereas this is not possible in errorless learning. However, our findings do not fully support this idea: The word brick should
be activated by the cue *br___*, similarly to how rose is activated when seeing flower. If simply covert retrieval is sufficient to produce a benefit, trial-and-error learning should have been superior in both conditions.

Finally, as mentioned in the introduction, Bridger and Mecklinger (2014) used similar learning paradigms to ours, and found like we did that guessing benefited target memory when cues were words (*angel – church*) but reduced it when cues were word-stems (*bro – brother*). They suggest that this dissociation is the result of varying cue constraint, i.e., the number of response candidates activated by the cue. In the case of related word pairs, the variety of possible guesses is a function of the learners semantic network, and the cue (*angel*) imposes low constraint on the target (*church*). By contrast, their word stems highly constrained generation because they were selected to elicit only two but equally dominant responses (textitbrother or textitbroom). They argue that when cues highly constrain possible targets, it is more difficult to suppress the error at recall given that it is both self-generated and strongly activated (see also Grimaldi & Karpicke, 2012). However, we argue that their results could be better interpreted in terms of the type of processing conceptual versus nonconceptual afforded by error generation, than in terms of cue constraint. First, our word-stems in the lexical condition were not selected to elicit few responses, and were less constrained by virtue of having two (*br___*) rather than three (*bro___*) letters. To our knowledge, there are no published word completion norms for two-letter word stems. To address this in our study, we looked at the number of unique nouns generated as a first guess by participants in response to both types of cues, and found that lexical cues actually elicited a greater number of unique responses relative to conceptual cues, $F(1,124) = 22.39, F<.001$ (there were no age differences or Age X Condition interaction, $Fs <1$). Second, our recall data for prior guesses contradicts the idea that lexical guesses were more activated during retrieval: Participants were significantly less likely to recall lexical relative to conceptual guesses overall, regardless of age. Put otherwise, memory for lexical targets was not reduced at the expense of better memory for self-generated, highly activated errors. In fact, older adults are known to show more difficulty suppressing irrelevant information (Lustig & Hasher, 2001), and yet they were more likely to recall fewer, rather than more of their prior errors in the lexical condition. Although we did find that participants were more likely to intrude their guess as the target in lexical relative to conceptual cued recall, this amounted to less than half a word ($M=0.42$ and $M=0.11$ respectively).

In sum, our results are the first to suggest that relationship between errors and targets should be conceptual in nature for trial-and-error learning to enhance memory. This is the case not only for younger adults, but also older adults who are known to benefit from semantic processing at encoding. However, by which mechanism do conceptual and lexical errors act as catalysts and competitors respectively?
Errors as stepping stones

Why must errors be conceptually related to targets to incur benefits? We suggest that remembering conceptual errors can be helpful at retrieval, acting as stepping stones towards the target. This idea is similar to mediator hypotheses of testing effects (Carpenter, 2011; Pyc & Rawson, 2010), which states that information generated by the learner at encoding can support later retrieval. For instance, recalling that we guessed tulip to the cue a flower can scaffold us toward the target rose. In line with this idea, we found that when younger and older learners successfully recalled targets in the conceptual condition, they were also very likely to remember both of the guesses they generated during the learning phase, and most often in the correct temporal order. The younger adults maintained this general pattern in the lexical condition, albeit to a much lesser extent, whereas the older group displayed the reverse pattern: Retrieving a lexical target was more likely to be accompanied by a failure to recover prior guesses. In other words, it was better for older adults to have forgotten their errors in order to retrieve lexical targets. Speaking to error mediation more directly is our finding that successful retrieval of guesses better predicted target recall in the conceptual relative to the lexical condition, especially among older adults.

However, it is possible that the nature of our cued recall task encouraged the use of errors as mediators and buffered against their potential interference, especially among older adults. Requiring participants to retrieve errors along with targets may have promoted a “recall-to-reject” strategy at retrieval such that attempting to exhaustively recall the items that are not correct helped them isolate which item is correct. Relative to when only the target is queried, participants may shift from a reliance on fluency to a more diagnostic basis for memory when they must distinguish errors from targets. This would be especially beneficial for combatting age-related decreases in recollection, similar to how warning learners of lures in false memory paradigms can significantly reduce age differences in memory performance (Jacoby & Rhodes, 2006). This idea is supported by our previous work looking at conceptual error generation (Cyr & Anderson, 2012): We found that younger- but especially older- adults were better at correctly rejecting their prior guesses at recognition when the task was to identify targets that were acquired via trial-and-error relative to errorless learning. In other words, participants reinstated the trial-and-error context when searching memory for trial-and-error targets, allowing them to successfully disqualify their guesses. By contrast, searching memory for errorless targets did not orient individuals to distinguishing guesses from targets, and they were more likely to be falsely endorsed on the basis of high familiarity. As such, our cued recall task may have facilitated mediation of conceptual errors, and downplayed age differences in their benefit. However, our cued recall task should also have encouraged error mediation in the lexical condition, and this was not the case, especially among older adults who tended to recall less rather than more guesses along with lexical targets. Therefore, while our test format may have modulated the size of our error generation effects and age differences therein, it cannot account for the different memorial consequences of error recall between conceptual and lexical learning.
Taken as a whole, these findings reveal that lexical and conceptual errors are differentially dealt with retroactively at retrieval. One possibility is that conceptual errors can be integrated into a single memory trace by virtue of their preexisting association with the target, disarming them as potential competitors. This idea resonates with findings that the fan effect can be eliminated in both younger and older adults when competing pieces of information can be integrated into a single scene (Radvansky, Zacks, & Hasher, 1996). To this point, a recent study on proactive interference using a classic A-B, A-D paradigm found that when asked to recall word D, younger and older participants were more likely to answer correctly if they had detected the response change (from B to D); conversely, participants who had not reported the response change were more likely to exhibit proactive interference (Wahlheim & Jacoby, 2013). In other words, memory for the new pairing (A-D) was facilitated if the learner was reminded of the initial pairing (A-B). This proactive facilitation was replicated in a subsequent study by these authors, wherein participants indicated whether a word in a series belonged to the same category as the previously presented word (e.g., steel, iron). At test, participants were presented with unstudied semantic cues (a metal) and asked to recall the most recently presented instance of that category (iron) and whether another instance had appeared prior (steel). They found facilitation on the cued recall test when participants were reminded of the earlier instance and interference when such remindings did not occur. Interestingly, participants rarely confused the most recent and the prior instances at test (e.g., indicating that iron was seen before steel), suggesting that remindings preserve their temporal order in the memory trace. This is in line with our findings that conceptual guesses were likely to be recalled in the correct temporal order, especially among younger adults. Lexical errors, however, cannot be integrated into the memory trace and thus amount to noise without effortful semantic elaboration on the part of the learner. It is possible that the younger adults were retroactively and spontaneously relating lexical guesses and targets to some extent (e.g., “the brick is on the bridge”), explaining why they continued to recall more rather than less errors along with targets as in the conceptual condition. Known age-related declines in the ability to self-initiate these elaborative strategies (Craik & Rabinowitz, 1985; Rabinowitz et al. 1982) may explain why older adults failed to activate lexical guesses along with targets. This overall poor memory for lexical errors is consistent with our finding that older adults rarely produced their wrong guesses as intrusions during recall (less than one prior error intrusion on average) (see also Lubinsky et al., 2009; Anderson et al., 2012 for similar results).

Future studies are needed in order to ascertain whether this pattern of results is maintained after a longer delay, although there is evidence to suggest it would. With respect to conceptual learning, Kornell et al. (2009) found that the error generation benefit in memory for related word pairs was maintained in younger adults 24 hours later, and both younger and older adults show testing benefits on immediate and delayed memory tests for educational facts (Meyer & Logan, 2013). It is less clear how a delay would affect lexical learning, although studies on errorless learning of nonconceptual information among memory-impaired populations show that its memorial benefits
are typically attenuated after a delay (Ruis & Kessels, 2005; Hunkin et al., 1998; Squires et al., 1997). As such, we might expect the positive effect of conceptual errors to be maintained, and the negative effect of lexical errors to be reduced after a delay. In addition, it would be interesting to examine the effects of massed versus spaced retrieval during errorless and trial-and-error learning; it is well established that younger and older adults memory performance benefits from spaced retrieval (Balota, Duchek, & Paullin, 1989), and it is possible that the conceptual trial-and-error advantage would be exaggerated under spaced conditions. Finally, it would be worth considering error generation effects in light of pathological aging: It is unlikely that individuals suffering from mild cognitive impairment or dementia would show a conceptual error benefit due to compromised recollection-based memory (Anderson et al., 2008), which we have argued is key to resolving and integrating errors (see Cyr & Anderson, 2012).

In summary, we found evidence that conceptual guesses are usefully retrieved along with targets among both age groups. When learning emphasizes conceptual processing, error generation creates a richer memory trace through ordered relations between errors and targets, leading to better memory. By contrast, lexical errors, in addition to targets, are recalled significantly less, and are best forgotten in service of older adults memory for correct information. Our findings illustrate the words of Eli Siegel: “If a mistake is not a stepping stone, it is a mistake.”
Chapter 3

Influence of semantic proximity on error generation effects in memory

A consequence of learning information in a deep and meaningful way is that it triggers thinking of related ideas. For instance, when asked the question “What is the capital of Canada?” we do not only activate the single correct response, Ottawa, but also a wide array of concepts that are meaningfully related either semantically to the answer (e.g., other Canadian cities: Montreal, Vancouver) to personal experience (e.g., I used to think the capital was Toronto; I went ice skating on the Ottawa canal). Activation of these conceptually related – but not targeted – memories can facilitate retrieval of sought after information by highlighting pathways for the search, essentially scaffolding retrieval (Anderson & Bower, 1972). This elaborative nature of retrieval is purported to be at the heart of retrieval-based learning techniques such as the testing effect, i.e., the finding that memory for studied information is enhanced by retrieval rather than a second study opportunity (see Karpicke & Roediger, 2008). For instance, studies have shown that when attempting to retrieve target information (e.g., bread – basket), individuals generate related mediators (e.g., bread: crust, butter) and the ability to re-activate these related concepts on a later test is associated with better target memory (e.g., bread – crust – butter – basket) (Pyc & Rawson, 2010; Carpenter, 2007). In other words, recalling the concepts that sprung to mind during a previous retrieval attempt can be useful on a later memory test for targets.

The elaborative nature of retrieval is also exemplified in studies on error generation effects in episodic memory. This literature examines the memorial benefits of retrieval prior to studying novel information, in contrast to the testing literature which considers the benefits of retrieval following initial study. Error generation is typically explored by contrasting memory performance following trial-and-error learning, where participants must guess what the target is in response to a cue (e.g., bread - ?) prior to seeing the target (e.g., bread - basket), and errorless learning,
where the correct cue-target pair is studied in full from the onset (e.g., bread - basket). Numerous studies have found improved target memory following trial-and-error relative to errorless learning for word pairs (Kornell, Hays, & Bjork, 2009; Grimaldi & Karpicke, 2012; Huesler & Metcalfe, 2012; Vaughn & Rawson, 2012) and more ecological materials such as facts (Pashler, Rohrer, Cepeda, & Carpenter, 2007; Kang, Pashler, Cepeda, Rohrer, Carpenter, & Mozer, 2011; Richland, Kornell, & Kao, 2009; Pressley, Tanenbaum, McDaniel, & Wood, 1991; Kornell, 2014). This enhancement has been attributed to the idea that retrieving errors enriches encoding by forging semantic connections and orienting learners to the relationship between the cue and the target (Cyr & Anderson, 2012; Huesler & Metcalfe, 2012; Kornell et al., 2009).

This proposal that errors support target memory by means of semantic elaboration is supported by two main findings. First, studies find no error generation benefits when targets do not share a preexisting semantic relationship with the cue (e.g., bread – cloud) (Huesler & Metcalfe, 2012; Grimaldi & Karpicke, 2012; Knight, Ball, Brewer, DeWitt, & Marsh, 2012, but see Potts & Shanks, 2013). Second, error generation has been found to impair memory relative to errorless learning when cues do not invoke semantic processing, such as guessing on the basis of a perceptual feature (e.g., a word that begins with ‘br’) (Cyr & Anderson, in press; Bridger & Mecklinger, 2014). In sum, it is increasingly clear that semantic processing that links cues, targets, and errors is required in order to observe error generation benefits. Our recent work extends this advantage of semantic error generation to healthy older adults (Cyr & Anderson, 2012; Cyr & Anderson, in press), suggesting that errors can afford elaboration akin to other encoding strategies known to minimize age-related declines in episodic memory (Luo, Hendriks, & Craik, 2007).

One question that has not been addressed is whether it matters how closely related the generated errors and targets are semantically. Do errors that are out in left field still boost memory for the target? If an error supports memory when it can be meaningfully integrated within the cue-target memory trace, effectively becoming one component of a multipart representation (e.g., bread ↔ crust ↔ basket) (see Chan, 2009 for a similar proposal applied to testing effects), then how closely the error and target are related should matter. Errors should benefit memory to the extent that they overlap with targets in terms of cue-relevant semantic features. In this research, we manipulated two aspects of memoranda which we hypothesized would modulate error integration: cue constraint and semantic proximity between targets and errors. In a recent study Bridger and Mecklinger (2014) found that error generation leads to greater memory for weakly constraining semantic items (e.g., angel: church, wings) and poorer memory for highly constraining lexical items (e.g., bro: broom, brother) items. However, the paradigm in their study confounded high/low constraint with lexical/conceptual processing, rendering it impossible to untangle their respective contributions to the interaction. If cue constraint plays a role as Bridger and Mecklinger argue, then we should observe its influence within conceptual error generation effects. In Study 2A, we manipulated cue constraint by having younger and older participants learn the endings
to sentences that were either high-cloze (e.g., You cant buy anything for a ____ (nickel)) or low-cloze (e.g., There is something grand about the ____ (hotel)) under trial-and-error and errorless learning instructions. High-cloze sentences are constraining and participants have a relatively small pool of highly interrelated candidates to choose from (e.g., penny, dollar, dime), resulting in a strong relationship between errors and targets. Low-cloze sentences, being less constrained, invite retrieval from a larger and more varied pool of candidates (e.g., duchess, opera, scenery), resulting in a weaker relationship between errors and targets. We predicted that the error generation benefit would be greater with high relative to low constraint cues, because high-cloze errors (e.g., penny and nickel) will be easier to integrate relative to low-cloze errors (e.g., duchess and hotel). In Study 2B, we explored how altering the semantic proximity of guesses to targets in trial-and-error learning would affect memory for targets and prior errors among younger and older adults. For cues we used homographs (words associated with more than one meaning, e.g., band: a music group or a binding object). In the trial-and-error condition, participants generated a guess (e.g., concert) and were shown a target that was always related to the cue (e.g., band), but was either related (e.g., guitar) or unrelated (e.g., elastic) to the meaning of the generated error. Relative to errorless learning, we predicted that errors would afford greater increases in target memory when they matched the semantic meaning of the target (e.g., concert and guitar) on account of greater overlap and semantic integration relative to mismatched errors (e.g., concert and elastic). We also predicted that participants would be more likely to retrieve their prior error along with the target if they were matched relative to mismatched, reflecting the fact that their memory traces overlap and are thus likely to be co-activated at retrieval. This would be consistent with previous findings that successful target memory is mediated by the ability to remember ones errors (Knight et al., 2012), provided that errors and targets share a conceptual relationship (Cyr & Anderson, in press). Therefore, greater memory for semantically matched relative to mismatched targets may be reflected in better memory for prior wrong guesses.

Finally, both studies also included a group of healthy older adults as we were interested in aging as a potential modifier of these effects. Although older adults can benefit from conceptual error generation (Cyr & Anderson, 2012; Cyr & Anderson, in press), there is evidence that the magnitude of this effect may be smaller relative to younger adults (Cyr & Anderson, in press). It is possible that age-related increases in semantic memory (Park et al., 2002; Taylor & Burke, 2002) may partly account for this difference, such that aging affects the ability to integrate cues, errors and targets via semantic networks.

3.1 Study 2A

The purpose of Study 2A was to examine the effects of cue constraint on conceptual error generation effects. Healthy younger and older participants studied a list of either high or low-
cloze sentence stems, and target word endings were provided under errorless and trial-and-error instructions. Based on our previous work (Cyr & Anderson, 2012; in press), we predicted that target memory would be better following trial-and-error relative to errorless learning, and that improvements would be greater among younger relative to older adults. Moreover, we hypothesized that participants in the high relative to low-cloze condition would show a larger error generation benefit, suggesting that strongly constraining cues bring errors and targets closer together in service of target memory.

3.1.1 Method

Participants

Fifty six younger and 56 older adults volunteered to participate in the study. Individuals had to be free of major health conditions affecting cognition to be included in the study, (e.g., stroke, dementia, uncontrolled metabolic dysfunction). We also administered the Modified Telephone Interview Cognitive Screening (Welsh, Breitner, & Magruder-Habib, 1993) to screen for cognitive impairment, and participants had to score above the cutoff of 30/40 to be included. Participants were recruited from volunteer research participant pools or responded to advertisements posted in the community and were paid for their participation. Twenty eight participants within each age group were assigned to the high-cloze condition, and the other 28 to the low-cloze condition. One younger participant was excluded because their electronic data file was corrupted. The mean age was 72.93(SD=5.39) for the older adults and 22.62(SD=3.29) for the younger adults. The older group had completed a higher number of years of formal education relative to the younger group, 16.39(SD=2.81) and 15.45(SD=1.72) respectively, t(109) = 2.12, p = 0.04.\(^1\) Number of years of education did not vary as a function of condition, F<1, and there was no Age X Condition interaction, F<1.

Materials

Undetermined sentences were selected from the Bloom and Fischler (1980) sentence completion norms, which had been normed in 2005 among 97 healthy older adults over the age of 65 residing in Toronto. From this initial pool of 329 sentence stems, 54 were selected for each of the high and low-cloze conditions (i.e., 108 sentence stems in total), and divided into two sets of 28: one for errorless learning (EL) and one for trial-and-error learning (TEL). A mean number of 12.59(3.22) normed unique endings were associated with the sentence stems in the high-cloze set, significantly less than 29.13(7.47) in the low-cloze set, F(1,106) = 223.04, p<.001. For each sentence stem (e.g., You can't buy anything for a _____), three single-word endings were selected from the norms as possible targets (e.g., 1: penny; 2: dime; 3: dollar). Word frequency (Brysbaert & New, 2009; database

\(^1\)Conducting the analyses that follow with years of education as a covariate did not change any of the results.
available at subtlexus.lexique.org), word length, and the number of individuals in the normative sample who generated the word as a response did not differ as a function of condition (high vs low-cloze), sets or word positions within the sets (all \( p \) values < .05). As expected, the number of unique endings generated to sentences in the normative sample, was significantly lower for word endings associated with sentence stems in the high relative to the low-cloze condition, \( F(1,323) = 21.43, p < .001 \). The assignment of which word ending would be selected as an intended target (1, 2, or 3) was counterbalanced across participants.

Design and Procedure

This was a mixed study design with age and cloze condition as between-subjects variables, and learning instructions as the within-subjects variable. Participants were assigned either to the high-cloze or low-cloze condition and tested individually. Stimuli were presented and participants responses were recorded by E-prime software (Psychology Software Tools, Inc., Pittsburgh, Pennsylvania).

Each individual underwent two study-test cycles of blocked EL and TEL learning, each consisting of 27 different sentence stems. Order of study set (1 and 2) and learning instruction (EL and TEL) were counterbalanced across participants. In EL, the experimenter presented the sentence stem on the screen (e.g., You cant buy anything for a ____) and then immediately displayed the target in the sentence context (e.g., You cant buy anything for a dime). In TEL, participants were shown a sentence stem (e.g., New York is a very busy ____) and prompted to guess the target word (e.g., “Is it town?”) after which they were given feedback by the experimenter (e.g., “No, good guess, try again!”). Participants were required to generate two guesses, after which the experimenter provided them the correct target in the sentence context (New York is a very busy city). For both EL and TEL, participants were instructed to commit to memory the correct target word ending to each sentence for a later memory test. The experimenter always selected a target word ending that was not generated as a guess by the participant. This was done by assigning each word ending of a sentence stem to numbers one through three in E-Prime: If the participant guessed the word ending in position number one, and one was their assigned target word number (counterbalanced across participants), the experimenter pressed on key number two to display a word ending that had not been generated. As a result, all participants generated exactly two errors at study. All targets were shown within the sentence for three seconds and words were followed by a one second inter-stimulus interval. A 20-minute break followed the study phase during which participants played Tetris on the computer. Once the break had expired, participants began the cued recall task. Following the study under EL and TEL instructions, participants were shown the studied sentence stems and asked to recall and type in the correct word ending for each. There were no new sentence stems, and the cued recall task was self-paced. Following both study-test cycles, participants were debriefed and compensated. The entire session lasted approximately 75 minutes. This study was approved
by the University of Toronto and Baycrest Research Ethics boards.

### 3.1.2 Results and Discussion

An alpha level of .05 was used throughout.

We first conducted a 2(Learning Style: EL, TEL) X 2(Condition: High-cloze, Low-cloze) X 2(Age) mixed ANOVA. This yielded a significant main effect of Learning Style, $F(1,107) = 124.63$, $p < .001$, $\eta^2 = .54$, and a significant Learning Style X Age interaction, $F(1,107) = 24.14$, $p < .001$, $\eta^2 = .18$, revealing that memory performance was greater following TEL relative to EL, and that this improvement was greater among younger relative to older adults (see Figure 3.1). There was also a significant main effect of Condition which showed that memory performance was better in the High-cloze relative to the Low-cloze condition, $F(1,107) = 7.44$, $p < .01$, $\eta^2 = .07$. The Learning Style X Condition interaction was not significant, $F(1,107) = 1.35$, $p = .25$, nor was the 3-way interaction, $F < 1$. There was also no main effect of Age or an Age X Condition interaction, $F < 1$.

![Figure 3.1: Memory for target word endings as a function of age and condition in Study 2A (bars represent ± 1 standard error).](image)

We then examined whether learners produced their prior guesses as intrusions when target recall was unsuccessful following trial-and-error learning, which would suggest that they were confusing self-generated errors as the right answers. A 2(Age) X 2(Condition) ANOVA revealed that older adults made more prior error intrusions than younger adults, $F(1,107) = 12.41$, $p < .01$, $\eta^2 = .10$, although these were committed infrequently overall (Older: $M = 2.60$; Younger: $M = 1.80$). There was no effect of Condition and no Age X Condition interaction, $F < 1$.

The results of Study 2A showed a benefit of trial-and-error over errorless learning among both age groups, consistent with previous studies using conceptual learning paradigms (Cyr & Anderson, 2012; Cyr & Anderson, in press). Moreover, the magnitude of the error generation benefit in cued recall performance was greater among younger relative to older adults, consistent with our prior
Chapter 3. Influence of semantic proximity on error generation effects in memory

work (Study 1 of this dissertation). Both age groups performed better overall in the high relative to low-cloze condition, which we interpret in light of the facilitating effects of highly constraining cues.

However, we found no evidence that the error generation effect is moderated by cue constraint in either age group, as errors based on both high and low-cloze sentence stems yielded equivalent memory benefits. In other words, it does not appear to matter whether the cue activates both errors and targets strongly as in high-cloze sentences (e.g., He caught the ball with his ____ target: hand, error: glove) or more weakly as in low-cloze sentences (e.g., When he grows up, Johnny wants to be a ____ target: singer, error: father). A first possibility is that cue constraint influences both errorless and trial-and-error learning. Indeed, Figure 3.1 shows that target memory improved from low to high-cloze under both errorless and trial-and-error learning. Younger adults also may have hit a performance ceiling in trial-and-error learning, obscuring an interaction. A second possibility, however, is that cue constraint per se is not a critical determinant of the error generation benefit, provided that errors are meaningful. In the design of the current experiment, despite being weakly constrained by the cue When he grows up, Johnny wants to be a ____, the error father is easy to integrate into the cue-target memory trace given the amount of contextual support. Thus, perhaps high and low-cloze errors did not sufficiently differ in terms of how easily they could be meaningfully related to cues and targets.

3.2 Study 2B

In Study 2B, we decided to manipulate the strength of the error-target relationship more directly by using homograph word cues (e.g., organ) which allows for error generation from two separate taxonomies (e.g., a structure of the body or a musical instrument). In the trial-and-error condition, younger and older participants were shown a target that either matched or mismatched the meaning of their guess. We predicted that the error generation effect would be amplified when participants error meanings were matched, consistent with our notion that conceptual errors are beneficial as a function of the ease of integration at encoding. We also asked participants to report their prior error at cued recall to examine whether matched errors and targets are more likely to be recalled together relative to mismatched errors and targets.

3.2.1 Method

Participants

Thirty two older and 32 younger adults agreed to participate in this experiment. The data of one older adult had to be discarded because of a programming error. As in Study 2A, all were initially screened for major health conditions affecting cognition (e.g., stroke, dementia, uncontrolled
metabolic dysfunction) and screened using the Telephone Interview Cognitive Screening (Welsh et al., 1993). Participants were recruited from volunteer research participant pools or responded to advertisements posted in the community and were paid for their participation. The mean age was 72.06 (SD=5.27) for the older adults and 23.66 (SD=3.33) for the younger adults. The older and younger groups did not differ in years of formal education, 16.48 (SD=2.47) and 15.34 (SD=2.66) respectively, t(61) = 1.76, p = 0.08. There were no age differences in scores on the Mini-Mental Status Examination (MMSE; Folstein, Folstein, & McHugh, 1975), t(61) = 1.13, p = 0.26. Consistent with prior research comparing younger and older adults, vocabulary scores on the Shipley Institute of Living Scale (Shipley, 1940) were higher in the older than younger group, t(61) = 6.19, p < 0.001. Anxiety and depression symptomatology as measured by the Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983) revealed significantly higher anxiety scores among younger relative to older adults, t(61) = 2.41, p = 0.02, and no age differences in depression scores, t(61) = 1.79, p = 0.09, respectively. However, scores from neither groups neared the clinical cutoff score of 13.

Materials

Sixty four homographs were selected from the Alberta homograph meaning frequency norms (Twilley, Dixon, Taylor, & Clark, 1996) based on the following criteria: a) the homograph was a noun; b) the homograph was predominantly associated with two meanings only; and c) the homographs primary and secondary meaning were nouns. For each homograph (e.g., port), we then created two associates as possible targets for each primary (1: boat; 2: dock) and secondary (1: wine; 2: brandy) meaning. This pool of 64 homographs was then divided into four sets of 16 pairs. Word frequency and word length did not differ as a function of set or word position with the sets, Fs<1. There were also no differences across sets in terms of the proportion of normed responses to the homograph for the primary and secondary meanings, F(1,60) = 1.64, p = 0.19.

Next, a latin-square was used to assign two sets to each of the EL and TEL conditions. For the EL sets, we counterbalanced whether homographs in the first and second half of each set were assigned a primary or secondary meaning target (e.g., boat or wine). We then randomized the order of the word pairs. Therefore, under EL each participant studied 16 homographs paired with a primary meaning target, and 16 homographs paired with a secondary meaning target. For the TEL sets, we counterbalanced whether homographs in the first and second half of each set were assigned to be Match or Mismatch trials, and randomized the order of the word pairs. As such, in the TEL condition participants studied 16 homographs paired with a target that matched their guess (Match condition) and 16 homographs paired with a target that did not match their guess (Mismatch condition). The assignment of which word would be selected as an intended target (one or two) was counterbalanced across participants.

2Rerunning the analyses that follow with HADS anxiety score as a covariate did not change results.
Design and Procedure

This was a mixed study design with age as the between-subjects variable, and learning instructions and semantic proximity as the within-subjects variable. Participants were tested individually, and stimuli presentation and recording was done with E-prime software (Psychology Software Tools, Inc., Pittsburgh, Pennsylvania).

Each individual underwent 2 study-test cycles of blocked EL and TEL learning, each consisting of 32 different homograph word pairs. Order of learning instruction (EL and TEL) was counterbalanced across participants. In EL, the experimenter presented the homograph on the screen (e.g., calf) and subsequently displayed the target (e.g., ankle). In TEL, participants were shown a homograph (e.g., pitcher) and were prompted to guess at the target word once (“Is it baseball?”). If it was a Match trial, the experimenter would provide them with a target from the same meaning as their guess (“No, the target word is catcher.”). If it was a Mismatch trial, the experimenter would provide them with a target from the other meaning (“No, the target word is lemonade.”). In other words, if they generated a word related to the primary meaning of the homograph, the target would be another word related to the primary meaning in a Match trial, and a word related to the secondary meaning in a Mismatch trial. For both EL and TEL, participants were instructed to commit to memory the correct target words in each category for later. In TEL, the experimenter always selected a target word that was not generated as a guess by the participant. This was done by assigning each word from the primary or secondary meaning to numbers one or two in E-Prime: if the participant guessed the word in position number one, and one was their assigned target word number (counterbalanced across participants), the experimenter pressed on key number two to display a word that had not been generated. All targets were shown with the homograph (e.g., calf – ankle) for four seconds and words were followed by a one second inter-stimulus interval. Because this task was more difficult compared to Study 2A, participants were asked to write down the targets as they were shown, making sure to cover their responses with another sheet of paper as they went along to discourage rehearsal. We also shortened that delay relative to Study 2A: A 10-minute break followed the study phase during which the HADS or the MMSE was administered. Once the break had expired, participants began the cued recall task. Following the study under EL instructions, participants were shown the studied homographs and asked to recall and type in the correct target. In the TEL condition, they were shown the homographs and required to recall the target, followed by the guess they had generated at study. There were no new homographs, and the cued recall task was self-paced. Following both study-test cycles, participants were debriefed and compensated. The entire study lasted approximately 60 minutes, and was approved by the University of Toronto and Baycrest Research Ethics boards.
3.2.2 Results and Discussion

An alpha level of 0.05 was used for all statistical tests.

Participant generated errors

We first examined whether the meaning of participant generated errors (primary or secondary) varied as a function of age and trial-and-error condition. Both younger and older adults were more likely to generate a guess that was related to the primary relative to secondary meaning of the homograph, $t(63) = 20.63, p < .001$, and, $t(63) = 22.01, p < .001$, respectively. However, the proportion of generated guesses that were related to the primary relative to the secondary meaning of the homograph did not vary as a function of trial-and-error condition or age, $F$s $< 1$, and there were no age differences in terms of the number of unique responses provided in response to homographs, $t(126) = 0.33, p = .74$.

Cued recall performance: Targets

We first ran a 2(Age) X 3(Condition: EL, TEL Match, TEL Mismatch) mixed ANOVA. This yielded a significant main effect of Condition, $F(1,61) = 31.61, p < .001, \eta^2 = 0.34$, revealing that memory was greater following both Match and Mismatch TEL relative to EL, as evinced by non-overlapping confidence intervals, and greater following Match TEL compared to Mismatch TEL and EL. The Age X Condition interaction was also significant, $F(1,61) = 5.68, p < .01, \eta^2 = 0.09$, which showed that younger adults displayed this pattern more strongly (see Figure 3.2). Breaking down the interaction, we find that younger adults benefit more from Match TEL relative to Mismatch TEL, $t(31) = 4.39, p < 0.001$, whereas older adults do not, $t(30) = 1.14, p = 0.26$. There was no main effect of Age, $F(1,61) = 3.15, p = .08$. 
Next, we conducted a 2(Age) X 2(Meaning: Primary, Secondary) mixed ANOVA because we were interested in memory for EL targets as a function of whether they were associated with the primary or secondary meaning. This showed that primary targets were better remembered than secondary targets, $F(1,61) = 17.61, p < .001, \eta^2 = 0.22$, but that meaning did not interact with age, $F < 1$. There were also no age differences, $F < 1$.

**Cued recall performance: Errors**

Next we examined memory for prior errors at cued recall by conducting a 2(Age) X 2(Condition: Match, Mismatch) mixed ANOVA. There was no main effect of Condition or Age X Condition interactions, $F$s $< 1$, but a significant main effect of age revealing that younger adults remembered more of their prior guesses compared to their older peers, $F(1,61) = 12.85, p = .001, \eta^2 = 0.17$ (see Figure 3.3). We also wanted to investigate whether TEL targets that were successfully recalled were accompanied by 0, or 1 guess. Therefore, we conducted a 2(Age) X 2(Condition: Match, Mismatch) mixed ANOVA that only included data from TEL trials where the target was successfully recalled (we could not run a similar analysis with unsuccessful trials on account of some participants having perfect recall following TEL). This revealed no main effect of Condition, $F < 1$, or Age, $F(1,60) = 2.12, p = .15$, and no interaction, $F(1,60) = 1.30, p = .26$. 
Chapter 3. Influence of semantic proximity on error generation effects in memory

Figure 3.3: Memory for prior errors as a function of age and trial-and-error condition.

The results of Study 2B show a clear enhancement of the error generation benefit in the younger group when errors and targets belong to similar relative to disparate semantic families. This is consistent with our hypothesis that proximal (e.g., band: concert – guitar) relative to distant (e.g., band: elastic – guitar) errors are better integrated in service of episodic memory. However, this integration was not reflected in memory for prior errors: matched errors were not better remembered along with targets than mismatched errors. In a previous study (Study 1 of this dissertation), we had found that targets and errors were likely to be retrieved together when the error generation benefit was present. It is possible that younger adults prior error recall was at ceiling across trial-and-error conditions, making it impossible to unveil any effect of semantic proximity on scaffolding target retrieval.

However, this dissociation in target memory as a function of semantic proximity was not found among older adults, who showed an error generation benefit of equal magnitude for matched and mismatched targets. This may be explained by proposals that semantic networks become more flexible and overlapping with age and experience, leading to more widespread activation (Taylor & Burke, 2002). As such, integrating semantically disparate errors and targets may have been easier for older relative to younger adults.

3.3 General discussion

It is increasingly clear that both younger and older adults can benefit from their mistakes, provided that they occur in a meaningful context. The goals of the current study were to explore the boundary conditions of this conceptual error generation benefit in memory, namely the effects of cue constraint and semantic proximity. The findings of Study 2A tentatively suggest that cue constraint is not an important determinant of conceptual error generation, as errors produced in
both high and low constraint contexts afforded memory benefits in both age groups (although ceiling effects may have obscured an interaction in the younger group). This is consistent with our recent work (Study 1 of this dissertation) among younger and older participants where we found an error generation benefit in memory for semantically-based errors and an impairment for lexically-based errors, despite the fact that the semantic cues (e.g., a farm animal: cow, pig, horse), constrained generation to a smaller array of unique exemplars compared to the lexical cues (e.g., br_____: brick, bridge, bride). We argue that the memorial benefits of conceptual error generation vary more as a function of semantic relatedness than cue constraint, an idea that was supported by the results of Study 2B. Both age groups showed an error generation benefit for target memory when errors and targets semantically matched (e.g., organ: piano – church), consistent with proposals that semantic processing is the locus of error benefits among younger (Kornell et al., 2009; Huesler & Metcalfe, 2012; Bridger & Mecklinger, 2014) and older adults (Cyr & Anderson, 2012; Cyr & Anderson, in press). The fact that we also found an error benefit when errors and targets were semantically mismatched (e.g., organ: liver-church) is at first surprising in light of previous studies showing a lack of error generation benefits for unrelated cues-target word pairs (e.g., frog – bottle) (Huesler & Metcalfe, 2012; Grimaldi & Karpicke, 2012). However, in these cases, participants are given a target that is related neither to the cue nor their error, making integration difficult. In our mismatch condition, the target church may not be related to the error liver, but the cue organ can still effectively bind them all together: in other words, the intact cue-target semantic relationship was sufficient to salvage the error generation benefit. Among younger adults, however, this benefit on mismatched trials was nevertheless reduced relative to match trials where the error-target relationship could be additionally exploited in service of memory (e.g., piano – church). This is consistent with our hypothesis that errors buttress retrieval as a function of their semantic overlap with targets. However, this scaffolding was not reflected in memory for errors: Matched errors were just as likely as mismatched errors to be retrieved along with targets, possibly due to ceiling effects in memory for prior errors.

Error generation benefits among older adults, however, were not mediated by semantic proximity. Although more research is needed, this finding could be understood in light of the fact that because older adults have more experience with language, their semantic networks are more highly interconnected, facilitating the spread of activation (Taylor & Burke, 2002). This would explain findings of greater semantic priming during word recognition (see Laver & Burke, 1993). Support for this hypothesis comes from a study by Taylor and Burke (2002), where younger and older adults had to name items represented in a picture as quickly as possible (e.g., say ‘ball’ when shown a picture of a ball). Importantly, the word representations of the picture items were homographs (e.g., ball can refer to a toy or a lavish gathering). Before each picture was presented, a written word was flashed that was either related and appropriate (e.g., toy), related and inappropriate (e.g., prom) or unrelated (e.g., duck) to the picture. They then measured response time to picture naming as
a function of the type of distractor. For both younger and older adults, related and appropriate
distractors slowed picture naming relative to unrelated distractors, suggesting that they interfered
with retrieval of the picture words. In other words, the ‘toy’ node competes with the ‘ball’ node as
a result of semantic connections that link them. However, the related but inappropriate distractors
(e.g., prom) only slowed naming relative to unrelated distractors among the older adult sample.
This suggests that for older adults, activation is more widely distributed in semantic networks rel-
ative to younger adults, such that naming ‘ball’ is affected by the activation of ‘prom’ via semantic
connections. As such, a homograph (e.g., organ) may have triggered both meaning nodes (e.g., a
bodily structure and a musical instrument) among older adults due to widespread activation in ag-
ing semantic networks. This would result in greater semantic overlap between mismatched targets
and errors relative to younger adults, and a lessened influence of semantic proximity.

In summary, these studies add to the growing literature on error generation effects by elucidating
the factors that mitigate its effectiveness, and extending its benefits to healthy aging. This work
has clear implications for educational practice given that questions can vary enormously in terms
of how they orient retrieval. For example, it may be better to ask questions that guide learners
to guess in the right ballpark (e.g., Question: What kind of living thing is an earwig?; Answers:
reptile, bird, insect) as opposed to left field (e.g. Question: What is an earwig?; Answers: earring,
hairpiece, insect). Viewing error generation in light of healthy aging is encouraging for late-life
learning, and provides a model with which to elucidate the cognitive factors that underpin error
resolution.
Chapter 4

Influence of confidence on error generation effects in memory

We routinely make errors when learning information in our day to day lives, leading to memory errors that influence our knowledge of the world. Although many errors are harmless, others can have detrimental effects, such as lowering a score on an exam or altering our understanding of important concepts. Finding ways to revise knowledge through error correction is an important goal for educators and for all of us in service of our everyday memory.

When providing feedback about a mistake after a quiz, it is common to hear reactions that vary from “I’ll never remember that answer” to “Aha! I can’t forget that one!” suggesting that feedback resonates differently with learners depending on internal factors. While it is intuitive to assume that errors endorsed with higher confidence would be strengthened and therefore more difficult to correct, the opposite has been shown for younger adults (Butterfield & Metcalfe, 2001; Butterfield & Mangels, 2003, 2006; Fazio & Marsh, 2009; Finn & Metcalfe, 2010). This finding that deeply entrenched errors are more easily corrected than are loosely held errors is termed the hypercorrection effect (Butterfield & Metcalfe, 2001). Butterfield and Metcalfe (2001) asked participants to generate answers to general knowledge questions and to rate their confidence in the correctness of their answers, after which they were shown the correct answer. When participants were later asked the same questions, they were more likely to respond correctly to the questions that had produced higher- rather than lower-confidence errors.

To account for this counterintuitive finding, many have suggested that feedback following higher confidence errors captures attention more readily (Butterfield & Mangels, 2003; Butterfield & Metcalfe, 2001, 2006; Fazio & Marsh, 2009): participants are surprised by the answer and rally their resources to correct their misconception. Others have suggested that the subject matter is more familiar for higher – relative to lower – confidence errors (e.g., the capital of Australia is Canberra.
versus the capital of Congo is Kinshasa, respectively), making it easier to encode the correct answer (Butterfield & Metcalfe, 2006; Finn & Metcalfe, 2011; Sitzman et al., 2014).

The effects of aging on hypercorrection have never been examined, but it is unclear whether older adults can use feedback to overcome errors as effectively as do their younger counterparts. Research stemming from the neuropsychological literature suggests that unlike younger adults, older adults may not recover from errors with the same facility, even in the presence of feedback (Lubinsky, Anderson, & Rich, 2010; Guild & Anderson, 2012; Anderson & Craik, 2006; Baddeley & Wilson, 1994; for an exception see Cyr & Anderson, 2012). Older adults may not harness feedback effectively and may fail to show hypercorrection. Our primary goal in the studies reported here was to determine whether or not older adults can hypercorrect their errors as well as do younger adults.

The standard paradigm requires participants to freely recall an answer to every item, even if they have to guess. Two features of this paradigm may dampen older adults ability to hypercorrect their errors. First, tip-of-tongue (TOT) states, defined as the temporary inability to retrieve a word despite a persistent feeling of knowing it (Brown & McNeill, 1966), are the most commonly reported memory failure with age (Sunderland, Watts, Baddeley, & Harris, 1986) and have been shown in the laboratory (Brown, 1991; Brown & Nix, 1996;). Furthermore, TOT states are more common for proper nouns (Brown, 1991) which constitute most of the answers to general knowledge questions (e.g., names of persons and places). Age-related increases in TOT states would cause older adults to be forced to generate more answers that they know are incorrect, and which they would rate with lower confidence. Second, it has been proposed that older adults show effective monitoring of well-learned information but deficient monitoring of recently learned information (Dodson, Bawa, & Krueger, 2007), both of which are encountered in a hypercorrection paradigm (i.e., general knowledge is retrieved but also acquired). Indeed, older adults show a weaker relationship between confidence and accuracy in tasks of cued recall (Kelley & Sahakyan, 2003; Dodson et al., 2007), source identification (Dodson et al., 2007) and eyewitness memory (Dodson & Krueger, 2006) for information shown earlier in an experimental session, whereas they possess good calibration where general knowledge is concerned (Dahl, Allwood, & Hagberg, 2009; Pliske & Muter, 1996; Dodson et al., 2007). These two issues suggest that using free recall and requiring explicit confidence ratings may obscure age-related differences in the relationship between metacognition and error correction.

We report two studies. In Study 3A, we sought to replicate the finding that errors made with higher rather than lower confidence are more easily corrected among young adults and to determine whether the same effect is shown by older adults. If confidence modulates the efficacy of feedback processing for younger and older adults alike, we would expect to see equivalent hypercorrection effects between age groups. In Study 3B, participants answered multiple-choice questions and indicated how many options they had ruled out prior to making their selection. The approach taken in the second study avoids free recall and explicit confidence ratings, and is arguably more aligned
with how general knowledge is tested in educational settings, where multiple-choice test formats are increasingly utilized due to large class sizes and ease of grading. Older adults also encounter multiple-choice testing in pursuit of continued education and in routine driving tests. Students typically cross out alternatives that they know are incorrect and narrow down their options till they reach a satisfactory answer. It is a reasonable assumption that students are more confident in their answers to questions that they can narrow the choice to fewer alternatives (e.g., Its A or B) compared to questions for which they are unable to rule out any options; if this is true, then the former should lead to greater error correction.

4.1 Study 3A

In Study 3A, younger and older participants answered general knowledge questions and rated their confidence after which they viewed corrective feedback. Following a brief delay, the same questions were administered and participants answered them again.

4.1.1 Method

Participants

Twenty-two younger and 22 older adults volunteered to participate; however, the data of one younger and four older adults was not included in analyses because they made no errors on the initial test. Therefore, the data from 21 younger and 18 older adults is reported. Participants were initially screened for major health conditions affecting cognition (e.g., stroke, dementia, uncontrolled metabolic dysfunction). Participants were recruited from volunteer research participant pools or responded to flyers posted at the University of Toronto and at a Toronto retirement community and were paid for their participation. The mean age was 69.64 (SD=4.21) for the older adults and 20.86 (SD=2.46) for the younger adults. The older and younger groups did not differ in years of formal education, 15.71 (SD=1.57) and 15.09 (SD=0.70) respectively, t(37) = 1.60, p=.12, and younger adults scored higher on the Mini-Mental Status Examination relative to older adults (MMSE; Folstein, Folstein, & McHugh, 1975), 29.81 (SD=.51) and 28.47 (SD=1.70), respectively, t(37) = 3.43, p<.01. Consistent with prior aging research, vocabulary scores on the Shipley Institute of Living Scale (Shipley, 1940) were higher in the older than younger group, 35.71 (SD=4.24) and 31.52 (SD=4.70), respectively, t(37) = 2.19, p = .04. Endorsements of anxiety symptoms from the Hospital Anxiety and Depression Scale (HADS: Zigmond & Snaith, 1983) did not differ between younger 6.52 (SD=2.80) and older adults 7.86 (SD=3.62), t(37) = 1.01, p = .32; older adults endorsed more symptoms of depression than did younger adults, 6.57 (SD=2.80) versus 2.19 (SD=1.50), respectively, t(37) = 4.14, p<.001, but no participant exceeded the cut-off of 8 for mild depression.
Materials

150 general knowledge questions were selected from various sources including the Nelson and Narens (1980) norms and online trivia games. The questions covered a broad range of topics such as history, social and natural sciences, pop culture, and varied in difficulty. The answers to each question were verified by other sources (e.g., online encyclopedias). We piloted among both age groups to ensure that the questions yielded a sufficient number of errors and a distribution of responses on the confidence scale.

Procedure

Following consent, participants were administered the MMSE individually. Participants were then tested in groups of 7 to 15 individuals of the same age group (piloting ensured that this did not bias responding). Following a practice phase, the questions were presented using Microsoft PowerPoint and participants responded to each question individually in a provided booklet. They also rated how confident they were in their answer to each question on a scale of 1 through 7, where 1 indicated ‘Sure Im wrong’, 4 indicated ‘Unsure’, and 7 indicated ‘Sure Im right’. Once everyone had answered, the correct answer was shown for 2 seconds and the next question followed. Participants were told to guess if they did not know the answer. Once all questions were answered, there was a 15-minute break and participants completed the Shipley scale and the HADS. Afterwards, participants were given a test booklet containing all 150 questions in the same order presented earlier. Like the study phase, the task was self-paced and guessing was encouraged when memory failed. Unlike the study phase, they did not rate confidence or receive feedback. Finally, participants were debriefed and compensated.

4.1.2 Results and Discussion

A significance level of .05 was applied throughout.

All answers were scored by hand by a research assistant and when an ambiguity was identified, the principal investigator was consulted. Spelling errors did not affect scoring. All results reported were unchanged when the HADS depression score was included as a covariate, with the exception of the age difference in gamma correlations noted below.

Initial and final test performance

To ensure that confidence and accuracy were well calibrated, we computed the gamma correlation between initial confidence ratings and initial recall performance. Computed on each participant and then averaged over participants, the correlation was .80 (SE=.03) and .75 (SE=.03) for younger and older adults respectively and both were significantly different than zero (younger: $t(20) =$
30.38, p<.001; older: t(17) = 26.15, p<.001). To determine whether both groups used the scale similarly, a repeated-measures ANOVA was conducted on the number of errors at each confidence level. As expected, this revealed a significant main effect of confidence, $F(1,37) = 26.15, p<.001, \eta^2_p = .45$. Importantly, there was no main effect of age, $F(1,37) = 1.88, p = .18$, and no interaction, $F(1,37) = 1.47, p = .24$.

We conducted a repeated-measures ANOVA to determine whether younger and older adults showed the same rate of improvement from initial to final test. The Age X Test Delay interaction was significant, $F(1,37) = 7.42, p = .01, \eta^2_p = .17$, reflecting a greater increase in performance from initial to final test among younger relative to older adults (see Figure 4.1). This is congruent with research showing age-related declines in the ability to override errors (Baddeley & Wilson, 1994; Anderson & Craik, 2006).

![Figure 4.1: Mean accuracy on initial and final free recall as a function of age in Study 3A (bars represent ± 1 standard error).](image)

**The hypercorrection effect**

Hypercorrection reflects that higher-confidence errors are more likely to be corrected than lower-confidence errors. We first report an Age X Confidence Level repeated-measures ANOVA on final test accuracy. Final test accuracy was higher in younger than older adults, $F(1,37) = 6.16, p = .02, \eta^2_p = .21$, and increased with confidence in errors on the initial test, $F(1,37) = 6.15, p<.001, \eta^2_p = .20$, but these effects did not interact, $F(1,37)<1$ (see Figure 4.2).
However, given that not all participants made errors at every confidence level, within-subject measures of association should be considered more trustworthy (Butterfield & Metcalfe, 2001). The gamma correlation, based on the difference between concordant pairs and discordant pairs, is most commonly reported. However, its shortcoming is that it ignores pairs numerically tied on the predictor variable (e.g. accuracy); therefore, gammas can overestimate the strength of relationships (see Masson & Rotello, 2008). The Somers d coefficient overcomes this issue because it accounts for tied pairs while remaining otherwise identical to gamma (Pannu & Kaszniak, 2005). Given this, we report both measures.

The mean gamma correlation between confidence in the original error and retest accuracy was .54 (SE=.08) for younger adults, and .25 (SE=.08) for older adults, both of which were significantly greater than zero, $t(20) = 6.15$, $p<.001$ and $t(17) = 3.02$, $p = .01$, respectively. The Somers d analysis yielded similar correlations for younger, .52 (SE=.07), and older adults, .24 (SE=.08), again significantly greater than zero, $t(20) = 7.50$, $p<.001$ and $t(17) = 3.01$, $p<.01$, respectively. Although both groups showed an effect, correlations were significantly higher among younger relative to older adults both when using gamma $t(37) = 2.42$, $p = .02$ and Somers d, $t(37) = 2.62$, $p = .01$ (see Figure 4.3). This difference was marginally significant when the HADS depression score was included as a covariate for gamma ($p = .07$), but remained significant for Somers d ($p = .04$). The fact that no participant came close to meeting the cut-off for even mild depressive symptomatology suggests that depression does not explain the age-related reduction in the hypercorrection effect.
Figure 4.3: Mean Somers d coefficient between confidence in the initial error and retest accuracy as a function of age in Study 3A (bars represent ± 1 standard error).

Why did older adults show a weaker hypercorrection effect? One possibility is that their attention may not have been as markedly captured by unexpected negative feedback as it was for younger adults, consistent with evidence of an age-related asymmetry in the processing of valence information (Charles, Mather, & Carstensen, 2003; Mather & Johnson, 2000). Another possibility is that errors were more entrenched among older adults and thus more difficult to correct. This latter interpretation does not hold up to an analysis of repetitions of initial errors on the final test, where we found no differences between younger .24(.20) and older .30(.17) adults, $t(37) = 1.03$, $p = .31$. Thus, erroneous responding on the final test was not driven by the perseveration of errors per se.

In summary, the results of Study 3A successfully replicated the finding of a hypercorrection effect among younger adults and provided the first evidence of a smaller but significant effect among older adults. At the group level, both younger and older adults showed the same pattern of greater error correction with greater initial confidence (Figure 4.2). When this association was computed for each participant, however, older individuals showed a weaker relationship between correction and confidence on average compared to younger adults. As discussed, age-related increases in TOT states and reductions in metacognitive skills related to episodic memory may have contributed to this age-related reduction in the hypercorrection effect.
4.2 Study 3B

Study 3B sought to examine the relationship between error correction and confidence using a design that circumvents self-initiated retrieval of factual knowledge and explicit confidence ratings. Participants answered similar questions as in Study 3A but answers were presented in a multiple-choice format. Furthermore, instead of rating confidence, participants indicated how many answers they had narrowed their choice down to prior to answering. We predicted that incorrect answers on the initial test that were narrowed down to fewer choices would be more likely to be corrected on the final test, and that this effect would be comparable in younger and older adults.

4.2.1 Method

Participants

Twenty younger and 20 older adults volunteered to participate in this study; however, the data of one younger and three older adults were not included because they made no errors on the initial test. Therefore, the data from 19 younger and 17 older adults are reported. Participants were screened for major health conditions affecting cognition as in Study 3A and were recruited from volunteer research participant pools or responded to flyers posted at the University of Toronto and were paid for their participation. The mean age was 70.52 (SD = 3.27) for the older adults and 21.41 (SD = 2.24) for the younger adults. The groups did not differ in years of formal education, 15.12 (SD = 1.00) and 15.00 (SD = 1.00) respectively, $t(34) = .23$, $p = .82$, and younger adults scored higher on the MMSE (Folstein et al., 1975), 29.70 (SD = .77) and 28.95 (SD = 1.22), respectively, $t(34) = 2.19$, $p = .04$. Vocabulary scores on the Shipley Institute of Living Scale were higher in the older than younger group, 36.26 (SD = 2.05) and 31.59 (SD = 4.26), respectively, $t(34) = 4.27$, $p < .001$. Scores on the HADS did not differ between groups for the anxiety subscale, 5.41 (SD = 4.87) and 4.84 (SD = 3.27), respectively, $t(34) = .42$, $p = .68$, but older adults scored higher on the depression subscale relative to younger adults, 3.16 (SD = 2.32) and 1.47 (SD = 2.37), respectively, $t(34) = 2.16$, $p = .04$, although both groups are well below the cut-off for mild depression.\(^1\)

Materials

170 questions, similar to those used in Study 3A, were selected. For each question, 4 probable lures (i.e. wrong answers) were created. A lure was never a correct answer to another question. Lures were selected to be probable and familiar candidates so that they would compete with the correct answer. A Latin Square design was applied to ensure that there were an equal number of correct answers in positions 1 through 5 for each participant.

\(^1\)Conducting the same analyses with the HADS depression score as a covariate yielded the same results.
Procedure

Participants were tested individually using a computer and E-Prime software was used for stimuli presentation. The order of the questions was randomized for each participant. Participants saw a screen with a question centered at the top (What is the last name of the author of Canterbury Tales?) along with five answer alternatives numbered 1 through 5 (1-Shakespeare; 2-Doyle; 3-Chaucer; 4-Austin; 5-Dickens). Participants selected their answer by pressing the number key corresponding to their choice. Following their selection, their choice was underlined in red and the following question appeared on the same screen ‘How many alternatives did you narrow it down to?’ along with a scale ranging from 0 to 4. Participants made their selection by pressing a number key. For example, if a participant was torn between Chaucer and Dickens but opted for Dickens, they would answer ‘2’. If instead a subject had disqualified the alternative Shakespeare but had no idea among the remaining alternatives, they would select ‘4’ to indicate that they had only narrowed it down to 4 out of 5. A completely random guess would elicit the response ‘0’ (did not narrow it down or, put otherwise, narrowed it down to 0), and cases in which all but one option was disqualified would yield the response ‘1’. Next, regardless of accuracy, the correct answer was shown for 2 seconds. This sequence was repeated for all 170 questions. Participants were instructed to pay close attention to the feedback for a later test. Before beginning the actual task, participants completed practice questions to ensure proper use of the scale. Following study, there was a 15-minute break during which participants completed the Shipley scale, the HADS, and the MMSE. For the test phase, participants were shown the study questions on the computer screen in the same order that they studied them. For each question, participants typed in their answer and pressed the ENTER key to submit it and move to the next question. Both the study and test sessions were self-paced.

4.2.2 Results and Discussion

An alpha level of .05 was used for all statistical analyses.

Initial and final test performance

The mean gamma correlation between initial confidence ratings and initial recall performance, computed on each participant and then averaged over participants, was .54 (SE=.04) and .54 (SE=.07) for younger and older adults, respectively, and both were significantly different than zero (younger: t(16) = 13.05, p<.001; older: t(18) = 7.78, p<.001), demonstrating comparable calibration. A repeated-measures ANOVA on the number of errors committed at each confidence level revealed a significant main effect of confidence, $F(1,35) = 4.38, p<.01, \eta^2_p = .16$. The main effect of age was non-significant, $F(1,35) = 1.57, p = .22$, as was the interaction, $F(1,35) = 2.35, p = .14$. As in Study 1, a repeated-measures ANOVA revealed a greater increase in performance from
initial to final test for younger relative to older adults (see Figure 4.4), $F(1,35) = 30.18, p<.001, \eta^2_p = .46$.

![Graph showing mean accuracy on initial and final free recall as a function of age in Study 3B (bars represent ± 1 standard error).](image)

**The hypercorrection effect**

In Study 3B, a hypercorrection effect would be evident if errors made from a selection of fewer candidates were more likely to be corrected on the final test than errors endorsed from a greater number of candidates. To test this, we rescaled the original responses to a scale of 1 (no alternatives eliminated) through 5 (four alternatives eliminated). The Age X Confidence Level repeated-measures ANOVA showed a significant main effect of confidence in errors on the initial test and final test accuracy, $F(1,35) = 10.94, p<.001, \eta^2_p = .26$, no main effect of age, $F(1,35) = 2.50, p = .12$, and a significant interaction, $F(1,35) = 4.05, p = .04, \eta^2_p = .12$ (see Figure 4.5). As discussed earlier, however, the within-subjects measures below are more dependable.
The mean gamma correlations between confidence in the original error and retest accuracy were .38 (SE=.09) for younger and .33 (SE=.09) for older adults, both of which were significantly greater than zero, \( t(16) = 4.02, p = .001 \), and \( t(18) = 3.02, p < .01 \), respectively. The mean Somers \( d \) coefficients were .36 (SE=.09) for younger and .32 (SE=.10) for older adults, again significantly greater than zero, \( t(16) = 4.01, p = .001 \), and \( t(18) = 3.39, p < .01 \), respectively. Gamma correlations and Somers \( d \) coefficients, however, were not significantly different between groups, \( t(33) = .39, p = .70 \), and \( t(33) = .29, p = .77 \), respectively (see Figure 4.6).
Figure 4.6: Mean Somers d coefficient between confidence in the initial error and retest accuracy as a function of age in Study 3B (bars represent ± 1 standard error).

Replicating Study 3A, younger and older adults both displayed a hypercorrection effect; however, these procedural changes eliminated the age difference in its magnitude. When TOT states were avoided and confidence was queried in a way that bypassed explicit confidence ratings, younger and older adults were equally able to correct higher-confidence errors. This suggests that age-related differences in retrieval or metacognition may have accounted for the weaker hypercorrection effect in Study 3A among older adults.

4.3 General Discussion

In two studies we found that both younger and older adults hypercorrect their errors. The way in which confidence was queried in Study 3B is novel and likely taps confidence in a different way than do explicit ratings of confidence. Indeed, if asked what the capital of Australia is, a learner may eliminate Sydney, Perth and Melbourne knowing that the capital is a lesser known city, but then hold little confidence over the remaining two options, Canberra and Brisbane. This would result in low confidence as measured on a traditional scale but higher confidence when operationalized as a process of elimination. To address this, we looked at the distribution of confidence responses for errors among younger and older adults in Studies 3A and 3B. Typical studies of this nature report more lower relative to higher confidence errors, and indeed this was found in Study 3A for both younger and older adults ($F(6,126) = 17.71$, $p<.001, \eta^2_p = .87$ and $F(6,96) = 17.88$, $p<.001, \eta^2_p = .84$, respectively). However, this effect was not found in Study 3B among either younger ($F(4,56)<1$),
Eliminating alternatives may reduce interference with the correct answer. For instance, a learner may have little confidence that her selection Queensland is correct over Canberra, but when shown the correct answer (Canberra) may find it easy to update her memory. By contrast, if the learner has no knowledge of Australia and is unable to eliminate any options, the four incorrect alternatives are connected to the question with equivalent strength. This explanation is reminiscent of the fan effect which describes increasing interference as a function of the number of concepts associated with an item (Anderson & Reder, 1999). Regardless of the cognitive underpinnings of our results, the conclusion is still warranted that errors are hypercorrected in the multiple-choice paradigm.

To our knowledge, we are the first to use Somers d as an alternative to gamma for examining hypercorrection. We agree with others (Pannu & Koszniak, 2005) that the statistical features of Somers d make it more appropriate as it avoids artificially-inflated correlations. Therefore, we recommend that this measure be adopted moving forward. We maintain that within-subjects measures of association are superior to analyses of variance for examining hypercorrection as it accounts for variations in the number of errors committed at each confidence level and potential group differences in scale use.

In summary, our results suggest that older adults can indeed learn from their mistakes when certain learning parameters are met. This is the first demonstration of a hypercorrection effect in aging and runs counter to a body of neuropsychological literature which has cautioned against errors for individuals with episodic memory declines (Baddeley & Wilson, 1994; Anderson & Craik, 2006; Clare & Jones, 2008). Rather, these findings are consistent with the hypothesis that errors made in supportive, conceptually relevant learning paradigms can be effectively overwritten by older adults (Cyr & Anderson, 2012) and that the same factors that mediate feedback processing in younger adults are also at play as we age.
Chapter 5

General Discussion

Past wisdom from experimental and clinical work maintains that younger but not older adults learn from their mistakes. The objective of this dissertation was to investigate the veracity of this claim using learning paradigms that unconfounded task specific effects from aging effects. Across studies I found evidence that, by and large, older adults learn from their mistakes in a similar fashion as their younger peers. The results from Study 1 showed that for both age groups, trial-and-error learning increased episodic memory performance relative to errorless learning when it invoked conceptual processing, consistent with previous work among younger adults (Kornell et al., 2009; Knight et al., 2012; Kornell, 2014; Richland et al., 2009; Hays et al., 2013) and older adults (Cyr & Anderson, 2012). By contrast, trial-and-error decreased memory relative to errorless learning during nonconceptual lexical processing, replicating findings from the aging literature (Clare & Jones, 2008) and extending it to younger adults. For both groups, conceptual targets were likely to be recalled along with their prior guesses, supporting the notion that remembering ones errors can scaffold retrieval (Knight et al., 2012). Study 2 replicated this age-invariant benefit with conceptual errors that were strongly and weakly constrained by cues (Study 2A) as well as semantically near and far from targets (Study 2B). Finally, Study 3 showed that older adults are equally capable as younger adults of correcting entrenched errors when conditions support age-related memory declines, extending the hypercorrection effect (Butterfield & Metcalfe, 2001) to healthy aging. In sum, conceptual processing, semantic activation, as well as indices of personal relevance such as confidence and prior knowledge, are aspects of error generation that boost episodic memory across the lifespan.

Nonetheless, age differences in some aspects of memory performance were observed across studies. First, aging was often associated with decreased overall memory performance. Younger adults performed better following lexical learning in Study 1 relative to older adults, consistent with research showing age differences in memory when encoding emphasizes perceptual rather than concep-
tual processing (Troyer, Hafliger, Cadieux, & Craik, 2006). Older adults also showed less learning from the initial to the final general knowledge tests in Studies 3A and 3B, consistent with general age-related declines in episodic memory. Second, aging modulated the magnitude of error generation effects: Cued recall performance for targets increased less from errorless to trial-and-error learning among older relative to younger adults. Moreover, older adults committed more (albeit infrequent) prior error intrusions relative to younger adults (i.e., Studies 1, 2A and 2B), consistent with our prior work (Cyr & Anderson, 2012). Third, aging appeared to influence the mechanisms that underpin error generation effects. For instance, successful recall of lexical targets in Study 1 was associated with greater forgetting of prior guesses for older adults, whereas younger adults continued to remember them along with targets. This suggests that nonconceptual errors were retrieval competitors and allies for older and younger adults respectively, in line with age-related vulnerabilities to interference (Lustig et al., 2007). Moreover, the semantic proximity of errors to targets did not affect the magnitude of the error generation benefit among older adults; for younger adults, errors were more beneficial when they were closely relative to distantly related to the target.

To summarize, the error generation effect is beneficial for older adults, but this advantage is mitigated by age-related declines in the cognitive domains discussed in the Introduction: recollection, source memory, and inhibition. Decreases in recollection and source memory make it more difficult to constrain retrieval to aspects of memoranda that are important for discriminating non-targets such as errors from targets, driving target memory down and false alarms to prior errors up (Hay & Jacoby, 1999). An inhibitory account would suggest that this pattern is due to a failure to suppress erroneous and irrelevant information in favour of target memory (Hasher & Zacks, 1988). Although these accounts approach interference resolution differently, they are similar to the extent that they posit the role of a control mechanism to hone in on target information, by dialing up the signal and/or dialing back the noise. In the framework that follows, I favour the concept of cognitive control as it describes this regulatory action without making claims about what is being acted upon. Deficits in these domains, however, cannot account for the fact that the benefit of errors is not swayed by their semantic proximity to the target among older adults. Although this finding is difficult to interpret without more data, a possibility is that the maintained integrity (Park et al., 2002) and increased flexibility (Taylor & Burke, 2002) of semantic networks with aging allows older adults to effectively embed semantically disparate errors into memory traces.

In the next section, I present a framework which extracts the principles that determine the effects of error generation on episodic memory. This framework aims to explain the data in this dissertation, as well as make predictions about the roles of other potential factors.
5.1 Framework of episodic error generation

Errors introduce potentially detrimental interference, but varying external factors such as the amount of conceptual processing and semantic activation engendered by error generation, as well as internal factors like age, confidence and prior knowledge affect the application of cognitive control to combat it. I propose that the management of errors is accomplished by engaging cognitive control at different temporal points during learning. Braver put forth a dual mechanism account of cognitive control which outlines how control can be deployed proactively or reactively to manage interference, and that the weighting between them is determined by situational factors (Braver, Gray, & Burgess, 2007; Braver, 2012). Specifically, proactive control gates information in an anticipatory way to prevent interference before its onset, whereas reactive control relies on elaboration of information to resolve interference after its onset. In the context of episodic trial-and-error learning, proactive control could be engaged before feedback in anticipation that one's response is incorrect, thus gating potential interference. Also, reactive control could be engaged following error feedback in order to suppress the prior response and amplify the correct answer. Figure 5.1 (Panel A) illustrates how reactive control may retroactively elaborate on associations between the cue, error and targets to organize and enrich the memory trace in service of memory. In addition, Figure 5.1 (Panel B) depicts how proactive control may operate by pre-emptively dialing down the activation of the response. While both forms of control can be applied in service of target memory, reactive control is likely more useful when the contents of memory map onto pre-existing structures that can be quickly activated to efficiently organize them in a just-in-time manner. Conversely, proactive control is likely more critical to resolving interference when the contents of memory are difficult to meaningfully integrate at the back-end. In such cases, responses are best gripped by control processes from the front-end so that their influence can be minimized should they be incorrect (see Braver, 2012 for a similar discussion).

![Figure 5.1: Schemata representing the action of dual modes of cognitive control for error resolution.](image-url)

Braver's theory also makes aging predictions, showing that older adults are less able than their...
younger counterparts to engage proactive control, relying more on reactive control, while younger adults can use both modes of control flexibly (Paxton, Barch, Racine, & Braver, 2008; see also Velanova, Lustig, Jacoby, & Buckner, 2007; Vaden, Hutcheson, McCollum, Kentros, & Visscher, 2012; Karayanidis, Whitson, Heathcote, & Michie, 2012). This is attributed to the fact that older adults fail to engage appropriate top-down attentional sets at the front end of learning, and, to compensate, frontally mediated processing must be harnessed at later stages. This converges on Craiks theory of environmental support which proposed that due to age-related reductions in cognitive resources, older adults are less able to engage in self-initiated processes and are more reliant on environmental or contextual support for successful memory performance (Craik, 1986).

In the next section, I show how this framework fits the pattern of error generation effects among younger and older adults found across the dissertation studies as function of the main factors that were manipulated: 1) type of processing; 2) semantic proximity; 3) confidence/prior knowledge. To foreshadow, the framework postulates that resolution of errors that map onto existing semantic structures preferentially engage reactive control whereas errors that do not spur semantic activation are best managed with proactive control. These effects will be considered in light of age-related changes in proactive and reactive control.

### 5.1.1 Cognitive aspects of the framework

The findings of Study 1 showed an asymmetry of memorial effects as a function of the type of processing engendered by error generation. I propose that resolution of conceptual errors is effectively managed by reactive control because cues, errors and targets can be easily organized and elaborated upon retroactively within an existing semantic structure. Figure 5.2 (Panel A) illustrates how reactive control can exploit semantic scaffolding (dotted lines) to support error resolution. For example, upon learning that the correct target to the category a flower is rose and not violet as they had guessed, the learner may think ‘Roses are red, violets are blue’ to integrate the memory traces of the error and the target. By contrast, this is more difficult when resolving nonconceptual errors, which puts the onus primarily on proactive control to gate them from the onset. Figure 5.2 (Panel B) illustrates how in the absence of semantic support, proactive control must work to dampen the activation of the response in case it is incorrect. Indeed, upon learning that the correct target to the word stem ro____ is rose and not rope as one had guessed, it would take much more time and effort to retroactively embed them semantically into a single memory trace. Thus, while both conceptual and nonconceptual errors can engage dual modes of cognitive control, only conceptual error generation can enhance the utilisation of reactive control in service of target memory. Moreover, nonconceptual interference relies more on proactive control, which is much more resource-demanding, substantially reducing the available capacity of working memory for maintenance of information (Braver, 2012).
Factor 1: Type of processing

Aging does not modify this general pattern, but it does affect the magnitude of harms and benefits afforded by error generation. In Study 1, conceptual target memory increased from errorless to trial-and-error learning among older adults, but not to the same extent as among their younger peers: Conversely, nonconceptual target memory decreased from errorless to trial-and-error learning among older adults, and to a greater extent than among their younger peers. The framework fits these data by showing that conceptual error resolution relies more heavily on reactive relative to proactive control: This weighting capitalizes on the strengths of older adults, allowing them to benefit along with younger adults. However, age differences are not eliminated because younger adults can engage proactive control more so than older adults to potentiate this boost. By contrast, nonconceptual errors provide no semantic support, and their interference is more optimally managed by engaging proactive control which is reduced by aging: This plays to the weaknesses of older adults episodic memory, leading to larger memory impairment following nonconceptual error generation relative to younger adults.

In sum, the intact reactive (i.e. context-supported) control of older adults allows them to benefit more fully from the semantic support at encoding and retrieval afforded by conceptual errors; however, compromised proactive (i.e. self-initiated) control drives down the magnitude of this benefit and exacerbates the negative influence of nonconceptual errors. At first blush, this appears contradictory to my previous findings that older adults benefited more from conceptual trial-and-error relative to errorless learning than younger adults (Cyr & Anderson, 2012). However, this can be explained by the fact that those studies used recognition as opposed to cued recall to query episodic memory. It is likely that increasing environmental support at retrieval leveled the playing field between age groups, and that the memory of older adults was more rewarded by the combination of encoding conditions that enhanced elaborative processes and testing conditions that guided retrieval. This is supported by studies showing that compared to younger people, older adults show increased activation of regions associated with reactive control during word recognition.
relative to recall (Jimura & Braver, 2010; Cabeza, Anderson, Locantore, & McIntosh, 2002).

**Factor 2: Semantic proximity**

The findings from study 2 showed that conceptual errors were more helpful for younger adults with increasing proximity to the target within a semantic structure, i.e. when they were members of the same taxonomic family. I propose that errors that are semantically near relative to far from the target, like *liver* is to *heart*, exploit reactive control to a greater extent (see Figure 5.3 Panel A). A strong association means that retroactively organizing errors and targets is easier, allowing more time to elaborate on the association. By contrast, a weak association between the error and the target, like *cedar* is to *heart*, requires one to shift from one semantic structure to another, leading to less efficient scaffolding from errors to targets (see Figure 5.3 Panel B).

Figure 5.3: Schemata representing strength of associations as a function of semantic proximity.

**Factor 3: Confidence and prior knowledge**

Results from study 3 showed that both younger and older adults are more likely to correct errors that are held with higher than lower confidence. As discussed, another way of interpreting this effect is to equate confidence with cue familiarity, such that high confidence errors are also those that strongly co-activate many related concepts in semantic memory (see Figure 5.4 Panel A). By contrast, low confidence errors weakly activate related concepts due to lack of familiarity with cue (see Figure 5.4 Panel B). I propose that high confidence errors can be more effectively resolved with reactive control relative to low confidence errors, because they can be retroactively integrated into existing knowledge more readily. Moreover, feedback following high relative to low confidence errors captures attention more readily (Fazio & Marsh, 2009), which would more strongly engage control from the back-end. Thus, resolving high confidence errors not only relies on age-spared reactive control, but also ramps up proactive control, allowing older adults to catch up with their younger peers and show equivalent hypercorrection (i.e., Study 3A).
To summarize, error generation can foster both interference and elaboration, to negative and positive memorial effect respectively. A number of external and internal factors determine the direction of error generation effects as a function of how they engage proactive (i.e. self-initiated) and reactive (i.e. context supported) cognitive control mechanisms. In the next section, I map neural correlates to this framework based on imaging research on cognitive control, interference resolution and aging.

5.1.2 Neural aspects of the framework

The neural substrates of self-generated errors produced in the context of episodic learning have never been directly examined, but three primary candidate brain areas can be gleaned from the literature on cognitive control and interference resolution as these processes similarly require the inhibition of task irrelevant information such as errors. I will focus on the potential roles of the lateral prefrontal, the anterior cingulate, and the hippocampus in resolving episodic errors in younger and older adults.

First, functional magnetic resonance imaging (fMRI) research on Braver’s theory of dual modes of control points to the lateral prefrontal cortex (PFC) as a critical region for the flexible application of cognitive control (Paxton et al., 2008). Paxton et al. (2008) used a continuous performance task in which participants saw a series of cuedelayproberesponse sequences, where each cue and probe was a single letter, and were to press a button only if the probe was an X that was preceded by an A. The authors argued that a target cue (A) invokes proactive cognitive control in anticipation of a target probe (X), while a target probe (X) invokes reactive cognitive control to ensure it was preceded by a target cue (A). The temporal dynamics of fMRI blood oxygen level dependent (BOLD) signals in the lateral PFC fit a self-initiated control pattern in the younger adults (greater cue-based activation), whereas in older adults a reactive control pattern was found (greater probe-based activation). Specifically, activation of the lateral PFC was reduced in older relative to younger adults when self-initiated control was engaged, but increased when reactive control was engaged.
(see also Velanova et al., 2007; Grandjean et al., 2012). Other findings are generally congruent, although they further distinguish between ventrolateral and dorsolateral regions of the lateral PFC. For example, the ventrolateral PFC is recruited in service of interference resolution in the Stroop task (Nee, Wager, & Jonides, 2007; Nee, Jonides, & Berman, 2007) as well as successful inhibition of distractors (Wais, Kim, & Gazzaley, 2012; Atkins & Reuter-Lorenz, 2011). Moreover, recent research suggests that the ventrolateral PFC (Atkins & Reuter-Lorenz, 2011; Han, O’Connor, Eslick, & Dobbins, 2012) in addition to the dorsolateral PFC (Herrmann, Rotte, Grubich, Ebert, Schiltz, & Munte, 2001) are critical for resolving interference from related competitors, and older adults show reduced activation of the ventrolateral PFC in these situations (Jonides et al., 2000). In sum, studies on cognitive control and interference converge in finding an important role for the lateral PFC in harnessing cognitive control and inhibiting irrelevant information which I argue are necessary during trial-and-error learning.

Second, the involvement of the anterior cingulate cortex (ACC) in response to error processing has been well documented (Ueno et al., 2011; Botvinick et al., 2004). In the work above, the authors suggest that the transient activity in the lateral PFC during their continuous performance task might be triggered via the detection of interference through the engagement of conflict-monitoring regions such as the ACC (Paxton et al., 2012). The ACC is also recruited when inhibition of a prepotent response is required (Takeuchi et al., 2012; Zoccatelli, Beltramello, Alessandrini, Pizzini, & Tassinari, 2011) and when multiple memory representations compete for selection (Nee et al., 2001; Hermann et al., 2001; MacDonald, Cohen, Stenger, & Carter, 2000; Barch, Braver, Sabb, & Noll, 2000; Thompson-Schill, D’Esposito, Aguirre, & Farah, 1997).

Third, there is evidence that the ACC may detect the need for control and signal the dorsolateral PFC (MacDonald et al., 2000) to interact with the hippocampus to suppress retrieval of an unwanted memory (Anderson et al., 2004). This is consistent with research showing that the hippocampus is activated during retrieval of correct and incorrect information equally, the output of the latter being thought to trigger conflict detection in the ACC, which in turn upregulates the right dorsolateral PFC to bias retrieval towards correct information (Kuhl, Dudukovic, Kahn, & Wagner, 2007). Together, these findings indicate that the ACC detects errors and recruits the lateral PFC to engage cognitive control, thereby biasing processing in the hippocampus.

In the context of my framework, I propose a pattern where activation in the hippocampus spreads to the ACC, and from there to the lateral PFC, then back to the hippocampus. This proposed network starts in the hippocampus because it has been hypothesized to obligatorily retrieve information, and is activated regardless of the accuracy of the information, consistent with it being an automatic module (Kuhl et al., 2007; Moscovitch, 2008). We would expect the strength of the functional connection between the hippocampus and the ACC to be weaker for older than younger adults, reflecting age-related decrements in encoding and retrieval. Let us now consider how this network would be affected by feedback during trial-and-error learning. When the learner receives
negative feedback following a response, we would expect activity in the ACC to down-regulate activity across these nodes, reflecting the role of the ACC and lateral PFC in error detection and cognitive control to bias memory towards wanted information and gate processing of unwanted memories (Ueno et al., 2009; Botvinick et al., 2004). During positive feedback, we would expect activity in the ACC to up-regulate activity in the lateral PFC, which in turn would amplify activity in the hippocampus. We would expect similar results during retrieval, with positive connections across nodes for target words, and negative connections for prior error words. The weighting of reactive and proactive cognitive control required for the task (i.e., conceptual vs nonconceptual) would modulate when these areas come online to manage interference. Finally, we would expect age-related decrements in the strength of these paths when self-initiated error control processes are required (nonconceptual trial-and-error), and minimized when reactive error control processes are invoked (conceptual trial-and-error).

In sum, I propose that the processing invoked by error generation as well as the age of the learner alter the weighting between early self-initiated cognitive control and later context supported cognitive control for resolving error interference. On a neural level, this implicates the involvement of the lateral PFC, the ACC and the hippocampus which are known to be sensitive to interference and aging.

5.2 Emotional and motivational factors

My dissertation studies manipulated learning parameters to infer how errors are differentially processed at a cognitive level among younger and older adults. However, my studies do not speak to the potential role of emotional and motivational factors on episodic error processing, an important consideration given that they are known to guide attention and memory via brain regions that overlap with those used for cognitive control (for a review see Banich et al., 2008). In this next section, I briefly review three such factors known to mediate age-related changes in episodic memory, and discuss how they may contribute to error generation effects as predicted by my framework.

5.2.1 The positivity bias

There is increasing evidence that healthy aging is accompanied by an asymmetry in memory for valenced information such that younger adults better remember negative information (Charles, Mather, & Carstensen, 2003; Grady, Hongwanishkul, Kightley, Lee, & Hasher, 2007; Kensinger, Brierley, Medford, Growdon, & Corkin, 2002) whereas older adults show reduced memory for negative information (Charles et al., 2003; Grady et al., 2007) or enhanced memory for positive information (Charles et al., 2003; Knight, Maines, & Robinson, 2002; Thomas & Hasher, 2006). This positivity bias in memory is due to enhanced attentional processing of positive stimuli at
encoding at the expense of negative and neutral stimuli (for a review, see Mather & Carstensen, 2005). On a broader motivational level, this asymmetry has been attributed to an increase in top-down emotion regulation with age, with older adults looking to maintain positive mood by orienting attention away from negative and toward positive information (Carstensen, Isaacowitz, & Charles, 1999; Mather & Carstensen, 2005). Given these findings, it is possible that age differences in the error generation effect are partly attributable to older adults engaging less with negative feedback, resulting in less effective error suppression. However, extending the positivity effect to error generation is complicated by the fact that valence is conveyed differently in each context. The positivity studies above typically examined attention and memory for stimuli such as images, faces and words, wherein valence is determined by the items representation. For example, the word cancer triggers more negative affect, images, and thoughts relative to the word sunshine. By contrast, feedback indicating that ones response rose to the category a flower is incorrect during episodic learning is “negative” because it threatens current memory representations, not because it conjures negatively arousing representations. To my knowledge, no studies have investigated whether younger and older adults differentially process positive and negative feedback following episodic error generation.

However, many studies have looked for a positivity bias in the processing of feedback during probabilistic/ rule-based learning using EEG. Results across these studies are decidedly mixed, with some finding that relative to younger adults, older adults rely more on positive feedback (Eppinger et al., 2008; Pietschmann, Endrass, Czerwon, & Kathmann, 2011; Samanez-Larkin et al., 2007), negative feedback (Frank & Kong, 2008; Hummerer, Li, Miller, Lindenberger, 2010), or show no difference between negative and positive feedback for learning (Simon, Howard, & Howard, 2010; Bellebaum, Rustemeier, & Daum, 2012). However, Ferdinand and Kray (2013) point out that in these probabilistic paradigms, feedback becomes less useful and more expected as the individual learns, potentially confounding age differences in learning and feedback processing. In their study, they found that when feedback was unexpected, i.e. unconstrained by rules acquired throughout learning, there were no age differences as a function of valence. This suggests that aging may not affect how we react to learning that our response is correct or incorrect during episodic learning if the feedback is not predictable.

In sum, the positivity bias is a tenuous explanation for error generation effects given that there is mounting evidence that any valence asymmetry in feedback processing with age will be highly variable and sensitive to task demands (Ferdinand & Kray, 2013). Likewise, many studies have failed to find positivity effects in memory (Grady et al., 2007; Kensinger, Garoff-Eaton, & Schacter, 2007; Kensinger et al., 2002; Comblain, DArgembeau, Van der Linden, & Aldenhoff, 2004; Denburg, Buchanan, Tranel, & Adolphs, 2003; Gruhn, Smith, & Baltes, 2005) suggesting that it similarly hinges on stimulus and task parameters (for a review, see Reed, Chan, & Mikels, 2013). Indeed, applying the positivity literature to feedback conditions across my dissertation
studies cannot account for findings: For both age groups, negative feedback led to poorer memory in nonconceptual learning, but better memory in conceptual learning. Task parameters are important because the valence of items is not triggered by an external purely positive or negative signal like right or wrong in the case of feedback but rather internal subjective valuations. For instance, older adults show a stronger bias toward positive information when they view it as personally relevant (Tomaszczyk, Fernandes, & Macleod, 2008) and when they are not told that their memory will be tested (see Reed et al., 2013). These top-down motivational shifts with aging are known to affect memory and likely influence error generation effects as we will see in the following sections.

5.2.2 Environmental stress

It is accepted that when we administer a memory test, we also induce stress to some degree. The amount of stress occasioned by testing depends on a number of internal factors, namely the importance of the task to the learner and their perceived competence to perform it (i.e., how much it threatens ones ego) (Dickerson & Kemeny, 2004). While environmental stress can be detrimental to episodic learning and memory among younger adults, older adults are especially sensitive to its effects (see Lupien, Maheu, Tu, Fiocco, & Schramek, 2007). For instance, studies have demonstrated that when instructions for memory tasks are modified to decrease emphasis on the memory aspect of the task, differences in memory performance between younger and older adults that were found previously are eliminated (Hasher, Zacks, Rahhal, 1999; Rahhal, Colcombe, Hasher, 2001). This has been interpreted in light of negative stereotypes about aging which are activated by emphasizing a domain (e.g., memory) known to undergo age-related decline (Chasteen, Bhattacharyya, Horhota, Tam, & Hasher, 2005). Indeed, the effect of manipulating memory instructions hinges on the extent to which older participants value memory ability (Hess, Auman, Colcombe, & Rahhal, 2003). These effects are partly attributable to the fact that stress affects functioning in brain regions that mediate cognitive control, such as the prefrontal cortex, and that older brains are less resilient to its effects (McEwen & Morrison, 2013).

Given these findings, it is possible that older adults are more stressed by negative feedback following error generation than younger adults as it signals poor performance. Errors may be more threatening to older adults, especially during learning of information for which they perceive themselves as low competence. This idea is consistent with findings from my dissertation showing that age differences in target memory following trial-and-error learning are minimized or eliminated when error generation is constrained by semantic knowledge (e.g., hypercorrection of misconceptions, conceptual error generation), and exacerbated when it is largely unconstrained (e.g., nonconceptual error generation). Older adults may perceive themselves as more competent in the former situations which capitalize on their intact semantic knowledge, leading to lower stress and better cognitive control.
In sum, feedback-induced stress, like other kinds of environmental stress, should not be expected to undermine episodic memory performance in all learning contexts as it hinges on the interaction of task demands and individual beliefs. Thus far, I have discussed how older adults may be less motivated to engage with error feedback due to top-down goals of avoiding negative affect and stress. However, older adults are just as capable as their younger peers at engaging with negative information if they deem it relevant or important (Mather & Carstensen, 2005). In the next section, I discuss how relevance is especially important for older adults, and how this may influence error generation effects.

5.2.3 Personal relevance

Given the key role of memory for survival, it is unsurprising that attention and memory are captured by stimuli that are seen as valuable to individuals. For instance, both younger and older individuals are more likely to remember information when they are required to evaluate whether or not it self-referential (e.g., *does the word honest describe you?*) (Kuiper & Rogers, 1979; Dulas, Newsome, & Duarte, 2011; Hamami, Serbun, & Gutchess, 2011; Serbun, Shih, & Gutchess, 2011) or subjectively pleasing to them (e.g., *do you like this object?*) (Dulas et al., 2011; Gusnard, Akbudak, Shulman, & Raichle, 2001; Raposo, Vicens, Clithero, Dobbins, & Huettel, 2011).

However, older adults show a greater disparity in episodic memory for items with intrinsically high relative to low subjective value. Older adults strategically shift attention in a top-down fashion to recall high-value information at the expense of lower value information to maximize memory efficiency (Castel, 2008; Castel, Benjamin, Craik, & Watkins, 2002). Relevance also modulates memory for contextual features associated with items. In a seminal study, Brown, Jones, and Davis (1995) asked younger and older adults to participate and listen in on conversational exchanges between inquirers (I) and responders (R). On a later source memory task, participants were shown statements and had to recall the individual who had made them. They found typical age-related source memory deficits when the exchange pertained to categories (e.g., I: *a precious stone*; R: *emerald*) but no age differences when it pertained to personally relevant information (e.g., I: *Country you want to visit*; R: *Thailand*). Age differences in source memory are also eliminated when source dimensions are associated with meaningful features such as the veracity of a statement (Rahhal, May, & Hasher, 2002) or the safety of an item (May, Rahhal, Berry, & Leighton, 2005) relative to perceptual features. More recently, Germain and Hess (2007) found a pattern of better memory for more personally relevant information under conditions that manipulated the age of the target person in the narrative (i.e., older versus younger) as well as the topic of the narrative (e.g., rising health care costs versus university budget cuts). Interestingly, they demonstrated that increased relevance was associated not only with better memory performance, but with a greater ability to inhibit distracting, irrelevant information interspersed within the passage. Moreover, these effects
were stronger within the older adult sample, suggesting that interest or relevance has a larger positive impact on memory functioning and cognitive control with aging.

Overall, these findings suggest that while both age groups preferentially encode self-referential stimuli, older adults may be more captured by the intrinsic subjective relevance of information in their environment. Given these findings, we can predict that the perceived relevance of the learning task would modulate the error generation benefit among older adults. When errors occur in the context of learning interesting information, older adults may be better able to deploy cognitive control to suppress them, similar to how relevance increased their ability to inhibit distracters in the study above (Germain & Hess, 2007). While relevance was not directly manipulated in my dissertation studies, it is notable that there were no age differences in error correction during learning of general knowledge trivia, provided that encoding and retrieval conditions were supportive (Study 2B). Learning information that is constrained by existing knowledge capitalizes on the experience of older adults, and is undoubtedly more interesting than learning arbitrary cue-target pairs (e.g., chest – heart). Further, personally relevant items are also likely to be those that map most strongly onto existing knowledge, such that we know more about things we find interesting. It would be interesting to examine error generation effects during learning of information that is completely novel but that older adults still find valuable. This would help disentangle whether interest/relevance acts primarily by activating semantic associations that allow the information to be more easily contextualized, or by triggering self-initiated processing driven by curiosity (see Kang et al., 2009).

In sum, motivational and emotional factors likely play a role in determining the effects of error generation among younger and older adults. Findings from studies on the effects of valence, stress, and relevance all converge in positing a top-down shift in cognitive control with age to maximize processing efficiency and regulate emotions. Based on these findings, I propose that age differences in the error generation effect will be minimized when learning parameters are aligned with the areas of competence and interest of older adults. Increased interest and motivation will increase deployment of deficient proactive (e.g., self-initiated) control and capitalize on their vast existing knowledge to exploit intact reactive control.

5.3 Paradoxical effects of retrieval: Facilitation versus forgetting

In my dissertation, I have mainly interpreted the effects of error generation on target memory in light of the benefits of retrieving related information. I have shown that when errors spur retrieval of conceptually related concepts, memory for both errors and targets is enhanced. This is congruent with a wealth of research showing memorial benefits of testing for retrieved materials (Karpicke &
Roediger, 2008), but also studies finding that these testing benefits spill over to related but not retrieved materials (Carpenter, Pashler & Vul, 2006; Chan, McDermott, & Roediger, 2006; Chan, 2009). For instance, Carpenter et al. (2006) found that in a paired associates learning task (angle – corner), retrieval practice (angle – ?) enhanced later memory for the target word (corner) as well as the cue (angle). According to this retrieval-induced facilitation view, successful memory is a consequence of the associative structure of memory. However, there is an extensive body of literature suggesting that retrieval can equally generate significant interference, especially of items that are related to retrieved items. According to this retrieval-induced forgetting view, forgetting is a consequence of the associative structure of memory. In this section, I attempt to reconcile my framework of error processing with the literature on retrieval-induced forgetting.

In a seminal study, Anderson, Bjork, and Bjork (1994) examined the effects of retrieval practice on memory for tested and nontested items. During the study phase, participants studied a list of category-exemplar pairs (e.g., fruit – apple, fruit – orange, metal – brass, metal – iron). This was followed by a retrieval practice phase, where participants performed a cued recall test on half of the exemplars in a select set of categories (e.g., they might be tested on fruit-ap but not fruit-or, but no items from the metal category). During the final test phase, participants were given a cued recall test for all items from the study phase, i.e., retrieval practiced items (Rp+), nonpracticed items (Rp-) and items from the nonpracticed category (Nrp). Results showed that the recall probability of the Rp- items was lower relative to Nrp items, suggesting that the act of retrieving Rp+ items impaired the recall of related items. This is the typical finding of retrieval-induced forgetting, which Anderson and colleagues attribute to an executive mechanism which actively inhibits related competitors to enhance recall of targets (for a review, see Anderson, 2003). As such, the inhibitory-deficit hypothesis (Hasher & Zacks, 1988) would predict that retrieval-induced forgetting should decline with increasing age; however, most studies have found it preserved among older adults (Hogge, Adam, & Collette, 2008; Gomez-Ariza, Pelegrina, Lechuga, Suarez, & Bajo, 2009; Ortega, Gomez-Ariza, Roman, & Bajo, 2012) with only a few finding age-related inhibitory deficits (Aslan & Bauml, 2012). It should be noted, however, that the inhibitory account of retrieval-induced forgetting is contested, with some arguing that interference is the result of strengthening practiced items which blocks access to nonpracticed items (see Verde, 2012). As such, the effects of aging on retrieval-induced forgetting remain unclear.

At first glance, the retrieval-induced forgetting literature appears incompatible with my framework of error processing, which posits retrieval-induced facilitation. However, I have argued that error generation binds the cue, error and target, making them components of a single multipart representation as opposed to individual competitors. Indeed, Anderson and McCulloch (1999) found that retrieval-induced forgetting was eliminated when participants in the retrieval-practice paradigm above were asked to inter-relate the exemplars within a category during the study phase (e.g., relate apple and orange) relative to the standard encoding condition (e.g., relate fruit and
This is congruent with findings that the fan effect is significantly reduced when the facts associated with a concept can be meaningfully integrated into one scene (Radvansky & Zacks, 1991). In a more recent study, Goodmon and Anderson (2011) found that retrieval-induced forgetting was abolished when there were many pre-existing semantic associations between the practiced and the nonpracticed set. These findings suggest that when encoding conditions promote integration of targets and competitors, or when there are pre-existing semantic associations between them, interference is reduced or eliminated.

This integration effect aligns perfectly with the presented framework of error processing, and accounts for our finding that conceptual error generation rewarded episodic memory for both errors and targets as they were presumably embedded into the same memory trace. When errors and targets shared closer pre-existing relations (e.g., Studies 1, 2B and 3), this integration was facilitated and interference was further reduced in service of memory. By contrast, nonconceptual targets and errors could not be semantically integrated, leaving multiple memory traces to compete at retrieval. Therefore, errors were suppressed in favour of targets at retrieval, accounting for the fact that they were less likely to be recalled than in the conceptual condition. Overall, these findings across studies support the initial hypothesis that errors can act as stepping stones toward the target when learning enables semantic processing. However, one question is whether findings from the retrieval-induced forgetting literature can really be extended to nonconceptual error generation given that the retrieval-practice paradigm typically manipulates semantic/conceptual associations. In fact, it has been suggested that retrieval-induced forgetting can only be seen for conceptual representations in memory (Perfect, Moulin, Conway, & Perry, 2002). However, one study did look at the effects of retrieval-induced forgetting in perceptually-driven memory, using lexical materials similar to those in the nonconceptual condition of study 1 of this dissertation (Bajo, Gomez-Ariza, Fernandez, & Marful, 2006). Participants underwent the standard retrieval-practice procedure with either category-exemplar pairs (e.g., animal-horse) or word stem-exemplar pairs (e.g., PE – pelota). As such, practiced and nonpracticed items were conceptually related in the former case (e.g., horse, dog, goat) and perceptually related in the latter case (e.g., pelota, peninsula, pedazo). Contrary to Perfect et al.s (2002) claim, they found significant retrieval-induced forgetting in both conditions, suggesting that resolving lexical and conceptual competition relies on similar mechanisms.

In sum, whether retrieval induces facilitation or forgetting is a function of how the information is processed at study (Chan, 2009). In the case of conceptual trial-and-error learning, error generation is in itself an act of integration which draws attention to the semantic relationship between cues, targets, and related concepts. This idea that targets and nontargets can be usefully bound converges on findings of change recollection, i.e. that proactive interference is best resolved in service of target memory among younger and older adults when they can remember the interfering items (see Wahlheim, 2014). Future studies are needed in order to elucidate the boundary conditions of integration in error generation as well as the influence of aging.
5.4 Conclusion

It is increasingly appreciated in psychology that aging is a dynamic process, and that its effects on memory cannot be explained by appealing to a single mechanism. This is because aging is associated with changes in cognitive hardware (e.g., brain structures) and software (e.g., information processing), but also shifts in the priorities that guide attention and memory. The question of whether older adults also learn from their mistakes captures the importance of carefully considering these spheres. Past research on errorless learning has almost exclusively viewed healthy aging as a model for understanding pathology, leading to conclusions that skirt its realities. My thesis aimed to recast the question of aging and errors in light of what we know to be beneficial among younger adults as opposed to detrimental among memory-impaired patient populations. This approach will become increasingly important as aging researchers continuously push back the lower limit of aging to better elucidate the early causes of age-related cognitive declines.

My dissertation findings are good news in a societal climate where older adults will be increasingly expected– and eager– to participate in late-life learning. My results suggest that aging does not fundamentally change how learning parameters affect memory, only the magnitude of change. This is encouraging for learning programs and technologies aimed at engaging individuals throughout the lifespan. However, it will be important for future research to broaden the scope of inquiry to include motivational and emotional factors that likely influence how older adults engage with mistakes.
Chapter 6

References


